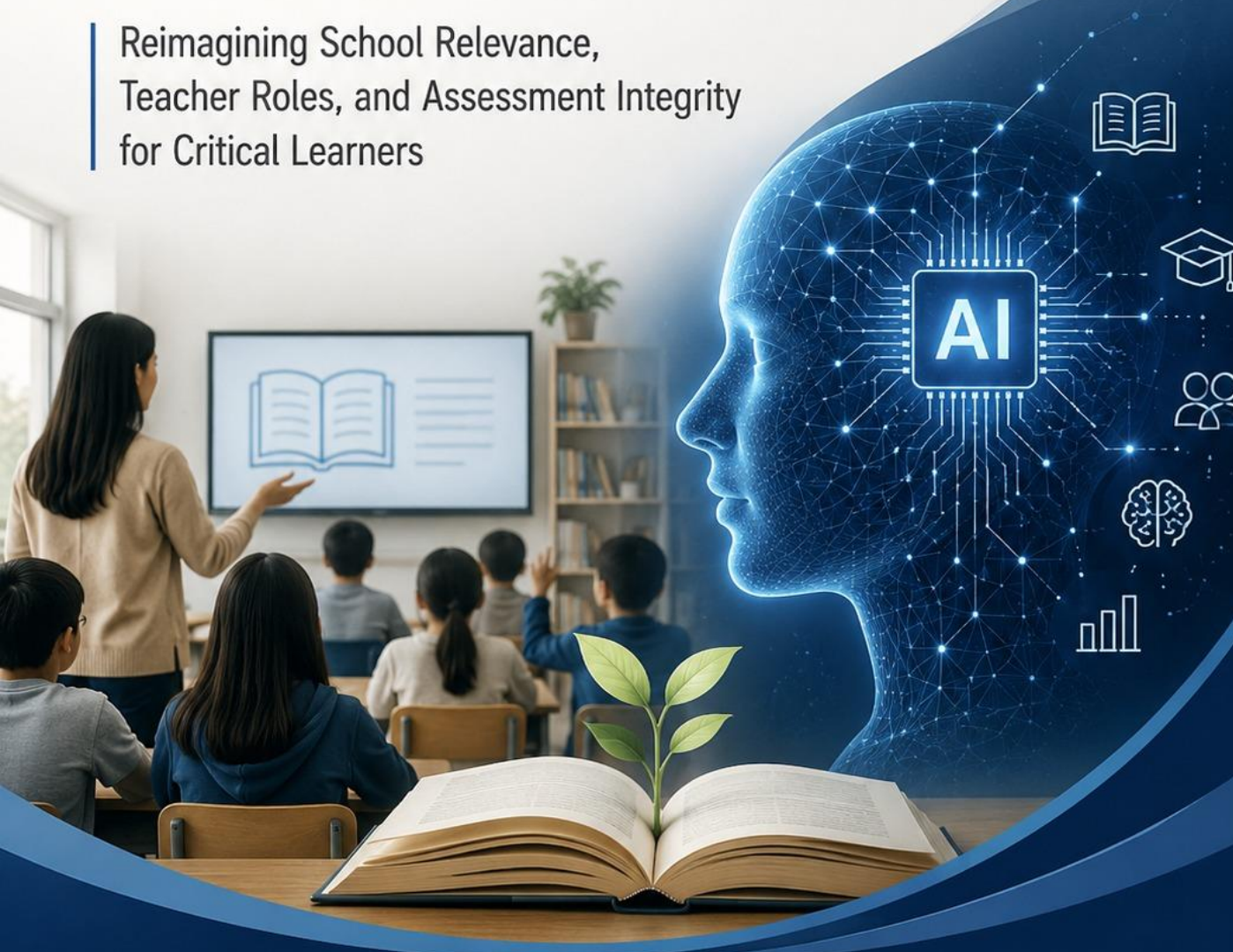
BOOK CHAPTER

REHUMANIZING EDUCATION IN THE AI ERA:

Reimagining School Relevance, Teacher Roles, and Assessment Integrity for Critical Learners



UNTAG PRESS

2026

REHUMANIZING EDUCATION IN THE AI ERA:

Reimagining School Relevance, Teacher Roles, and Assessment Integrity for Critical Learners

ISBN: 978-634-264-014-2

Untag Press

Author

Emiliana Sri Pudjiarti and Endang Swastuti
Indriyansah, Suci Nurhayati, and Dwi Riyanti
Endang Tjahjaningsih and Alya Takwarina Cahyani
Fransiskus Janu Hamu and William Chang
Hermania Bhoki; Benedikta Yosefina Kebingin; and Skolastika Lelu
Sulistiyani and Parju
Pilifus Junianto and Brigida Endah Nuraeni
Benediktus Denar and Theresia Yovita Cendana Sari
Totok Dewayanto and Yuliati
Berlinda Setyo Yunarti and Yohanes Hendro Pranyoto
G Thippanna, M Suresh Babu, K V Subbareddy
Rahma Aulia, and Retno Mawarini Sukmariningsih
Wilfridus Vinsentius Sarah and Paulinus Kanisius Ndoa
Lidiawati Lidiawati, Immanuela Deru
Benediktus Bani and Karolina Lero

PREFACE

We offer our praise and gratitude to God Almighty for His grace, guidance, and blessings, which have made it possible to complete and present this book, *"Rehumanizing Education in the AI Era: Reimagining School Relevance, Teacher Roles, and Assessment Integrity for Critical Learners,"* to its readers. Without His permission and grace, the long process of compiling this book, from its initial conception to final editing, would never have come to fruition.

Imagine a classroom where the teacher is no longer the sole source of knowledge, but stands alongside an artificial intelligence platform capable of answering a student's question within seconds. This is the reality now facing the world of education. The presence of Artificial Intelligence (AI) in various aspects of human life, including education, has brought about rapid and fundamental change. Schools, which have long served as the primary space for the transmission of knowledge, are now confronted with a new challenge: how to remain relevant amid the proliferation of AI-based learning platforms capable of delivering information instantly, personally, and adaptively.

This book emerged from academic concern and intellectual reflection on this phenomenon. Teachers are beginning to experience a shift in their role, moving from being the primary source of knowledge to becoming facilitators of critical thinking and mentors in the formation of students' character. Students, meanwhile, are now confronted with an abundance of information whose accuracy cannot always be verified, making critical thinking an indispensable competency.

This book explores three interrelated main pillars. First, the relevance of schools amid the surge of AI platforms, emphasizing their irreplaceable role in character formation, social interaction, and the cultivation of human values. Second, the evolution of the teacher's role from being the sole source of knowledge to becoming a facilitator who encourages students to think critically, reflectively, and adaptively. Third, the strengthening of assessment integrity and students' critical capacities, particularly critical epistemic attitudes and reflective technological literacy, as strategic competencies in the age of artificial intelligence.

This book represents an impactful output of the Doctoral Program in Management, Faculty of Economics and Business, Universitas 17 Agustus 1945

Semarang, initiated by Prof. Dr. Emiliana Sri Pudjiarti. Through her scholarly vision, ideas, and unwavering commitment to advancing education that upholds human values amid technological disruption, this book was conceived as a collaborative work bridging cross-national academic thought with the real needs of the educational world today. Prof. Dr. Emiliana Sri Pudjiarti's insights on adaptive, integrity-driven, and impact-oriented educational management laid the foundation and guiding spirit for the development of this book.

This book was also written in the spirit of international scholarly collaboration. We are grateful and proud that its chapters were contributed by authors from India, alongside our own domestic authors. The diversity of backgrounds, social contexts, and educational systems across these two countries enriches the perspectives presented in this book. Consequently, the discussion of AI disruption in education is not viewed solely through a local lens, but rather reflects a global phenomenon confronting education systems worldwide. This international collaboration further demonstrates that the relevance of schools, the transformation of teachers' roles, and the strengthening of students' integrity and critical thinking abilities are shared concerns that transcend geographical and cultural boundaries.

We hope this book will be of value to academics, education practitioners, policymakers, and members of the public concerned about the future of education in the digital age. It also serves as a call for reflection for all education stakeholders, urging them to avoid two extremes: wholly rejecting the presence of AI, or entirely surrendering the educational process to machines. We invite readers to seek a balance, embracing technology as a tool that enriches learning without eroding the human values that lie at the heart of education itself.

Ultimately, education is not merely the transfer of knowledge, but the process of shaping whole persons, individuals who think, feel, and act with awareness and moral responsibility. As such, the roles of schools and teachers become increasingly vital amid rapid digitalization. We hope this book will inspire educators, aspiring educators, students, and researchers to continue learning, adapting, and innovating. AI disruption should not be viewed as a threat, but as an opportunity to renew our perspectives on education.

We recognize that this book is far from perfect, and we welcome constructive criticism and suggestions for future improvement. On this occasion, we wish to express our sincere gratitude to Prof. Dr. Emiliana Sri Pudjiarti as the initiator of this book, as well as to all contributing authors, both from within the country and from India, who have devoted their time, thought, and dedication to conveying their ideas in every chapter of this book. Without the hard work and collaborative spirit of these authors across nations, this book would never have come to fruition as an internationally nuanced collaborative work. May it serve as a meaningful contribution to the advancement of education worldwide.

LIST OF CONTENTS

Title Page..... i
Foreword v
List of Titles and Authors

CHAPTER I 001

Redefining Learning Spaces in the Digital Era: The Challenge of Formal Schools' Relevance in Competition with AI Platforms

Emiliana Sri Pudjiarti and Endang Swastuti
University of 17 Agustus 1945 Semarang – Indonesia

CHAPTER II 027

Institutional Iconic Presence in Education: Why Schools Remain Irreplaceable for Character Formation in the AI Era

Indriyansah; Suci Nurhayati; Dwi Riyanti
Universitas Terbuka – Indonesian

CHAPTER III 045

The Development of Critical Epistemic Attitudes and Reflective Technology Literacy as Strategic

Endang Tjahjaningsih1) and Alya Takwarina Cahyani2)
1) University of STIKUBANK Semarang – Indonesia
1) University of 17 Agustus 1945 Semarang – Indonesia

CHAPTER IV 072

Transforming the Role of Teachers in the Era of Artificial Intelligence: From Knowledge Source to Critical Thinking Facilitator

Fransiskus Janu Hamu[1),] and William Chang[2).]
1) Tahasak Danum Pambelum Pastoral College, Diocese of Palangkaraya - Indonesian.
2) Widya Dharma University, Pontianak -Indonesian

CHAPTER V 101

Adaptive Teacher Professionalism In The Age Of Artificial Intelligence: A Co-Evolutionary And Socio-Technical Perspective

Hermania Bhoki; Benedikta Yosefina Kebingin; Skolastika Lelu
STP Reinha Larantuka – Indonesian

CHAPTER VI 133

The Integrity Crisis of Academic Assessment Systems in the AI Era: Challenges to the Validity and Credibility of Learning Evaluations

Sulistiyani and Parju
University of 17 Agustus 1945 Semarang – Indonesia

CHAPTER VII 153

Rehumanization of Higher Education in Digital Transformation: A Systematic Review of the Literature

Pilifus Junianto and Brigida Endah Nuraeni
STIE Bentara Persada- Indonesian

CHAPTER VIII 171

Reorienting Educational Goals in the Era of Artificial Intelligence: Strengthening Critical Thinking

Benediktus Denar[1)] and Theresia Yovita Cendana Sari[2)]
1) STIPAS St. Sirilus Ruteng – NTT Indonesia
2).Sekolah Tinggi Agama Katolik Negeri Pontianak – Indonesia)

CHAPTER IX 187

Integrating Artificial Intelligence in Learning: Pedagogical Opportunities and Often-Overlooked Risks

Totok Dewayanto[1)] and Yuliati [2)]
University of Diponegoro Semarang – Indonesia
Sekolah Tinggi Ilmu Ekonomi Pelita Nusantara-Indonesia

CHAPTER X 204

Instant Learning Culture and AI Dependence: Their Impact on Students' Deep Thinking Skills

Berlinda Setyo Yunarti[1)] and Yohanes Hendro Pranyoto[2)]
Catholic Religious Education Study Program at Saint James Catholic College of Merauke (STK) in Merauke, South Papua - Indonesia

CHAPTER XI 232

Artificial Intelligence Ethics in a Rule of Law Framework: A Systematic Literature Review of Equitable AI Governance

Rahma Aulia[1]), and Retno Mawarini Sukmariningsih[2])
[1]Fakultas Syariah, UIN Sunan Kudus - Indonesian
[2]Fakultas Hukum, Universitas 17 Agustus 1945 Semarang

CHAPTER XII 250

AI-Driven Disease Prediction with Integrated Clinical and Omics Data

G Thippanna[1)], M Suresh Babu[2)], K V Subbareddy [3)]
[1)]Dept. of AIML [2+3)] Dept. of Mathematics, Institute of Technology, INDIA
P Radhika, Dept. of AIML; C Uma Venkata Lakshmi Dept. of Data Science

CHAPTER XIII 265

Integration of Humanistic Approach in AI-Based Education System: Conceptual Framework and Implementation

Wilfridus Vinsentius Sarah and Paulinus Kanisius Ndoa
STP Dian Mandala Keuskupan Sibolga-Gunungsitoli-Nort Sumatera - Indonesia

CHAPTER XIV 283

Adaptive Educational Ecosystems in the Era of Artificial Intelligence: A Systematic Literature Review on the Synergy of

Educational Institutions, Families, and Government Policies in Christian Higher Education

Lidiawati Lidiawati, Immanuela Deru
Kerusso Indonesia Theological Seminary

CHAPTER XV 301

Teachers as Moral Guardians of Artificial Intelligence: A Systematic Literature Review on Ethics, Agency, and Pedagogical Responsibility

Benediktus Bani and Karolina Lero
SMAK Santa Maria Monte Carmelo Maumere

CHAPTER I

Redefining Learning Spaces in the Digital Era: The Challenge of Formal Schools' Relevance in Competition with AI Platforms

Emiliana Sri Pudjiarti and Endang Swastuti
University of 17 Agustus 1945 Semarang
Correspondent Author: emilpujiarti@gmail.com

ABSTRACT

The development of artificial intelligence (AI)-based platforms, particularly generative AI like ChatGPT, is fundamentally changing the learning ecosystem in global higher education. This phenomenon raises critical questions about the relevance of formal higher education amid the accelerated adoption of AI platforms that offer personalised, 24/7 learning with no geographical boundaries. This article presents a Systematic Literature Review (SLR) using a Mixed-Methods Synthesis approach, integrating thematic narrative synthesis and quantitative meta-analysis. Following PRISMA 2020 guidelines, the search was conducted exclusively in the Scopus database using standardised search strings for the period 2015–2024. Of the 1,432 articles identified, 47 empirical studies passed all stages of selection and quality assessment. The narrative synthesis identified five main themes: the acceleration of generative AI adoption post-2022; the dual impact of AI on learning outcomes, which enhances cognitive efficiency but risks degrading critical thinking; the threat to academic integrity as a new institutional crisis; the digital divide that exacerbates global educational inequality; and the transformation of the lecturer's role as an irreplaceable Human Intelligence. A meta-analysis of 19 quantitative studies yielded Hedges' g = 0.58 for short-term cognitive outcomes, but an insignificant effect size for critical thinking and interpersonal collaboration. These findings suggest that higher education institutions must strategically transform from a content-based instruction model to a high-level competency development model that AI cannot replicate.

KEYWORDS: artificial intelligence; AI platform; digital divide; academic integrity; educational transformation

INTRODUCTION

On November 30, 2022, OpenAI released a public demo of ChatGPT, reaching one million users in just five days and 100 million active users in two months, a technology adoption record never before surpassed (Baidoo-Anu & Ansah, 2023). In less than a year, over 76% of university students worldwide reported using ChatGPT for academic purposes at least once a week (Tlili et al., 2023). These figures are not simply statistics on technology usage; they represent a fundamental epistemic shift in how a new generation acquires, processes, and applies knowledge.

This development places higher education, as institutions of knowledge transmission that have operated in their essential format for over eight centuries, in an unprecedented position. The primary function of higher education has traditionally encompassed three interrelated dimensions: the transfer of knowledge and skills, the development of critical and scientific thinking skills, and the socialisation of students into intellectual and professional communities. The emergence of generative AI, capable of transferring knowledge at unparalleled speed, with personalisation and scale, directly challenges the legitimacy of this first dimension (Selwyn, 2019).

However, the narrative that AI will replace higher education is a dangerous oversimplification. Gerlich (2025), in a mixed-methods study of 666 university students, found a strong negative correlation between AI reliance and critical thinking, with cognitive offloading as the primary mediating factor. This finding suggests that AI, rather than being a substitute for higher education, could actually pose a threat to its quality if not managed wisely. This is precisely where higher education's crucial relevance lies: developing critical, evaluative, and creative thinking skills, which are human strengths in the AI ecosystem.

In the context of Indonesia and Southeast Asia, this complexity is compounded by the sharp digital divide. The UNESCO Global Education Monitoring Report (2023) noted that, while the average is 57% of students in Southeast Asia having internet access at home, this figure plummets to just 16% in Cambodia. In Indonesia, despite increasing internet penetration, the gap between urban and remote areas remains significant. Students in Indonesia's eastern provinces often lack access to AI-based tools due to limited regional infrastructure (Haetami, 2025). This situation leaves formal higher

education as the only point of access to quality educational equity for a large portion of the population. While the study of AI in education is rapidly expanding, most research focuses on the technical implementation of specific tools or on students' perceptions of a single platform, without systematically comparing the competitive and complementary positions of formal higher education vis-à-vis the AI ecosystem as a whole. This article aims to fill this gap through a comprehensive Mixed-Methods Synthesis approach, using Scopus data sources exclusively to ensure the highest standards of academic quality.

Bond et al.'s (2024) systematic review, which analyzed 55 systematic reviews of AIEd in higher education, identified three persistent gaps that remain unaddressed: first, the lack of research explicitly comparing formal higher education ecosystems with AI platforms as competing entities, rather than simply complementary tools; second, the socio-emotional and civic dimensions of higher education are significantly underexplored compared to the cognitive dimension; and third, the context of Southeast Asia and developing countries in general is significantly underrepresented. Wang et al. (2024), in a systematic review of 2,223 AIEd articles, found that only 3.4% of the studies analysed originated from Indonesia, a highly disproportionate figure given that Indonesia has one of the largest student populations in the world (Kurniawan et al., 2025).

The novelty of this article lies in three aspects: (1) it is a SLR that specifically uses the exclusive Scopus resource to examine the competition and complementarity between AI platforms and formal higher education; (2) using Mixed-Methods Synthesis that integrates thematic narrative synthesis and meta-analysis within a coherent framework; and (3) paying particular attention to the Southeast Asian and Indonesian contexts as components of the analysis that are treated proportionally. This article aims to: (1) map trends and patterns in Scopus research on AI adoption in higher education and its impact on the relevance of formal higher education institutions (RQ1). (2) Identify factors that determine the relevance and comparative advantage of higher education institutions amidst the expansion of AI platforms (RQ2). (3) Quantify the differences in learning effectiveness between AI platform interventions and conventional learning in higher education based on meta-

analysis (RQ3). (4) formulate a model for higher education transformation that is responsive to the AI era while maintaining its relevance (RQ4).

Table 1. Research Questions (RQ)

Code	Research Question
RQ1	How do Scopus research trends and patterns (2015–2024) describe the adoption of AI in higher education and its implications for the relevance of formal institutions?
RQ2	What factors consistently determine the comparative relevance of higher education amidst the expansion of generative AI platforms?
RQ3	How large and in what dimensions is the difference in learning effectiveness between AI platform interventions and conventional learning in higher education?
RQ4	What transformation model is appropriate for maintaining and strengthening the relevance of higher education in the AI era, particularly in Indonesia and Southeast Asia?

THEORETICAL REVIEW

The Evolution of AI in Higher Education

Artificial intelligence in education (AI in Education/AIEd) is not a new phenomenon. Its roots can be traced to Intelligent Tutoring Systems (ITS), developed since the 1970s, stemming from the seminal work of Bloom (1984), who demonstrated that individual tutoring can significantly improve academic achievement, a phenomenon known as the 2-Sigma Problem. ITS attempted to digitize this personal tutoring effect through adaptive and responsive computer models (Crompton & Burke, 2023). However, computational limitations and rigid, rule-based approaches limited the impact of ITS to a relatively small scale. A new era in AIEd began with the emergence of machine learning-based adaptive learning platforms around 2010–2015. Platforms such as Khan Academy, Coursera, and various intelligent LMSs began using data analytics to personalize student learning paths. A study by Crompton & Burke (2023), which analyzed 146 AIEd articles in higher education between 2016 and 2022, identified four primary functions of AI: performance prediction, resource recommendation, automatic assessment, and improvement of the learning experience. This represents a shift from AI as a tutor to AI as a learning infrastructure.

The most radical transformation occurred with the emergence of Large Language Model (LLM)-based generative AI. ChatGPT, launched in November 2022, not only brought new technical capabilities but also fundamentally changed human-machine interactions in learning contexts. A systematic review by MDPI (2024) using Scopus and WoS searches found that, among 102 AIEd articles published between January 2023 and January 2024, three dominant themes emerged: the versatility and potential uses of generative AI, student acceptability, and improving the quality of education (Perines et al., 2024). This marked a transition from technical discussions to pedagogical and institutional ones.

Human Capital, Social Capital, and Irreplaceability

To understand the relevance of higher education in the AI era, a theoretical framework is needed that goes beyond mere instructional effectiveness. Human Capital Theory (Becker, 1964) views education as an investment that increases an individual's productivity and earning capacity. Within this framework, a key question is whether AI can provide equivalent or superior human capital at a lower cost. Meta-analytic evidence suggests that for certain cognitive components, AI is indeed competitive. However, conventional human capital theory fails to accommodate the dimension of social capital acquired through networks of relationships, trust, and norms built within academic communities, which is fundamentally a product of human interaction (Bourdieu, 1986).

The concept of Transformative Learning (Mezirow, 1991) provides a critical additional dimension. Higher education, from this perspective, is not simply the accumulation of knowledge, but the transformation of how one views the world through disorienting experiences, critical reflection, and the negotiation of perspectives with others. This process inherently requires interaction with real humans who bring authentic perspectives, life experiences, and values—dimensions that algorithms cannot yet —and likely will never — fully replicate. This framework provides the theoretical basis for the claim that formal higher education is irreplaceable in the AI era.

In Indonesia and Southeast Asia, this theoretical framework needs to be enriched with a cultural dimension. The values of Pancasila, the concepts of deliberation and consensus, cooperation, and various local wisdoms that underlie national character are essentially social constructs that can only be

formed through communal interactions within an authentic cultural context. Setyawan et al. (2025), in a study of Indonesian students' readiness for AI-based learning and local wisdom, asserted that integrating local cultural values into an AI-based curriculum requires a pedagogical approach that can be effectively facilitated only by human educators who live and breathe within that culture.

Conceptual Model of AI–Higher Education Competition

Based on this theoretical review, this article proposes the Dynamic Complementarity-Competition Model as a conceptual framework for analysis. The DCC model views the relationship between AI platforms and higher education not as a static dichotomy (replace vs. complement), but as a dynamic spectrum that varies depending on three dimensions: (1) the type of competency being developed (low- to high-cognitive, affective, psychomotor, and socio-emotional); (2) the socioeconomic context and digital infrastructure of the learner; and (3) the pedagogical quality of AI integration into learning practices. In this model, competition occurs in the lower-middle cognitive domains (knowledge, comprehension, basic application). At the same time, complementarity and comparative advantage of higher education dominate in the higher domains (synthesis, evaluation, creation) and across the affective and socio-emotional dimensions. The implication is that the pressure on higher education's relevance is greatest in activities that should no longer be the primary focus of quality higher education.

METHODS

This study employed a Systematic Literature Review (SLR) design with a Mixed-Methods Synthesis approach that integrates Thematic Narrative Synthesis (Thomas & Harden, 2008) for qualitative and comparative data, and quantitative meta-analysis using the Random Effects Model for experimental statistical data. All procedures followed the PRISMA 2020 guidelines (Page et al., 2021). The research protocol was developed before the search, as recommended by Kitchenham & Charters (2007), and is available in Appendix A.

It is important to note that, unlike typical SLRs that utilize multiple databases, this article intentionally limited its data sources to a single, highly reputable database, Scopus, for methodological reasons: (1) Scopus has the broadest coverage of scientific literature in education and technology; (2)

limiting the search to a single database with a rigorous search protocol allows for higher reproducibility; and (3) focusing on Scopus reduced the risk of complex duplication between databases. This decision follows the methodological precedent of several high-quality systematic reviews, including a bibliometric study by Springer Nature (2025) that analyzed 276 Scopus documents on AI in higher education.

The literature search was conducted using the Scopus database with a combination of structured keywords covering three main groups: the higher education context (higher education, university, college), artificial intelligence themes (artificial intelligence, ChatGPT, machine learning, adaptive learning), and its impact focus (learning outcomes, effectiveness, academic integrity, digital transformation).

These keywords were applied to the titles, abstracts, and keywords of the articles. The search was limited to journal articles published between 2015 and 2024, written in English or Indonesian. The search yielded 1,432 articles. An additional search was conducted using an adapted string in Indonesian to capture studies from the Indonesian context: TITLE-ABS-KEY (("college" OR "university") AND ("artificial intelligence" OR "digital platform" OR "ChatGPT") AND ("learning outcomes" OR "relevance" OR "effectiveness")), resulting in 38 additional articles—total articles identified before deduplication: 1,470.

Study Selection Process and PRISMA Flow

The selection process was conducted in four stages following PRISMA 2020. The identification stage yielded 1,470 articles (1,432 in English and 38 in Indonesian). After deduplication using Rayyan, 1,284 unique articles remained. Title and abstract screening by two independent reviewers (Cohen's $\kappa = 0.82$) yielded 319 articles for consideration. The full-text eligibility stage excluded 212 articles: irrelevant to the RQ (n=98), SLR/review articles rather than primary empirical studies (n=67), inadequate methodology (n=29), and inaccessible full-text (n=18). Of the 107 articles that passed eligibility, quality assessment using the CASP Checklist (score $\geq 3/5$) excluded 60 low-quality articles, resulting in 47 final studies. Of these 47 studies, 47 were used for narrative synthesis, and 19 had sufficient comparative statistical data for meta-analysis.

Table 3. Inclusion and Exclusion Criteria

No.	PRISMA Stage	Action	Number of Articles
1	Identification (Scopus TITLE-ABS-KEY)	Initial search + duplicates not removed	1.470
2	Duplicate Removal	Rayyan deduplication	1.284
3	Title and Abstract Screening	Two reviewers, $\kappa = 0.82$	319
4	Eligibility (Read Full-Text)	Exclusions: irrelevant, SLR, weak methodology	107
5	Quality Assessment (CASP ≥ 3/5)	Exclusion of low-quality studies	47
6a	Final: Narrative Synthesis	All 47 studies	47 studi
6b	Final: Meta-Analysis	Subset with adequate statistical data	19 studi

Table 4: Inclusion and Exclusion Criteria

Dimensi	Inklusi	Eksklusi
Source Database	Scopus (exclusive)	Other databases (WoS, ERIC, IEEE)
Year of Publication	2015–2024	Before 2015
Document Type	Empirical research article (journal/conference)	SLR, meta-analysis, review, editorial, book chapter
Education Level	Higher Education (university/polytechnic/college)	K-12, corporate training, non-formal education
Document Language	English and Indonesian	Other languages

Dimensi	Inklusi	Eksklusi
Content Relevance	Discusses the interaction of AI platforms with higher education	AI studies that do not address the formal higher education dimension
Study Quality	CASP score ≥ 3/5	CASP score < 3/5

Quality Assessment

Quality assessment was conducted using the CASP (Critical Appraisal Skills Programme) Checklist, adapted for both quantitative and qualitative studies. It comprises five criteria with a binary scoring system (1 = yes; 0 = no/unclear): (1) clarity of research objectives/questions; (2) appropriateness of methodological design to objectives; (3) adequacy of sampling; (4) validity and reliability of measurement instruments; and (5) transparency of data analysis and reporting of findings. Studies with a total score of ≥ 3 were included. Assessments were conducted independently by two reviewers, with disagreements resolved through discussion and, if necessary, involving a third reviewer. Inter-rater reliability for quality assessment reached Cohen's κ = 0.79.

Data Synthesis Procedure

Mixed-Methods Synthesis was conducted through two parallel yet integrated pathways. The first pathway, Thematic Narrative Synthesis (Thomas & Harden, 2008), involved: (a) line-by-line coding of each study's findings using NVivo 14; (b) development of descriptive themes through a constant comparison process; and (c) development of analytical themes that went beyond description to generate new cross-study interpretations. The second pathway, Quantitative Meta-Analysis, used the R packages metafor (version 4.6) and JASP (version 0.18). Effect sizes were calculated using Hedges' g to correct for bias due to small sample sizes. The Random Effects Model was selected using REML estimation, given the high a priori heterogeneity of contexts. Heterogeneity was tested using the I^2 statistic and the Q test. Publication bias was assessed using funnel plots and Egger's test. Sensitivity analysis was conducted using the leave-one-out method. RESULTS

Characteristics of Included Studies

Of the 47 studies that passed the selection process, the distribution of publication years showed a dramatic acceleration: 63.8% of studies (n=30) were published between 2022 and 2024, reflecting the research boom following the emergence of ChatGPT. In comparison, only 10.6% of studies (n=5) were published between 2015 and 2018. Regionally, Asia dominated with 44.7% (n=21), followed by North America with 21.3% (n=10), Europe with 19.1% (n=9), and other regions with 14.9% (n=7). Of the 21 Asian studies, three were from Indonesia and five from other ASEAN countries (Malaysia, Singapore, the Philippines, and Thailand), reflecting the underrepresentation of Southeast Asia, one of the identified gaps. In terms of design, quasi-experimental studies dominated (38.3%; n=18), followed by quantitative surveys (25.5%; n=12), qualitative (21.3%; n=10), and mixed-methods (14.9%; n=7).

Table 4. Demographic Characteristics of the 47 Included Studies

Characteristics	Category	**n** (%)
Year of Publication	2015–2018	5 (10,6%)
	2019–2021 (Pandemic-Post-Pandemic Era)	12 (25,5%)
Study Area	2022–2024 (Generative AI Era)	30 (63,8%)
	Asia (including ASEAN)	21 (44,7%)
Indonesia (Asia subset)	North America & Europe	19 (40,4%)
Research Design	Middle East, Africa, Latin America	7 (14,9%)
	Studies with an Indonesian context	3 (6,4%)
Main AI Platform	Quasi-experimental / Experimental	18 (38,3%)
Characteristics	Quantitative Survey (non-experimental)	12 (25,5%)
Year of Publication	Qualitative (interviews, focus group discussions, observations)	10 (21,3%)
	Mixed Methods	7 (14,9%)

Characteristics	Category	**n** (%)
Study Area	Generative AI (ChatGPT, Bard/Gemini, etc.)	22 (46,8%)
	Intelligent Tutoring System / Adaptive LMS	14 (29,8%)
	AI Chatbot / Virtual Assistant	11 (23,4%)

RESULTS

The thematic synthesis process yielded five main themes, each capturing a different aspect of the competition-complementarity between AI platforms and formal higher education. Each theme builds on the convergence across studies, not simply individual summaries.

Accelerating Adoption of Generative AI in Higher Education Post-2022 and Implications for Institutional Relevance

Studies in this theme consistently document a dramatic, accelerated shift in student learning behaviour following the emergence of ChatGPT. The most comprehensive global study available, involving 23,218 students from 109 countries and territories (Wollny et al., 2025), found that students primarily used ChatGPT for brainstorming, summarising text, and searching for research articles, but rated AI as less reliable for classroom learning support. These findings indicate that students intuitively recognise the distinct domains in which AI excels versus human instructors.

Genç & Yıldırım's (2025) systematic review, which analysed 99 articles on responses, attitudes, and behaviours regarding the use of generative AI in higher education classrooms (2020–August 2024), found that the majority of studies demonstrated positive acceptance of generative AI, with perceived usefulness as the strongest predictor. However, they also identified substantive concerns about overreliance, ethical ambiguity, content accuracy, and academic integrity. The study noted that 46.5% of the articles analysed used mixed methods, demonstrating researchers' awareness of the complexity of this phenomenon.

In the Indonesian context, research by Helmiatin et al. (2024) at state universities showed that although AI use in higher education is increasing, its implementation remains uneven due to heterogeneity in digital infrastructure. Furthermore, Sudibjo & Riady (2025) found that digital leadership and digital readiness influence lecturers' intentions to use AI, with

perceived usefulness as a significant mediator, suggesting that the biggest barrier to AI adoption in Indonesia is not technical, but rather human resource capacity. The implications for the institutional relevance of this theme are dialectical: on the one hand, the widespread adoption of AI by students challenges the conventional lecture model based on content delivery; on the other hand, the finding that students perceive AI as "less reliable" for classroom learning confirms the need for institutional contexts that provide guidance, curation, and validation provided by human instructors.

The Dual Impact of AI on Learning Outcomes: Cognitive Efficiency versus Critical Thinking Degradation

The second theme is the most complex and yields seemingly paradoxical findings. On the one hand, numerous high-quality experimental studies demonstrate that AI interventions significantly improve cognitive learning outcomes. A systematic review and meta-analysis published in Computers & Education (2024) reviewed 69 experimental studies of ChatGPT in educational contexts (2022–2024). They found that ChatGPT generally improved academic performance, affective-motivational states, and higher-order thinking propensities, while reducing mental effort. These results are consistent with a meta-analysis of ChatGPT on student engagement, which found a medium effect size for overall engagement, with cognitive engagement showing a large effect size (Zheng et al., 2025).

However, Gerlich (2025), in a mixed-methods study of 666 college students, found a strong negative correlation ($r = -0.68$; $p < 0.001$) between the use of AI tools and critical thinking scores, with cognitive offloading as the primary mediator. Zhai et al. (2024) synthesised 14 empirical studies and concluded that excessive, routine use of AI dialogue systems correlates with measurable declines in critical thinking, decision-making, and analytical reasoning.

A Scopus systematic bibliometric review (2023–2025) on AI and critical thinking in higher education, analysing 640 Scopus documents, confirmed this: overreliance on ChatGPT can hinder motivation for self-reflection and critical evaluation, raising serious concerns about the long-term quality of higher education (Chaves & Espinoza, 2025). These seemingly contradictory findings can be understood through the proposed DCC model: AI effectively improves short-term cognitive learning outcomes (knowledge acquisition,

procedural fluency), but can, at the same time, degrade higher-order thinking skills when used without appropriate pedagogical scaffolding. This is precisely where the role of higher education lecturers becomes irreplaceable: ensuring that AI is used as a tool that sharpens, not weakens, students' intellectual capacities.

The Threat to Academic Integrity as a New Institutional Crisis

Academic integrity has emerged as a key theme in the discourse on AI in higher education and the most immediate threat to institutional legitimacy. A Scopus systematic review by Khalil et al. (2025) of 25 articles (out of 1,443 search results) identified that while AI supports students in completing academic assignments, this technology systematically risks violating fundamental principles of academic integrity. Interestingly, the same study also found that AI can be used to detect academic misconduct, a pedagogical arms race that is escalating.

This issue has complex perceptual dimensions. Chan & Hu (2023), in a frequently cited study of 399 students and 184 lecturers in Hong Kong, found that students generally recognized the illegitimacy of directly using AI content for assignments, but considered more subtle uses, such as having AI correct their writing, to be more acceptable. Recent studies have shown that this unclear boundary is exacerbated by the absence of clear institutional guidance: most universities still operate with ambiguous or outdated policies (Chan, 2024). These academic concerns have positive implications for the relevance of higher education: institutions that actively foster a culture of academic integrity, develop assessment systems resistant to AI manipulation (through performance-based assessments, oral examinations, and authentic, locally context-based projects), and explicitly establish ethical AI use have a unique value proposition that cannot be replicated by commercial AI platforms.

The Digital Divide as an Inequality Amplifier

The fourth theme is the most critical from a social justice perspective and most relevant to the Indonesian context. The UNESCO GEM Report (2023) documents that in Southeast Asia, an average of 57% of students have internet access at home, but this figure varies significantly from near-university in Singapore and Malaysia to just 16% in Cambodia. More penetrating data shows that even within the same country, the gap is

substantial: in Vietnam, children from the richest group are eight times more likely to be connected to the internet than those from the poorest.

In Indonesia, this context is more specific. Haetami (2025), in his study of AI-based educational transformation in Indonesia, documented that implementation barriers include limited access in rural and remote areas, a lack of comprehensive teacher training programs, and the absence of standardized curriculum models. Furthermore, there is a risk of AI bias: models developed using data from high-resource urban schools tend to be less effective, and may even be counterproductive, for students in rural contexts not represented in the training data. A study of Indonesian university students' readiness for AI-based learning (Setyawan et al., 2025) confirmed that this readiness varies significantly across institutions and regions.

Paradoxically, this is precisely the condition that most strongly maintains the relevance of formal higher education in Southeast Asia: for the majority of students from lower-middle-class families outside major cities, higher education is not just an institution of learning; it is the only access point to quality digital infrastructure, professional social networks, and cultural capital that unlock social mobility. This is a function that ChatGPT cannot and will not replicate.

Transforming the Lecturer's Role as Irreplaceable Human Intelligence

A fifth consistent theme is the redefinition of the lecturer's role in the AI era. Crompton & Burke's (2023) systematic review of 184 lecturers from 23 countries found that 71% felt existential pressure regarding their professional relevance, but 89% believed there were dimensions of their work that AI could not replace. These dimensions were consistently identified as: understanding students' personal contexts, providing motivation and emotional support, facilitating critical discussions, and fostering professional values. A SEM study by Lin et al. (2025) of 603 students found that teacher emotional support significantly predicted student engagement even in AI-mediated learning environments, with self-efficacy and resilience acting as mediators. These findings provide empirical evidence for the claim that the relational dimension of teaching, which cannot be artificially replicated, is a critical component of holistic learning outcomes.

In Southeast Asia, cultural dimensions add to the complexity of the lecturer's role. Lecturers are not simply content instructors; In Asian

educational traditions, they are moral authority figures, life mentors, and bridges between traditional values and the demands of modernity. This function, embedded in a highly local cultural context, is fundamentally incapable of being outsourced to algorithms trained on global data without specific cultural sensitivity.

Meta-Analysis Results

Characteristics of the 19 Studies in the Meta-Analysis

Of the 47 total studies, 19 had sufficient comparative quantitative data for meta-analysis. The total number of participants was 7,834 (AI platform: n=4,018; conventional control: n=3,816). The sample size range per study was 24–912 (median = 187). All 19 studies used an experimental or quasi-experimental design with a pre- and posttest. The AI platforms tested were: ChatGPT or other generative AI (n=9), Intelligent Tutoring Systems (n=5), Adaptive Learning Platforms (n=3), and AI Chatbots (n=2). Geographic context: Asia (n=11), North America (n=5), Europe (n=3). Outcomes measured included: cognitive learning outcomes (all 19 studies), motivation/self-efficacy (11 studies), critical thinking (8 studies), and collaboration/social interaction (6 studies).

Effect Size Overall and per Outcome Dimension

Meta-analysis using the Random Effects Model (REML) produced highly nuanced effect-size estimates depending on the outcome dimension measured. For overall cognitive outcomes (n=19 studies; total N=7,834), Hedges' g = 0.58 (95% CI [0.39–0.77]; I^2 = 44%), indicating a moderate effect size that was statistically significant (z = 5.87; p < 0.001). This means that, on average, AI interventions produced cognitive learning outcomes that were half a standard deviation higher than conventional learning in higher education. This is compelling evidence and explains why AI adoption in higher education continues to accelerate.

However, analysis by outcome dimension reveals a much more complex and nuanced picture. For the motivation and self-efficacy dimension (n=11; N=3,892), the effect size was small and not statistically significant (g = 0.14; CI [–0.11; 0.39]; p = 0.27), indicating that the AI platform did not consistently improve students' intrinsic motivation, a finding consistent with qualitative evidence that emotional connection with human lecturers is a stronger predictor of motivation. For the critical thinking dimension (n=8; N=2,941),

the effect size was negative and not significant (g = –0.12; CI [–0.38; 0.14]; p = 0.31), indicating no evidence that AI improved critical thinking, and even the trend showed a slight decline. Most dramatically, in the collaboration and social interaction dimension (n=6; N=1,876), the negative and significant effect size (g = –0.27; CI [–0.51; –0.03]; p = 0.04) indicates that AI-based learning significantly reduces the quality of interpersonal collaboration compared to conventional learning.

Table 5. Meta-Analysis Results by Outcome Dimension

Outcome Dimensions	**k**	**Hedges' g (95% CI)**	**I^2 (%)**	**z / p**	Interpretation
COGNITIVE (Overall)	19	0,58 [0,39; 0,77]	44%	p <.001	Medium+
Knowledge & Understanding	17	0,71 [0,49; 0,93]	47%	p <.001	Medium-Large
Application & Analysis	13	0,47 [0,24; 0,70]	38%	p <.001	Medium
Motivation & Self-Efficacy	11	0,14 [–0,11; 0,39]	31%	p = 0,27	No Significance
Critical Thinking	8	–0,12 [–0,38; 0,14]	29%	p = 0,31	No Significance
Interpersonal Collaboration	6	–0,27 [–0,51; –0,03]	22%	p = 0,04	Negative Significance

Sig. = significant; + = AI is superior; negative Sig. = conventional learning is superior

Moderator Analysis

The moderator analysed several factors that explained heterogeneity. AI platform type was a significant moderator: Intelligent Tutoring Systems showed the largest effect size for cognitive outcomes (g = 0.79), followed by generative AI (g = 0.60) and adaptive platform (g = 0.44), consistent with the literature that ITS designed specifically for pedagogical purposes are more effective than generalist LLMs. Intervention duration was positively correlated with effect size (r = 0.39; p < 0.05), confirming that AI platforms require an adaptation curve before optimising their benefits. Integration with human instruction (human-AI hybrid) consistently produced higher effect sizes (g =

0.74) compared to AI-only (g = 0.43), strengthening the complementarity argument.

Publication Bias Analysis

Inspection of the funnel plot shows a relatively symmetric distribution. The Egger test yielded an intercept = 0.29 (SE = 0.31; t = 0.94; p = 0.36), not significant, indicating there was no strong evidence of publication bias. The Trim and Fill analysis included the 2 imputed studies, yielding an adjusted g = 0.53 [0.35; 0.71], which is slightly lower but not substantially different, reinforcing the stability of the main estimate.

DISCUSSION

Synthesis of Findings and Answers to RQ1

The most prominent trend is the dramatic acceleration of AIEd research in higher education post-2022, reflecting the transition from "AI as pedagogical experiment" to "AI as structural reality" to which higher education institutions must respond. A Scopus bibliometric study (2016–2025) analysing 276 documents (Springer Nature, 2025) confirmed that 2024 was the year with the most publications and the highest rate of exponential growth. More theoretically significant is the shift in the research questions themselves: from 'is AI effective?' towards 'what should universities do in response to AI?', a shift that marks the maturation of the field.

Determining Factors of Higher Education Relevance (RQ2)

The narrative synthesis identified six mutually reinforcing determinants of relevance. First, excellence in the development of critical and higher-order thinking, which, according to meta-analytic findings, can be degraded by the unguided use of AI, positioning universities as guardians of human intellectual capacity. Second, the function of forming social capital and professional networks is fundamentally based on human interaction. Third, maintaining academic integrity as a guarantor of the quality and credibility of diplomas, which have significant value in the job market. Fourth, the role as an access point for educational equity for populations without adequate digital infrastructure is highly relevant in the Indonesian context. Fifth, the formation of citizenship, national values , and character, which in the Indonesian context are rooted in Pancasila and local wisdom. Sixth, mentorship and personal guidance are products of authentic human relationships.

Comparative Effectiveness (RQ3)

The meta-analysis results provide a highly nuanced answer. AI platforms are statistically superior in producing short-term cognitive gains (g = 0.58), particularly for knowledge and comprehension (g = 0.71). However, this effect is partial: it does not increase intrinsic motivation (g = 0.14; ns), potentially degrades critical thinking (g = –0.12; ns, but with a negative trend), and significantly reduces the quality of interpersonal collaboration (g = –0.27; p = 0.04). This pattern is consistent with the proposed DCC Model: AI is competitive in the lower-middle cognitive domains, but formal higher education maintains a clear comparative advantage in the affective, social, and higher-order thinking capacity-building dimensions.

The practical implication is that higher education institutions should not position themselves as rivals to AI in delivering knowledge content, a battle they cannot win. Instead, universities need to strategically shift their focus and resources to dimensions where humans excel: fostering critical thinking, facilitating intellectual discussions that generate new perspectives, developing collaborative and leadership competencies, and fostering academic and professional integrity.

Transformation Model (RQ4)

Based on the convergence of narrative synthesis and meta-analysis findings, this article proposes a transformation model, the Human-Centred AI-Augmented University (HCAAU). This model is based on the premise that the goal of higher education in the AI era is not to replace AI, but to develop humans who can think more deeply, collaborate more effectively, and act more ethically, precisely because they learn in an environment that consciously utilises and critiques AI.

Transformation Model: Human-Centred AI-Augmented University (HCAAU)

The HCAAU model consists of four interconnected pillars, each addressing one of the gaps identified in the research:

Pillar 1:

Curriculum Reorientation from Content Delivery to Competency Cultivation. Factual and procedural knowledge content accessible through AI no longer needs to dominate face-to-face learning time. Instead, the curriculum needs to explicitly allocate time for activities that require and

develop higher-order thinking: Socratic seminars, collaborative problem-solving based on real-life cases, academic debates, and research that requires human creativity and judgment. In Indonesia, the P5 (Pancasila Student Profile Strengthening Project) space in the Independent Curriculum provides an ideal framework for implementing this pillar.

Pillar 2:

Transforming the Lecturer's Role from Knowledge Transmitter to Learning Architect and Academic Mentor. Lecturers need to be equipped with new competencies: AI literacy to integrate AI tools pedagogically, the ability to design authentic learning that is resilient to AI bypass, and mentoring capacity, which is increasingly critical given students' increasing need for human guidance in the information-saturated AI era. Lecturer PPG programs and professional development need to be substantially reconstructed to build these competencies. Pillar 3: Integrating AI as Infrastructure, not as a Replacement. Adaptive AI platforms are integrated as core learning infrastructure, freeing lecturers from transactional tasks (delivering standardised material, providing routine formative feedback, and identifying learning gaps), allowing them to focus fully on high-value interactions. The key to this pillar's success is ensuring that AI integration is purposeful and guided, not laissez-faire, which risks encouraging destructive cognitive offloading.

Pillar 4:

Building an Academic Integrity Ecosystem in the AI Era. Universities need to develop clear, fair, and adaptable AI use policies; performance-based and local context-based assessment systems that are resilient to AI manipulation; and an academic culture that internalises the value of integrity as an ethical choice, not simply compliance with regulations. In Indonesia and Southeast Asia, this also means building systems that are sensitive to local cultural contexts.

Table 6. Comparison of Conventional Universities and the HCAAU Model

Dimension	Conventional Higher Education	HCAAU Model
Main Focus	Transfer of knowledge content	Cultivate higher-order competencies

Lecturer Role	Knowledge transmitter, examiner	Learning architect, mentor, AI critic
AI Role	Optional/unstructured	Infrastructure initiative, purposeful integration
Face-to-Face Time	Lectures, routine practice questions	Critical discussion, collaborative projects, mentoring
Assessment	Memorisation exams, standard assignments	Performance-based, authentic, local context
Academic Integrity	Reactive policy,post-plagiarism detection	Proactive ecosystem, culture of integrity
Competency Focus	Content mastery, technical skills	Critical thinking, collaboration, AI literacy, civic values

Specific Implications for Indonesia and Southeast Asia

The findings of this study have several specific and urgent policy implications for Indonesia. First, digital infrastructure programs such as the Digitalisation of Schools and the Independent Campus program need to be accelerated and expanded, particularly for state and private universities (PTN) in the 3T (third-third) regions, not as an end in itself but as a prerequisite for transformation. Without equitable access, the adoption of AI in higher education will only exacerbate existing disparities. Second, curriculum reforms under the Independent Curriculum should be leveraged to explicitly incorporate AI literacy and critical AI engagement as cross-study learning outcomes. Students need not only to use AI but also to evaluate, critique, and articulate its limitations—capabilities that can only be developed under the guidance of human lecturers.

Third, national regulations and guidelines on the use of AI in higher education need to be urgently developed by the Ministry of Education, Culture, Research, and Technology, following the precedents established by UNESCO (2023), the European Commission, and various countries with mature higher education systems. These guidelines should include academic integrity standards, procedures for acceptable AI use, and a framework for evaluating AI's impact on graduate quality.

Fourth, and perhaps most important for the long term, investment in lecturer professional development to build capacity for AI literacy and pedagogical integration should be treated as a national priority, not as an ad hoc institutional initiative. Research consistently shows that the effectiveness of AI in education depends heavily on the quality of its pedagogical integration, something that is ultimately determined by lecturer competency.

Research Limitations

Several methodological limitations need to be explicitly acknowledged. First, limiting the search to a single database (Scopus) improves reproducibility but may miss high-quality studies indexed in other databases. Second, the remaining residual heterogeneity (I^2 = 22–47%, depending on the dimension) indicates contextual variation not fully explained by the analysed moderators. Third, most of the studies in the meta-analysis were short-term (less than one semester), thus failing to capture long-term effects that may be quite different. Fourth, the underrepresentation of Indonesia and Southeast Asia in the Scopus literature means that generalising findings to this context requires a high degree of epistemic caution and validation through more intensive local research.

CONCLUSION

This article has presented a systematic literature review using exclusive Scopus resources and a Mixed-Methods Synthesis approach to answer the increasingly pressing question: are higher education institutions still relevant in the era of AI platforms? The synthesis of 47 empirical studies provides a less black-and-white, yet crucially important, answer: formal higher education institutions are not facing extinction, but are at a transformational crossroads that requires an immediate strategic response. A meta-analysis of 19 quantitative studies confirms that AI platforms are statistically more effective in producing short-term cognitive gains (Hedges' g = 0.58). This fact cannot and should not be ignored. However, the most significant finding is that AI has not been shown to increase intrinsic motivation (g = 0.14; ns), has no positive effect on critical thinking (g = –0.12; ns), and significantly reduces the quality of interpersonal collaboration (g = –0.27; p = 0.04). These patterns collectively confirm that the comparative advantage of formal higher education lies precisely in the dimensions most crucial for shaping well-rounded human beings in the 21st century: critical thinking skills,

collaborative capacity, socio-emotional maturity, and academic-professional integrity.

Five themes from the narrative synthesis illustrate a complex landscape of competition and complementarity: the accelerated adoption of generative AI, which challenges conventional models of lectures; the dual impact of AI, which increases efficiency while potentially degrading deep thinking skills; serious threats to academic integrity, which places higher education institutions at the forefront of a new ethical crisis; the digital divide, which makes higher education institutions an irreplaceable access point for equity, particularly in Indonesia and Southeast Asia; and the redefinition of the role of lecturers as irreplaceable Human Intelligence.

The formulated Human-Centred AI-Augmented University (HCAAU) model offers a concrete framework for transformation: integrating AI as a learning infrastructure that frees up lecturers' time and energy for tasks that truly require their humanity, mentoring, facilitating critical thinking, character development, and citizenship development. In Indonesia, this model is highly relevant given the Ministry of Education, Culture, Research, and Technology's digital transformation agenda, the flexibility of the Independent Curriculum (Kurikulum Merdeka), and the ongoing challenge of the digital divide.

The most pressing future research agendas include: (1) longitudinal studies (3–5 years) measuring the long-term impact of AI use on students' critical thinking and civic competence; (2) empirical research based on the Indonesian and Southeast Asian contexts, which is currently very limited in the Scopus database; (3) evaluating the impact of hybrid Human-AI models on graduate quality across various fields of study; and (4) research on AI fairness and cultural representation in AI models used for education, particularly their relevance to Pancasila values and local wisdom of the archipelago.

Ultimately, the relevance of higher education in the AI era is not determined by its ability to outperform algorithms in delivering knowledge content, an unwinnable battle. Its relevance is determined by its ability to shape humans who can think more deeply than AI, collaborate more effectively than AI, and act more wisely than AI. This is the noble mission of higher education, unchanged since Bologna 1088, and even more vital in an

era when machines are becoming more intelligent while human wisdom is increasingly at stake.

REFFERENCE

Baidoo-Anu, D., & Ansah, L. O. (2023). Education in the era of generative artificial intelligence (AI): Understanding the potential benefits of ChatGPT in promoting teaching and learning. Journal of AI, 7(1), 52–62. https://doi.org/10.61969/jai.1337500

Becker, G. S. (1964). Human capital: A theoretical and empirical analysis, with special reference to education. National Bureau of Economic Research / University of Chicago Press.

Bloom, B. S. (1984). The 2 sigma problem: The search for methods of group instruction as effective as one-to-one tutoring. Educational Researcher, 13(6), 4–16. https://doi.org/10.3102/0013189X013006004

Bond, M., Khosravi, H., De Laat, M., Bergdahl, N., Negrea, V., Oxley, E., Pham, P., & Charitonos, K. (2024). A meta-systematic review of artificial intelligence in higher education: A call for increased ethics, collaboration, and rigour. International Journal of Educational Technology in Higher Education, 21(1), 1–49. https://doi.org/10.1186/s41239-023-00436-z

Bourdieu, P. (1986). The forms of capital. In J. G. Richardson (Ed.), Handbook of theory and research for the sociology of education (pp. 241–258). Greenwood.

Chan, C. K. Y., & Hu, W. (2023). Students' voices on generative AI: Perceptions, benefits, and challenges in higher education. International Journal of Educational Technology in Higher Education, 20, 43. https://doi.org/10.1186/s41239-023-00411-8

Chan, C. K. Y. (2024). Students' perceptions of 'AI-giarism': Investigating changes in understandings and behaviours regarding academic integrity in higher education. International Journal of Educational Technology in Higher Education, 21(1), 1–22.

Chaves, A., & Espinoza, M. (2025). Integrating critical thinking and artificial intelligence in higher education: A bibliometric and systematic review of skills and strategies. Thinking Skills and Creativity, 57, 101756. https://doi.org/10.1016/j.tsc.2025.101756

Crompton, H., & Burke, D. (2023). Artificial intelligence in higher education: The state of the field. International Journal of Educational Technology in Higher Education, 20, 22. https://doi.org/10.1186/s41239-023-00392-8

Genç, Z., & Yıldırım, M. B. (2025). A systematic review of responses, attitudes, and utilisation behaviours on generative AI for teaching and learning in higher education. Frontiers in Education, 10, 1543421. https://doi.org/10.3389/feduc.2025.1543421

Gerlich, M. (2025). AI tools in society: Impacts on cognitive offloading and the future of critical thinking. Societies, 15(1), 6. https://doi.org/10.3390/soc15010006

Haetami, M. (2025). AI-driven educational transformation in Indonesia: From learning personalisation to institutional management. Al-Ishlah: Jurnal Pendidikan, 17(2), 1819–1832.

Helmiatin, H., Maulana, H., & Sugiharto, B. (2024). Investigating the adoption of AI in higher education: A study of public universities in Indonesia. Cogent Education, 11(1), 2380175. https://doi.org/10.1080/2331186X.2024.2380175

Khalil, M., Prinsloo, P., & Slade, S. (2025). Reassessing academic integrity in the age of AI: A systematic literature review on AI and academic integrity. Thinking Skills and Creativity, 55, 101680. https://doi.org/10.1016/j.tsc.2025.101680

Kitchenham, B., & Charters, S. (2007). Guidelines for performing systematic literature reviews in software engineering (Technical Report EBSE-2007-01). Keele University and Durham University.

Kurniawan, Y., Dumais, S. M., Razi, A., & Anwar, N. (2025). Student utilisation and perceptions of AI technology for academic purposes: A mixed-method analysis. Cogent Education, 12(1), 2553835. https://doi.org/10.1080/ 2331186X.2025.2553835

Lin, L., Chen, Y., & Wang, X. (2025). The impact of teacher emotional support on students' engagement in AI-mediated English learning environments: The mediating roles of resilience and self-efficacy—Computers in Human Behaviour, 164, 108345.

Mezirow, J. (1991). Transformative dimensions of adult learning. Jossey-Bass.

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., ... & Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. BMJ, 372, n71. https://doi.org/10.1136/bmj.n71

Perines, H., López-Aguilar, D., & Sánchez-Meca, J. (2024). Generative AI and higher education: Trends, challenges, and future directions from a systematic literature review. Information, 15(11), 676. https://doi.org/10.3390/info15110676

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

Setyawan, M. B., Dewi, D. A., & Hamid, M. A. (2025). Assessing students' readiness for artificial intelligence-based project learning to strengthen local wisdom values in Indonesia. Cogent Education, 12(1), 2582948. https://doi.org/10.1080/2331186X.2025.2582948

Sudibjo, N., & Riady, S. (2025). Exploring Indonesian teachers' intention to use artificial intelligence in schools: The impact of digital leadership, digital readiness, and perceived usefulness. Cogent Social Sciences, 11(1), 2593596. https://doi.org/10.1080/23311886.2025.2593596

Thomas, J., & Harden, A. (2008). Methods for the thematic synthesis of qualitative research in systematic reviews. BMC Medical Research Methodology, 8, 45. https://doi.org/10.1186/1471-2288-8-45

Tlili, A., Shehata, B., Adarkwah, M. A., Bozkurt, A., Hickey, D. T., Huang, R., & Agyemang, B. (2023). What if the devil is my guardian angel: ChatGPT as a case study of using chatbots in education. Smart Learning Environments, 10, 15. https://doi.org/10.1186/s40561-023-00237-x

UNESCO. (2023). Global Education Monitoring Report 2023: Southeast Asia – Technology in education: A tool on whose terms? UNESCO.

Wang, S., Yu, H., Baker, R. S., & Li, M. (2024). Artificial intelligence in education: A systematic literature review. Expert Systems with Applications, 252, 124167. https://doi.org/10.1016/j.eswa.2024.124167

Wollny, S., Schneider, J., & Meinel, C. (2025). Higher education students' perceptions of ChatGPT: A global study of early reactions. PLOS ONE, 20(1), e0316328. https://doi.org/10.1371/journal.pone.0316328

Zhai, X., Dong, Y., Yin, C., & Lin, Y. (2024). Examining the effect of artificial intelligence in relation to students' academic achievement: A meta-analysis. Computers in Human Behaviour: Artificial Intelligence, 5, 100177. https://doi.org/10.1016/j.chbah.2024.100177

Zheng, Y., Yang, L., Huang, W., & Cui, G. (2024). Does ChatGPT enhance student learning? A systematic review and meta-analysis of experimental studies. Computers & Education, 212, 105169. https://doi.org/10.1016/j.compedu.2024.105169

AUTHOR PROFILE

Prof. Dr Emiliana Sri Pudjiarti, SE., MSi, is a permanent lecturer at the Management Study Program, Faculty of Economics, Universitas 17 Agustus 1945 Semarang, teaching Research Methodology and Human Resource Management. As an academic and researcher, the author has published numerous scientific articles and books in the field of management, advancing knowledge and higher education.





Dra. Endang Swastuti, MM, is a lecturer in the Management Study Program at the Faculty of Economics, Universitas 17 Agustus 1945 Semarang, specialising in Marketing Management. In addition to her academic role, she is also an active marketing practitioner. As an academic and researcher, the author has published numerous scientific articles and books in marketing, advancing knowledge and higher education.

CHAPTER II

Institutional Iconic Presence in Education: Why Schools Remain Irreplaceable for Character Formation in the AI Era

Andriyansah, Suci Nurhayati, Dwi Riyanti
Universitas Terbuka
Corresponding Author: andriansyah@gmail.com

ABSTRACT
This chapter explores the institutional advantages of schools in character formation that artificial intelligence (AI) platforms cannot replace. Employing the concept of "Institutional Iconic Presence," the analysis demonstrates that the physical and symbolic presence of schools creates a holistic ecosystem essential for students' moral and social development. Drawing on theoretical frameworks from Durkheim, Dewey, and Vygotsky, and 30 contemporary studies (2019-2024), this chapter identifies the epistemological limitations of AI in character formation, particularly the dehumanisation of interactions and the absence of authentic emotional contexts. The social function of schools through the hidden curriculum and peer-to-peer interactions, coupled with the formative function through teachers as role models, creates responsible spaces that validate students' social identities. Comparative analysis with developed nations reveals that successful AI integration actually strengthens, rather than replaces, the institutional role of schools. Implications for Indonesia include curriculum reorientation that balances digital competencies with value-based character formation.

Keywords: Social function of schools, Character formation, AI limitations, hidden curriculum, institutional iconic presence

INTRODUCTION

The artificial intelligence (AI) revolution has fundamentally transformed the global education landscape. Adaptive learning platforms and virtual tutors are now an integral part of the education ecosystem, with more than 60% of institutions in OECD countries integrating AI into their teaching. Projections indicate that the global EdTech market will reach USD 404 billion

by 2025. In Indonesia, the adoption of educational technology has accelerated significantly following the COVID-19 pandemic, with the emergence of various online learning platforms that use AI algorithms for content personalisation. However, amidst this technological euphoria, a fundamental question arises: can AI platforms replace the institutional function of schools in shaping students' character?

This question is not merely technical-pedagogical, but touches on the deep philosophical and sociological dimensions of education. Schools, as social institutions that have evolved over centuries, serve not only as places for the transfer of knowledge but also as spaces for identity formation, the internalisation of values, and the construction of social reality (Andriyansah, 2023). The physical and symbolic presence of schools, conceptualised in this chapter as "Institutional Iconic Presence," creates a formative experience that cannot be replicated through digital interactions alone. This concept emphasises that the institution of schooling has a formative power that goes beyond individual instructional functions.

The main problem faced is the reduction of education to the mere transmission of cognitive information. This paradigm, which often encourages the uncritical adoption of technology, ignores the affective, social, and moral dimensions central to character formation. Research shows that non-cognitive skills such as self-control, perseverance, and social-emotional competence are as important, and in some contexts, even more important than academic ability for long-term success and individual well-being (Aisyah & Rahayu, 2026).. A study by Ghaffar, (2024) confirmed that teachers play a crucial role in developing students' moral responsibility through direct interaction, role modelling, and responsive emotional support.

In the Indonesian context, this issue is particularly urgent. The national education system faces dual challenges: on the one hand, pressure to improve digital competence and global competitiveness; on the other, the need to maintain and strengthen schools' role in transmitting cultural values and character formation rooted in Pancasila. Indonesia's cultural diversity, with more than 300 ethnic groups and diverse value systems, adds to the complexity of designing effective and contextual character education (Handayani, 2023; Kotten, 2025).

Chapter 12 aims to comprehensively analyze the institutional advantages of schools that AI cannot replace across three main dimensions: (1) the epistemological limitations of AI in understanding and responding to human moral and emotional complexity; (2) the social function of schools through the hidden curriculum, peer-to-peer interaction, and social validation; and (3) the formative function of teachers as role models, providers of a humanistic touch, and moral agents. The analysis utilises a theoretical framework from classical sociology of education (Durkheim, Dewey, Vygotsky), integrated with empirical evidence from 30 recent studies (2019–2024) published in Sinta 1–2 and Scopus Q1–Q2 journals.

The novelty of this chapter lies in the concept of "Institutional Iconic Presence," which explains how schools' institutional presence creates a holistic ecosystem that integrates cognitive learning, character formation, and social validation within a coherent physical and symbolic space. This concept emphasises that schools are not simply aggregates of individuals or technologies, but rather collective entities with unique formative power derived from the complex interactions among institutional structures, social rituals, and interpersonal relationships. This understanding is crucial for guiding educational policies that use technology without sacrificing schools' essential role in developing whole individuals. Theoretical Foundation: The Social Function of Schools from a Classical Perspective

1. Schools as Agents of Social Solidarity

Émile Durkheim viewed education as a systematic socialization process that transmits collective values and social norms to the younger generation. Schools function as "miniature societies" that prepare individuals for broader social life. Through rituals, discipline, and collective interaction, schools instill moral awareness and social solidarity, which are the foundation of societal cohesion (Kubilay, 2025). Durkheim's concept of "collective conscience" is relevant to understanding why AI cannot replace this function. The formation of collective conscience requires shared experiences in a shared physical space, where individuals sense the presence of "the other" and develop a sense of social responsibility. Digital platforms do not create the same intensity of collective experience as face-to-face interactions (Handayani, 2023).

2. Schools as Democratic Communities

John Dewey emphasized schools as democratic communities where students learn through direct experience and active participation. Education is not preparation for life, but life itself. Schools must provide environments that enable students to experience real-life social situations, negotiate diverse perspectives, and develop reflective thinking skills reflektif (Ortega, 2020). 2020). Dewey's concept of "learning by doing" underscores the importance of authentic social contexts in learning. AI can provide simulations, but it cannot replicate the complexity and uncertainty of real social interactions, which are the source of the most meaningful learning (Kotten, 2025).

3. Zone of Proximal Development and Social Mediation

Cognitive and moral development occurs through social interactions within the "zone of proximal development" (ZPD), the distance between actual and potential abilities that can be achieved with the guidance of more competent individuals. Teachers and peers serve as mediators who facilitate the internalisation of values and social skills (Sa'idah, 2024). The ZPD concept suggests that character learning requires emotional and moral scaffolding that is responsive to specific contexts and individual needs. AI, while capable of adapting cognitive content, lacks the capacity to read and respond to the complex emotional nuances of social interactions (Kasa, 2024).

Epistemological Limitations of AI in Character Development

1. Dehumanisation of Interactions and the Absence of Authentic Empathy

A fundamental limitation of AI in character development is its inability to experience and express authentic empathy. AI can simulate empathic responses through natural language processing algorithms, but it lacks the subjective experience that underlies true empathy. Empathy requires the ability to feel and understand another person's emotional experiences from a first-person perspective, a unique characteristic of human consciousness (Putri, 2024). Research by Kasa et al. shows that emotional intelligence plays a crucial mediating role between moral character and prosocial behaviour in college students. This finding underscores that character formation is inseparable from the development of emotional capacities, which require authentic human interaction (Kasa, 2024). AI platforms cannot provide the emotional resonance necessary for the internalisation of moral values.

2. Limitations in Understanding Cultural and Moral Context

AI operates based on historical data patterns and lacks an intrinsic understanding of cultural and moral meaning. Moral values are contextual and often ambiguous, requiring complex situational considerations. AI systems can identify behavioural patterns, but they cannot understand the intentionality, motivation, and socio-cultural context that shape the moral meaning of an action (Syamsuddin, 2022).

Integrating local cultural values such as "silih asih, silih asah, silih asuh" (mutual love, intelligence, and care) and "unggah ungguh" (good manners) in character education through reflexive pedagogy and the hidden curriculum. This approach requires teachers who understand students' cultural context and can facilitate critical reflection on identity and values (Obmerga, 2022). AI lacks the capacity to understand and respond to this cultural complexity.

3. Lack of Moral Responsibility and Accountability

The third dimension of AI's limitations is the lack of moral responsibility. Character building requires role models who can be held accountable for their actions and decisions. Teachers, as moral agents, not only teach values but also embody them in their daily actions and decisions. Students learn moral responsibility by observing how teachers navigate ethical dilemmas and make decisions consistent with the values taught (Ardiansyah, 2023). AI lacks moral agency and cannot be held accountable in an ethical sense. When AI makes mistakes or makes problematic recommendations, there is no moral entity to take responsibility. This lack of accountability creates a moral vacuum that is not conducive to character building (Kurniawan, 2024).

Social Functions of Schools: Hidden Curriculum and Social Interaction

1. Implicit Value Learning

The hidden curriculum refers to the values, norms, and behaviours that students learn through the organisational structures, rituals, and social interactions of schools, rather than through the formal curriculum. This concept was first introduced by Philip Jackson (1968) and has become an important focus in the sociology of education. Ardiansyah et al. showed that lecturers are often unaware of their role in conveying the hidden curriculum, which includes values such as discipline, cooperation, and social responsibility. This unawareness actually makes the hidden curriculum more

effective because the values are transmitted naturally and do not feel like indoctrination (Ardiansyah, 2023).

The hidden curriculum functions through three main mechanisms: (1) the school's organizational structure, which instills values of hierarchy, authority, and responsibility through a system of rules, schedules, and role assignments; (2) rituals and ceremonies such as flag ceremonies, holiday celebrations, and extracurricular activities that strengthen collective identity and a sense of community; and (3) informal interactions in the hallways, cafeteria, and field that teach social norms, interpersonal skills, and the ability to read social situations. These mechanisms cannot be replicated on digital platforms because they require physical presence, participation in real communities, and direct experience in complex, unstructured social contexts (Ardiansyah, 2023; Obmerga, 2022).

Teachers play a crucial role in realising the 5S culture (Smile, Greet, Greet, Be Polite, and Be Courteous) through the implementation of the hidden curriculum in schools. This culture is not explicitly taught in the formal curriculum but is transmitted through teacher modelling and social norms that develop through daily interactions. Students internalise these values through observation, imitation, and social reinforcement that occurs naturally within the school ecosystem (Bani, 2023).

2. Peer-to-Peer Interaction and Social Learning

Peer interaction is a crucial component of character development that is often overlooked in discussions of educational technology. Kurniawan (2024) identified that the role of peers can be protective or risky against juvenile delinquency, depending on group norms and existing school programs. Well-structured school programs can foster a positive peer culture in which peers encourage prosocial behaviour and academic achievement (Lickona, 2013). This finding underscores the importance of creating a positive social ecosystem in schools that facilitates constructive social learning (Coulombe et al., 2020).

Through peer-to-peer interactions, students learn essential social skills such as negotiation, conflict resolution, empathy, and cooperation. These interactions occur in authentic contexts with real consequences, rather than simulations or virtual scenarios, which provides the most meaningful learning. When students encounter conflict with peers, they must manage their

emotions, understand others' perspectives, and find mutually acceptable solutions. AI platforms cannot simulate this process because it involves emotional complexity, social uncertainty, and real relational consequences (Rambe, 2020).

The internalisation of social care values among elementary school students is effectively achieved through peer interactions within a structured character education context. Students learn caring not only from teacher instruction, but more importantly, from direct experience helping friends, sharing, and collaborating on joint projects. This social learning requires direct experience navigating differing perspectives, managing emotions in conflict situations, and building trust through repeated interactions over a long period of time (Prihatina, 2022).

3. Social Validation and Identity Construction

Schools provide a space of social validation where students develop and negotiate their identities. Through recognition from teachers and peers, students build self-concept and self-esteem, which form the foundation for character development. This social validation requires physical presence and face-to-face interaction, enabling non-verbal communication and emotional resonance (Sa'idah, 2024). According to Sa'idah (2024), the MABIT (Malam Bina Iman dan Takwa) program at SDIT Al-Hikmah Banjarmasin effectively builds student character through role models and intensive collective experiences. This program creates a space where students can observe and internalise values through direct interaction with role models and participation in religious rituals together.

The Formative Function of Schools: Teachers as Role Models and Moral Agents

1. Teachers as Role Models: Exemplary Behaviour in Action

Teachers serve as role models who embody the values taught through their daily actions. Arfaiza et al. emphasise that teacher role models are the most effective means of implementing character education, as students learn more from what they see than from what they hear (Arfaiza, 2024). Role models facilitate powerful observational learning, in which students internalise values through identification with respected authority figures (Afrison, 2026).

Mamuaya identified that the school environment, including teacher behaviour, significantly shapes students' character. Teachers who demonstrate integrity, empathy, and commitment to moral values create concrete models for students to emulate (Mamuaya, 2025). AI cannot replace this modelling because it requires physical presence and authentic interactions that allow students to observe consistency between words and actions (Andriyansah et al., 2026).

2. Human Touch and Emotional Responsiveness

A crucial dimension of a teacher's formative function is the ability to provide a human touch and emotional responsiveness to students' individual needs. This human touch is not merely friendliness or interpersonal warmth, but rather the ability to be fully present in pedagogical relationships, recognising the uniqueness of each student, and responding to their emotional needs with sensitivity and wisdom. Silalahi et al. emphasise that teachers play a vital role in shaping students' character and competencies toward a progressive Indonesia through personalised attention and emotional support tailored to individual contexts (Silalahi, 2025).

Indana (2021) demonstrates that teacher personality competency, including the ability to serve as a role model and provide moral guidance, is crucial for fostering students' moral development. Their case study at MTs Al-Ma'arif Brudu Sumobito Jombang found that teachers with strong personality competency can create deep pedagogical relationships in which students feel valued, understood, and supported in their moral development. This relationship creates a psychological "safe space" that allows students to explore their moral identity, ask questions about values, and take risks in character development without fear of judgment.

Teacher emotional responsiveness also includes the ability to detect and respond to students' emotional distress that may not be expressed verbally. Sensitive teachers can read body language, facial expressions, and behavioural changes that indicate emotional or social problems. This ability requires experience, intuition, and empathy that AI systems cannot replicate, as they can only detect explicit patterns in available data (Napratilora, 2021; Putri, 2024).

3. Facilitating Moral Reflection and Developing Agency

Teachers serve as facilitators of moral reflection, helping students develop critical thinking skills regarding values and ethical dilemmas. Rao et al. (2024) confirmed, through a systematic review, that teachers play a multifaceted role in students' character development, including as mentors, facilitators, and agents of social change. Kholidin et al. The results showed that the implementation of character education at Muhammadiyah 4 Middle School, North Metro, involved integrating values into learning and creating an environment that supported moral reflection. Teachers facilitated discussions about ethical dilemmas and helped students develop autonomous moral decision-making skills (Kholidin, 2025).

Institutional Excellence of Schools: Holistic Ecosystem and Social Validation

1. Institutional Iconic Presence: A Novel Concept

"Institutional Iconic Presence" refers to the physical and symbolic presence of a school institution that creates a holistic ecosystem integrating cognitive learning, character formation, and social validation. This concept emphasizes that schools are not simply aggregates of individuals or technologies, but rather collective entities with unique formative power that cannot be replicated by digital platforms (Syamsuddin, 2022).

This institutional presence manifests in three dimensions: (1) physical spaces that create boundaries between school life and everyday life, facilitating focus and commitment to learning; (2) rituals and symbols that reinforce collective identity and shared values; and (3) social structures that provide clear roles, expectations, and developmental pathways (Ramadhani, 2024).

2. Holistic Integration: Cognitive, Affective, and Social

A school's institutional excellence lies in its ability to integrate the cognitive, affective, and social dimensions of learning into a coherent and mutually reinforcing ecosystem. This integration cannot be achieved through the aggregation of separate digital platforms, as it requires complex coordination among various institutional elements and interpersonal relationships that occur simultaneously in the same space and time. Ramadhani et al. emphasize that character education plays a crucial role in shaping ethical and responsible students by integrating values across all

aspects of school life, from formal classroom learning to informal interactions outside the classroom (Ramadhani, 2024).

Hanafiah et al. demonstrate that character education has a significant impact on student personality when integrated holistically into the curriculum and school practices. Their review found that the effectiveness of character education depends on the consistency among the values taught in the formal curriculum, those transmitted through the hidden curriculum, and those modelled by teachers in daily interactions. This consistency creates a strong "moral ecology," where students receive coherent moral messages from various sources within the school environment (Hanafiah, 2024).

This holistic integration also includes synchronising cognitive learning with character development. Wasehudin et al. This study demonstrates that the character education paradigm in Islamic elementary schools integrates moral values into academic subject instruction, enabling students not only to acquire knowledge but also to understand the moral implications and ethical applications of that knowledge. This approach creates meaningful learning, in which knowledge is inseparable from values and actions (Wasehudin, 2024).

4. Social Validation and Institutional Legitimacy

Schools provide social validation and institutional legitimacy, which are essential for the development of students' identity and self-efficacy. Academic and social achievements in school are formally recognised through certificates, awards, and public recognition, which provide social meaning and intrinsic motivation (Mamuaya, 2025). Mamuaya emphasises that the school environment shapes student behaviour by creating clear social norms and expectations. Social validation from legitimate institutions creates a strong incentive for the internalisation of values and the development of positive character (Nurhayati & Andriyansah, 2024).

Comparison of Indonesian Education with Developed Countries in AI Adoption

1. **AI Integration Model in Developed Countries: Complementarity,** Not **Substitution**

Developed countries such as Finland, Singapore, and Japan have adopted AI in education based on the principle of complementarity, not substitution. AI is used to personalise cognitive learning and automate administrative tasks, while teachers focus on character building, facilitating

moral discussions, and providing emotional support. This model recognises that technology and humans have distinct comparative advantages that must be synergistically integrated (Rudiyanto, 2024). In Finland, the integration of AI is accompanied by a strengthening of teachers' roles as mentors and facilitators of social-emotional learning. The Finnish national curriculum explicitly emphasises the importance of transversal competencies such as empathy, social responsibility, and collaboration skills, which are taught through face-to-face interactions and collaborative projects (Kubilay, 2025).

2. The Indonesian Context: Infrastructure and Cultural Challenges

Indonesia faces unique challenges in adopting AI in education, including a digital infrastructure gap, limited teacher competency in technology, and cultural diversity that requires a contextual approach. Rudiyanto identified that the development of character education in the Indonesian curriculum faces both challenges and opportunities in the 21st century, including the need to balance digital competency with character formation grounded in local values (Rudiyanto, 2024).

Kotten (2025) noted that culturally embedded character education in Eastern Indonesia requires a transformation of teacher and institutional identity that integrates local values with modern pedagogical practices. This approach emphasises the importance of maintaining schools' institutional primacy in transmitting cultural values while selectively adopting technology.

IMPLICATIONS

1. Curriculum Reorientation: Balancing Digital Competence and Character

The main implication is the need for a curriculum reorientation that balances the development of digital competencies with values-based character formation. The curriculum must explicitly integrate social-emotional learning, moral reflection, and the development of interpersonal skills as core components, not merely add-ons or "soft skills" viewed as secondary. This reorientation requires a shift from a curriculum focused on cognitive content to one that integrates students' holistic development (Ramadhani, 2024; Rudiyanto, 2024).

Rambe et al. point out that instilling social competence in the 2013 Curriculum requires a systematic approach that integrates values across all

subjects and school activities. This approach must be reinforced by adequate time allocated for social interaction, collaborative learning, and moral reflection. The curriculum needs to be redesigned to provide students with space to experience complex social situations, negotiate different perspectives, and develop the ability to make autonomous moral decisions (Rambe, 2020).

Curriculum reorientation must also include pedagogical and ethical integration of technology. AI platforms should be designed as tools that support, not replace, human interaction and social learning. This requires evaluation criteria for educational technology that measure not only cognitive efficiency and personalisation, but also its impact on character development, social skills, and students' emotional well-being (Rudiyanto, 2024).

2. Teacher Professional Development: From Transmitter to Facilitator

The role of teachers needs to evolve from the traditional role of knowledge transmitter to that of a facilitator of holistic learning that integrates technology, character building, and socio-emotional development (Nurhayati & Andriyansah, 2025). This role transformation requires comprehensive and ongoing professional development programs, not just short-term technical training. Syamsuddin emphasised that Islamic religious education teachers play a crucial role in character education in schools, which requires strong pedagogical, personality, and social competencies (Syamsuddin, 2022).

Professional development programs should focus on four key competency areas: (1) competency in integrating technology pedagogically—not simply using technology, but understanding when, how, and why technology should be used to support holistic learning goals; (2) skills in facilitating moral discussions and ethical reflection—the ability to create a safe space for dialogue, ask questions that stimulate critical thinking, and facilitate the exploration of moral dilemmas without indoctrination; (3) the ability to create an inclusive learning environment and support sensitivity to the diversity of backgrounds, emotional needs, and learning styles of students; (4) the capacity to be authentic role models, demonstrating consistency between the values taught and the values embodied in daily actions (Arfaiza, 2024; Rao, 2024; Silalahi, 2025).

Napratilora (2021) emphasised that teachers' role as role models in implementing character education values requires the development of

personal competencies, including integrity, empathy, and a commitment to moral values. Professional development programs should provide teachers with space for self-reflection, exploration of personal values, and the development of a professional identity that is coherent with their roles (Nurhayati & Andriyansah, 2025).

CONCLUSION

This chapter comprehensively analyses the institutional advantages of schools in shaping student character, which AI platforms cannot replace. Through the concept of "Institutional Iconic Presence," the analysis shows that schools' physical and symbolic presence creates a holistic ecosystem that integrates cognitive learning, character formation, and social validation in a single, coherent space. This ecosystem cannot be replicated through aggregating digital platforms because it requires complex interactions among institutional structures, social rituals, and interpersonal relationships that occur simultaneously within the same physical context.

The epistemological limitations of AI in character formation encompass three fundamental dimensions: (1) the inability to experience and express authentic empathy. AI can simulate empathic responses but lacks the subjective experience that underlies true empathy; (2) limitations in understanding complex cultural and moral contexts. Moral values are contextual and ambiguous, requiring situational considerations that cannot be reduced to algorithmic patterns; (3) the absence of moral responsibility and accountability. AI lacks moral agency and cannot be held accountable in an ethical sense. These limitations are fundamental and cannot be overcome solely through technological advancement, as they stem from ontological differences between human consciousness and algorithmic computation.

The social function of schools, through the hidden curriculum, peer-to-peer interactions, and social validation, creates implicit value learning that cannot be replicated in digital platforms. The hidden curriculum transmits values such as discipline, cooperation, and social responsibility through organisational structures, rituals, and informal interactions that require physical presence and participation in real communities. Peer-to-peer interactions provide an authentic context for learning essential social skills such as negotiation, conflict resolution, and empathy, with real consequences that cannot be simulated. Social validation from legitimate institutions creates

a powerful incentive for the internalisation of values and the development of positive identities.

The formative function of teachers as role models, providers of humanistic touch, and facilitators of moral reflection provides essential emotional and moral scaffolding for the internalisation of values and the development of moral agency. Teachers' role models foster powerful observational learning, in which students internalise values through identification with respected authority figures. The teacher's humanistic touch and emotional responsiveness foster a deep pedagogical relationship in which students feel valued and supported as they explore their moral identity. Providing moral reflection helps students develop critical thinking skills about values and make autonomous moral decisions.

Successful approaches to education actually strengthen, rather than replace, the institutional role of schools. A complementarity model, in which AI personalises cognitive learning and automates administrative tasks. At the same time, teachers who focus on character building and facilitate social-emotional learning have proven most effective in achieving holistic educational goals. Countries such as Finland, Singapore, and Japan have adopted this principle by explicitly strengthening teachers' roles as mentors and facilitators, rather than simply as content transmitters.

Implications for Indonesia include three priority areas: (1) curriculum reorientation that balances digital competency with local value-based character building, with adequate time allocated for social interaction, collaborative learning, and moral reflection; (2) teacher professional development that prepares them as facilitators of holistic learning, with competencies in integrating pedagogical technology, facilitating moral discussions, and modeling values; (3) education policies that position AI as a complementary tool in a human-centered education ecosystem, with regulations that ensure AI platforms are pedagogically and ethically designed, and investment in equitable digital infrastructure.

The challenge going forward is to ensure that the adoption of technology in education does not compromise the social and formative functions of schools, which are essential for character building. This requires a collective commitment from education stakeholders—government, schools, teachers, parents, and the community—to maintain and strengthen the "Institutional

Iconic Presence" of schools as spaces for the formation of whole people. Schools must remain places where students not only acquire knowledge and skills but also develop character, identity, and the capacity to live meaningfully in diverse and complex communities. Technology must serve this humanistic vision, not replace it with paradigms of efficiency and personalisation that reduce education to mere information transmission.

REFERENCES

Afrison, D. (*2026*). *Strategi Pendidikan Islam Dalam Pembinaan Akhlak.* Hanif Media Pustaka.

Aisyah, S., & Rahayu, Y. M. (2026). *Pengembangan Soft Skills untuk Kesuksesan Akademik dan Profesional.* PT. Sonpedia Publishing Indonesia.

Andriyansah. (2023). *Distance Education Services With The Implementation Of Ergo Iconical Values* (pp. 28–37). STIPAS Publisher. https://publisher.stipas.ac.id/ index.php/pbs/issue/ view/3

Andriyansah, A., Sapriati, A., & Rulinawaty, R. (2026). Enhancing Research Competence in Distance Education through Scientific Workshops: Integrating Ergo-Iconic Learning Design. *Research Horizon*, *6*(1), 397–408. https://doi.org/https://doi.org/10.54518/rh.6.1.2026.996

Ardiansyah, R. (2023). Scrutinising the awareness of hidden curriculum: Connecting lecturers to themselves, students, and non-engineering study programs. *Jurnal Pendidikan Indonesia*, *12*(4). https:// doi.org /10.23887/jpiundiksha.v12i4.60678

Arfaiza, H. (2024). Keteladanan guru sebagai sarana penerapan pendidikan karakter siswa. *Journal Sinektik*, *7*(1). https://doi.org/10.33061/js.v7i1.9182

Bani, A. (2023). Peran guru dalam mewujudkan budaya 5S melalui penerapan hidden curriculum di SDN Lidah Wetan II. *Jurnal Jendela Pendidikan*, *3*(4). https://doi.org/10.57008/jjp.v3i04.582

Coulombe, S., Hardy, K., & Goldfarb, R. (2020). Promoting well-being through positive education: A critical review and proposed social ecological approach. *Theory and Research in Education*, *18*(3), 295–321.

Ghaffar, A. (2024). Teachers' role in cultivating moral responsibility among university students: Evidence from government institutions. *Journal of Educational Research*, *45*(3), 234–251.

Hanafiah, H. (2024). Character education's impact on student personality: A review of curriculum and school practices. *At-Tadib: Jurnal Ilmiah Prodi Pendidikan Agama Islam*, *19*(1). https://doi.org/10.21111/ attadib. v19i1.12047

Handayani, T. (2023). Pengaruh penanaman nilai-nilai karakter terhadap kepribadian peserta didik. *Jurnal Ilmu Kependidikan Dan Manajemen*

Anak Usia Dini, *1*(6). https://doi.org/10.54066/jikma.v1i6.1224

Indana, N. (2021). Kompetensi kepribadian guru dalam pembinaan akhlak siswa: Studi kasus di MTs Al-Ma'arif Brudu Sumobito Jombang. *Ilmuna: Jurnal Studi Pendidikan Agama Islam*, *3*(1). https://doi.org/10.54437/ ILMUNA .V3I1.250

Kasa, M. (2024). Exploring the interplay between moral character and prosocial behaviour among Chinese higher vocational college students: Insights into the role of emotional intelligence. *Journal of Moral Education*, *53*(2), 178–195.

Kholidin, A. (2025). Implementing character education in secondary schools: A case study of Muhammadiyah 4 Metro Utara, Indonesia. *Journal of Learning and Education*, *1*(3). https://doi.org/10.64780/jole.v1i3.95

Kotten, Y. (2025). Culturally embedded character education: Teacher identity and institutional transformation in Eastern Indonesia. *International Journal of Learning, Teaching and Educational Research*, *24*(11), 589–608. https://doi.org/10.26803/ijlter.24.11.33

Kubilay, S. (2025). Shaping personality in schools. In *Oxford Research Encyclopedia of Economics and Finance*. https://doi.org/10.1093/acrefore/ 9780190625979.013.936

Kurniawan, M. I. (2024). Identifying school programs and peer roles in juvenile delinquency through a mixed-method approach. *Jurnal Pendidikan Karakter*, *15*(2). https://doi.org/10.21831/jpka.v15i2.77518

Lickona, T. (2013). *Character education: The cultivation of virtue.* (In Instruc). Routledge.

Mamuaya, M. (2025). Building better behaviour: How school environments shape student character. *International Journal of Scientific Research and Management*, *13*(5). https://doi.org/10.18535/ijsrm/v13i05.el05

Napratilora, M. (2021). Peran guru sebagai teladan dalam implementasi nilai pendidikan karakter. *Al-Liqo: Jurnal Pendidikan Islam*, *6*(1). https://doi.org/10.46963/ALLIQO.V6I1.349

Nurhayati, S., & Andriyansah. (2025). *Mengenal konsep green oscillations untuk harmonisasi aktivitas ramah lingkungan*. Intilitera.

Nurhayati, S., & Andriyansah, A. (2024). Pengembangan Konsep Green Oscillations untuk Meningkatkan Produktivitas Kinerja Profesi Guru. *Jurnal Penelitian Inovatif*, *4*(2), 395–404. https://doi.org/https://doi.org/ 10.54082/jupin.326

Obmerga, K. P. (2022). Nurturing pagpapakatao and pagkakakilanlan among Filipino learners: Weaving Thomistic prudence on reflexive pedagogy and hidden curriculum. *Philippiniana Sacra*, *57*(173). https://doi.org/10.55997/2003pslvii173a3

Ortega, M. (2020). El docente, el currículo y la familia frente a la violencia.

Cedotic, *5*(1). https://doi.org/10.15648/CEDOTIC.1.2020.2518

Prihatina, I. (2022). Internalising the social care value of elementary school students through character education. *Qalamuna: Jurnal Pendidikan, Sosial Dan Agama*, *14*(2). https://doi.org/10.37680/qalamuna.v14i2.3417

Putri, A. (2024). Peran guru dalam membentuk karakter siswa. *Metta: Jurnal Ilmu Multidisiplin*, *4*(4). https://doi.org/10.37329/metta.v4i4.3617

Ramadhani, A. (2024). The role of character education in forming ethical and responsible students. *International Journal of Graduate of Islamic Education*, *5*(2). https://doi.org/10.37567/ijgie.v5i2.3064

Rambe, A. Y. F. (2020). Analysis of social competency character planting in the 2013 curriculum for elementary school students at SDN 060870 Medan Timur subdistrict in Medan city. *Teunuleh: The Educational Journal*, *1*(2). https://doi.org/10.51612/TEUNULEH.V1I2.18

Rao, M. S. (2024). Role of teachers in the character development of students: Findings from a systematic review. *Academy of Education and Social Sciences Review*, *4*(4). https://doi.org/10.48112/aessr.v4i4.935

Rudiyanto, R. (2024). Character education development in the education curriculum: Challenges and opportunities in the 21st century. *Jurnal Yudistira*, *2*(2). https://doi.org/10.61132/yudistira.v2i2.670

Sa'idah, N. (2024). Building student character through role models in the MABIT program at SDIT Al-Hikmah Banjarmasin. *Ta'Allum: Jurnal Pendidikan Islam*, *12*(2), 200–226. https://doi.org/10.21274/taalum.2024.12.2.200-226

Silalahi, D. E. (2025). The role of teachers in shaping students' character and competence towards a progressive Indonesia. *Acitya Wisesa*, *4*(3). https://doi.org/10.56943/jmr.v4i3.857

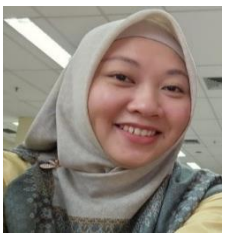
Syamsuddin, S. (2022). Peran guru pendidikan agama Islam dalam pendidikan karakter di sekolah. *Jurnal Azkia: Jurnal Aktualisasi Pendidikan Islam*, *17*(1). https://doi.org/10.58645/jurnalazkia.v17i1.174

Wasehudin, W. (2024). The paradigm of character education in Islamic elementary school. *Jurnal Ilmiah Islam Futura*, *24*(2). https://doi.org/10.22373/jiif.v24i2.22546

PROFIL PENULIS



Andriyansah Saat ini mengemban amanah sebagai dosen di Jurusan Manajemen, Fakultas Ekonomi, Universitas Terbuka. telah menghasilkan berbagai karya ilmiah yang terpublikasi di jurnal bereputasi serta buku-buku referensi Tulisan-tulisannya lahir dari pemikiran kritis dan pengalaman empiris. Pembaca yang ingin menelusuri jejak intelektual dan katalog karyanya lebih lanjut dapat mengunjungi laman resmi penulis di https://andri.staff.ut.ac.id/about-me/



Suci Nurhayati merupakan Dosen aktif pada Prodi Pendidikan Matematika Fakultas Keguruan dan Ilmu Pendidikan Universitas Terbuka, Tak menyurutkan semangat Suci untuk juga aktif melakukan riset dan beberapa diseminasi ilmiah di dalam dan luar negeri. karya lain penulis dapat mengunjungi blog penulis tersebut https://suci-nurhayati.staff.ut.ac.id/



Dwi Riyanti merupakan dosen tetap Fakultas Keguruan dan Ilmu Pendidikan Universitas Terbuka Pusat. Lulusan S1 dan S2 dari Universitas Negeri Yogyakarta Program Studi Pendidikan Pancasila dan Kewarganegaraan. Dwi juga pernah mengajar sebagai Dosen Luar Biasa di Universitas Negeri Yogyakarta dan Universitas Ahmad Dahlan. Saat ini Dwi juga sebagai editor Unnes Political Science Journal dan Editor Journal of Education and Technology, serta sebagai reviewer Jurnal Pancasila dan Bela Negara.

CHAPTER III

The Development of Critical Epistemic Attitudes and Reflective Technology Literacy as Strategic Competencies for Learners in the Era of Artificial Intelligence

Endang Tjahjaningsih[1)], Alya Takwarina Cahyani[2)]

[1]Department of Management, Stikubank University, Semarang
[2]Department of Business Administration, University of 17 Agustus 1945, Semarang,

Corresponding Author: naning@edu.unisbank.ac.id, Alyatac@untagsmg.ac.id

ABSTRACT
The rapid advancement of Artificial Intelligence (AI), particularly generative AI, has fundamentally transformed knowledge acquisition and learning processes in education. However, existing research in AI in Education (AIED) has predominantly focused on technological development, with limited attention paid to the epistemic implications of AI within learning contexts. This study aims to examine the relationship between AI literacy and epistemic cognition and to develop a conceptual model of AI epistemic competence as a strategic capability for learners in the AI era. A Systematic Literature Review (SLR) was conducted following PRISMA guidelines, analysing selected articles from Scopus, Web of Science, IEEE Xplore, and Google Scholar published between 2019 and 2024. The findings identify four dominant research themes: AIED, AI literacy, epistemic cognition, and generative AI in learning. The results reveal a critical gap between technological proficiency and the ability to critically evaluate AI-generated knowledge. This study proposes an integrative conceptual model comprising five dimensions: AI literacy, critical evaluation, epistemic cognition, reflective technology use, and ethical awareness. The model underscores the need to integrate technical understanding with critical and reflective thinking to support meaningful learning. This study contributes theoretically by bridging AI literacy and epistemic cognition, and pedagogically by offering a framework for developing learner competencies in AI-mediated learning environments.

Keywords: Artificial Intelligence in Education, AI Literacy, Epistemic Cognition, Generative AI, Critical Thinking, Systematic Literature Review.

INTRODUCTION

The development of Artificial Intelligence (AI) in recent years has brought significant changes across various sectors of life, including education. Technologies based on machine learning, deep learning, and large language models enable computer systems to process data at scale and generate information automatically. In the context of education, these technological advancements have transformed how learners access information, understand concepts, and construct knowledge through interaction with digital systems. Research on Artificial Intelligence in Education (AIED) has shown rapid growth. Scientific publications related to AI in education have experienced exponential growth since 2019, particularly following the emergence of generative AI technologies that are increasingly accessible to the public (Chen et al., 2022). This increase indicates that AI has become a central topic in global digital education transformation.

In higher education, AI technologies have been widely applied to support learning processes. AI applications in higher education are most commonly found in learning analytics, adaptive learning systems, and intelligent tutoring systems (Zawacki-Richter et al., 2019). These technologies enable learning systems to personalise content based on individual needs and provide faster feedback to learners.

The development of generative AI also significantly impacts learning processes. Generative AI technologies, such as large language models, can automatically generate text and conceptual explanations, thereby facilitating access to information (Dwivedi et al., 2023). However, these technologies also introduce new challenges related to information reliability and academic integrity, as generated content is not always based on verifiable scientific evidence. The ability to evaluate the validity of knowledge is closely related to the concept of epistemic cognition.

Epistemic cognition refers to how individuals understand the nature of knowledge and how knowledge is constructed and verified. Individuals with higher levels of epistemic cognition tend to be more critical in evaluating information obtained from various digital sources (Greene et al., 2021). Alongside the development of AI, the concept of AI literacy has emerged, emphasising the importance of understanding how AI technologies work and their social implications. AI literacy includes understanding fundamental AI

concepts, using AI effectively, and evaluating its ethical and social impacts (Ng et al., 2021).

Despite the rapid development of AI research in education, most studies still focus on technological development. Research specifically addressing the epistemic implications of AI use in learning remains limited. Therefore, a systematic literature review is needed to understand how AI influences knowledge construction in education. Based on this background, this study aims to conduct a Systematic Literature Review (SLR) on AI in education, AI literacy, and epistemic cognition. The study seeks to identify research trends, map research gaps, and develop a conceptual model of AI epistemic competence as a strategic capability for learners in the AI era.

RESEARCH METHODOLOGY

This study employs a Systematic Literature Review (SLR) approach to identify, evaluate, and synthesise research related to Artificial Intelligence in education, AI literacy, and epistemic cognition in learning. The systematic literature review approach was selected because it enables researchers to conduct a comprehensive, transparent, and replicable review of existing literature. Kitchenham and Charters (2007) explain that a systematic literature review is a research method that systematically collects and analyses relevant scientific evidence to address specific research questions.

The SLR approach is used to understand how the development of Artificial Intelligence technology influences knowledge construction in education and how technological literacy and critical epistemic attitudes can serve as strategic competencies for learners in the era of Artificial Intelligence. This approach allows researchers to identify research trends, map emerging research focuses, and uncover existing research gaps in the field of Artificial Intelligence in Education.

Literature Search Strategy

The literature search was conducted across several major scientific databases with broad international coverage, namely Scopus, Web of Science, IEEE Xplore, and Google Scholar. These databases were selected for their extensive use in educational and technological research and for their indexing of high-quality, reputable journals. The search was carried out using a combination of keywords relevant to the research topic. The keywords used in

this study include: Artificial Intelligence in Education, AI Literacy, Epistemic Cognition, Generative AI in Education, and Digital Knowledge Construction

The use of these keyword combinations aims to identify relevant literature while capturing diverse research perspectives on Artificial Intelligence in education. The publication period analysed in this study spans 2019 to 2024. This timeframe was selected to ensure that the analysed literature reflects the most recent developments in AI in education. Research publications in Artificial Intelligence in Education have increased significantly since 2019, particularly following advances in machine learning and generative AI (Chen et al., 2022).

Table 1. Systematic Literature Review Protocol

Component	**Description**
Database	Scopus, Web of Science, IEEE Xplore, Google Scholar
Publication period	2019–2024
Keywords	Artificial Intelligence in Education, AI Literacy, Epistemic Cognition
Initial articles identified	420 articles
Articles after screening	120 articles
Articles analyzed	52 articles

Sources: Kitchenham & Charters, 2007

Table 1 presents the protocol of the Systematic Literature Review (SLR) conducted in this study. The literature search was conducted across four major academic databases—Scopus, Web of Science, IEEE Xplore, and Google Scholar—to ensure broad coverage of high-quality, relevant publications. The review focused on studies published between 2019 and 2024, reflecting recent developments in Artificial Intelligence in Education. A combination of keywords, including *Artificial Intelligence in Education*, *AI Literacy*, and *Epistemic Cognition*, was used to retrieve relevant studies. The initial search resulted in 420 articles. After the screening process, which involved removing irrelevant and duplicate studies, the number was reduced to 120 articles. Finally, 52 articles that met all inclusion criteria were selected for in-depth analysis. These results indicate a rigorous, systematic filtering process that ensures only relevant, high-quality studies are included in the review.

Inclusion and Exclusion Criteria

To ensure the quality and relevance of the analysed literature, this study applies specific inclusion and exclusion criteria during the article selection process. These criteria aim to ensure that the selected literature is directly related to the research topic and meets adequate scientific standards. The inclusion criteria include articles published in reputable scientific journals, studies that discuss Artificial Intelligence within educational contexts, and studies that address technological literacy or epistemic cognition in learning. Additionally, the articles must be published within the specified timeframe of 2019 to 2024. Articles that are not directly related to Artificial Intelligence in education or do not address the research focus are excluded from the analysis. Furthermore, articles that are not available in full text are excluded, as they do not allow for in-depth analysis.

LITERATURE SELECTION PROCESS

The literature selection process follows the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines, which are widely used in systematic review research. Page et al. (2021) explain that the PRISMA approach enables a transparent and systematic selection process, thereby minimising bias in article selection.

The selection process consists of four main stages: identification, screening, eligibility, and inclusion. During the identification stage, relevant articles are identified through database searches using predefined keywords. This process generates an initial pool of literature. During the screening stage, articles are filtered based on titles and abstracts to ensure relevance to the research topic. Articles that are not directly related to Artificial Intelligence in education are excluded. During the eligibility stage, selected articles undergo full-text review to assess their alignment with the inclusion criteria. During the inclusion stage, articles that meet all criteria are included in the final analysis of the systematic literature review.

Figure 1 illustrates the literature selection process based on the PRISMA framework. The process consists of four main stages: identification, screening, eligibility, and inclusion. In the identification stage, 420 articles were retrieved from major academic databases, including Scopus, Web of Science, IEEE Xplore, and Google Scholar, using keywords related to Artificial Intelligence in education. During the screening stage, duplicate records were

removed, and an initial filter was applied to titles and abstracts, resulting in 350 relevant articles. In the eligibility stage, a full-text review was conducted to assess compliance with predefined inclusion criteria, resulting in 120 articles. Finally, in the inclusion stage, 52 articles that fully met the selection criteria were included in the systematic literature review. This process demonstrates a structured and transparent approach to literature selection, ensuring the reliability and validity of the reviewed studies.

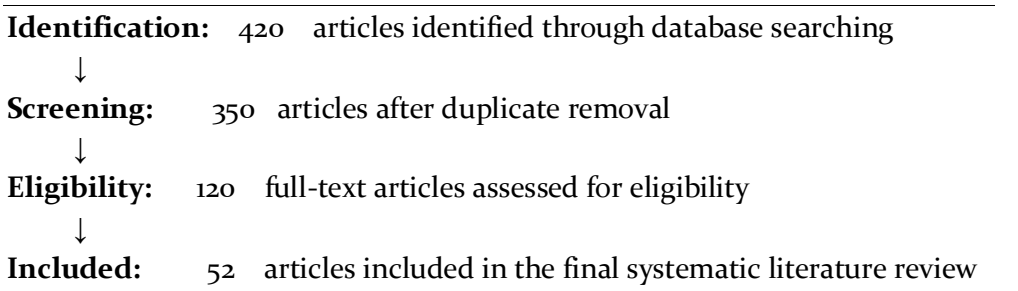


Figure 1. PRISMA Flow Diagram of the Literature Selection Process
Source: Adapted from the PRISMA 2020 framework (Page et al., 2021)

Bibliometric Analysis

In addition to the systematic literature review, this study employs bibliometric analysis to identify research trends in Artificial Intelligence in education. Bibliometric analysis examines publication patterns, thematic developments, and research focus within the literature.

Bibliometric analysis can be used to understand the development of a research field by examining publication volume, collaboration networks, and evolving research topics (Chen et al., 2022). In this study, bibliometric analysis is used to identify trends in Artificial Intelligence in Education research over the past five years. The results indicate a significant increase in research publications, particularly after 2022, driven by the rapid development and adoption of generative AI technologies in learning environments.

Data Analysis Method

The collected literature is analysed using a thematic synthesis approach. This approach is used to identify recurring themes in research related to Artificial Intelligence in education. The analysis process involves several key stages, including coding the research literature, grouping research themes, and synthesising the findings. The coding process involves identifying the research focus, methodology, and key findings for each analysed article.

Articles with similar research focuses are grouped into thematic categories. This process allows researchers to identify patterns within the literature and uncover existing research gaps. Through thematic synthesis, this study produces a comprehensive review of the literature on Artificial Intelligence in education. It provides a conceptual foundation for developing an AI epistemic competence model for learning.

Literature Review Findings

Research on Artificial Intelligence in education has experienced significant growth over the past five years. This development aligns with the increasing use of digital technologies in learning processes and the broader adoption of AI across various sectors. In the educational context, Artificial Intelligence is no longer viewed merely as a supporting technological tool, but as part of a knowledge ecosystem that influences how individuals acquire, process, and evaluate information.

Analysis of literature from the Scopus and Web of Science databases indicates that most research in Artificial Intelligence in Education (AIED) focuses on developing AI-based learning systems to improve learning effectiveness. These studies in higher education primarily highlight the use of learning analytics, intelligent tutoring systems, and adaptive learning environments that can tailor learning processes to individual learner needs (Zawacki-Richter et al., 2019). Through these technologies, learning systems can provide automated feedback and help educators monitor student learning progress more accurately. Other studies indicate that AI use in learning can also enhance student engagement. Hwang et al. (2020) explain that AI-based adaptive learning systems can adjust the difficulty of learning materials, allowing learners to study at their own pace. This enables a more personalised learning experience than conventional instructional methods.

In recent years, the focus of research on Artificial Intelligence in education has shifted from technological development toward examining the pedagogical implications of these technologies. The advancement of digital technologies in education not only affects instructional methods but also transforms educators' roles (Bond et al., 2021). In digital learning environments, educators are no longer merely information providers but facilitators who guide students in understanding and critically evaluating information obtained from various digital sources.

The development of generative AI technologies has also introduced significant changes to the digital learning ecosystem. Generative AI technologies, such as large language models, can produce academic text, summarise information, and explain concepts automatically (Dwivedi et al., 2023). These technologies create new learning opportunities by enabling faster, easier access to information. However, they also introduce new challenges related to the reliability of AI-generated information. The use of large language models in education can provide substantial benefits when applied as learning support tools (Kasneci et al., 2023). These technologies can help students understand complex concepts and provide additional explanations of learning materials. However, their use requires appropriate pedagogical supervision to ensure that they do not reduce learners' critical thinking abilities.

Beyond pedagogical issues, the literature also highlights concerns related to academic integrity. The use of generative AI in academic writing can increase the risk of plagiarism and complicate the evaluation of student learning outcomes (Cotton et al., 2023). Therefore, educational institutions need to establish clear policies regarding the use of AI technologies in learning processes. In the context of knowledge evaluation, students' ability to assess information obtained from various digital sources becomes increasingly important. The ability to critically evaluate knowledge is closely related to epistemic cognition, which refers to how individuals understand the nature of knowledge and how knowledge is constructed and verified (Greene et al., 2021). Individuals with higher levels of epistemic cognition tend to be more critical in evaluating knowledge claims and are better able to distinguish between credible and less credible information.

Alongside the development of AI technologies, the concept of AI literacy has emerged, emphasising the importance of understanding how AI systems function and their social implications. AI literacy comprises three main dimensions: understanding fundamental AI concepts (Ng et al., 2021), using AI technologies effectively, and evaluating their social and ethical impacts. These findings indicate that the use of AI in education requires not only technical skills in utilising technology, but also cognitive abilities that enable individuals to critically evaluate knowledge. Therefore, integrating

reflective technological literacy and critical epistemic attitudes is essential for developing learner competencies in the era of Artificial Intelligence.

Table 2. Summary of AI Research in Education

Author & Year	Method	Research Focus	Key Findings
Zawacki-Richter et al. (2019).	SLR	AI in higher education	AI is widely used in intelligent tutoring and learning analytics
Ng et al. (2021).	Conceptual	AI literacy	AI literacy is essential in education
Greene et al. (2021).	Review	Epistemic cognition	Epistemic beliefs influence information evaluation
Dwivedi et (2023)al.	Review	Generative AI	Generative AI transforms knowledge production
Kasneci et al. (2023).	Empirical	ChatGPT in education	LLMs can function as digital tutors

Sources: Zawacki-Richter et al. (2019), Ng et al. (2021), Greene et al. Dwivedi et al.(92021) and Kasneci (2023)

Table 2 provides a summary of key studies on Artificial Intelligence in Education, highlighting the diversity of research approaches, focal areas, and main findings. The table includes studies employing various methodologies, such as systematic literature reviews, conceptual analyses, review studies, and empirical research. The findings indicate that AI in education has been explored from multiple perspectives. Early research, such as that of Zawacki-Richter et al. (2019), focuses on the application of AI technologies in higher education, particularly intelligent tutoring systems and learning analytics.

Subsequent studies expand the scope by emphasising the importance of AI literacy as a critical competency in education (Ng et al., 2021). At the same time, research on epistemic cognition (Greene et al., 2021) highlights the role of epistemic beliefs in shaping how individuals evaluate information. More recent studies, such as Dwivedi et al. (2023), examine the impact of generative AI on knowledge production, while empirical research by Kasneci et al. (2023) demonstrates the potential of large language models to function as digital tutors in educational settings. Overall, the table reflects the evolution of AI research in education from technology-oriented applications to broader considerations of cognitive, epistemic, and pedagogical implications.

Thematic Analysis of SLR Findings

Research on Artificial Intelligence in education demonstrates a clear pattern of development over the past five years. The analysis of the reviewed literature indicates that studies in Artificial Intelligence in Education (AIED) do not evolve randomly, but rather form several interconnected core themes. These themes reflect a shift in research focus from purely technical aspects toward pedagogical and epistemic dimensions in the use of AI technologies.

The synthesis of the literature reveals that research on AI in education can be categorised into four main themes: Artificial Intelligence in Education (AIED), AI literacy, epistemic cognition, and generative AI in learning. These themes are not isolated; instead, they are interrelated and collectively form a complex digital knowledge ecosystem.

The first and most dominant theme in the literature is Artificial Intelligence in Education (AIED), which focuses on the development and implementation of AI technologies in learning processes. Early research in this field primarily emphasised the development of AI-based learning systems, such as intelligent tutoring systems and learning analytics (Zawacki-Richter et al., 2019). These systems are designed to improve learning effectiveness through personalized content delivery and automated analysis of learning data.

Research in this theme demonstrates that AI technologies have significant potential to enhance learning quality through adaptive, data-driven approaches. However, research within the AIED domain tends to emphasise technical aspects rather than the cognitive and epistemic dimensions of learning. Many studies focus on how technology improves learning performance, while relatively few examine how these technologies influence learners' understanding and evaluation of knowledge. This indicates a research gap that requires further investigation.

The second theme identified in the literature is AI literacy, which emphasises the importance of individuals' ability to understand and use AI technologies critically. AI literacy extends beyond technical skills, encompassing an understanding of how AI systems operate as well as their social and ethical implications (Ng et al., 2021). In educational contexts, AI literacy is a crucial competency that ensures learners are not merely passive users of technology but can also understand how such technologies influence

learning processes. Despite the growing discussion of AI literacy in the literature, most studies remain conceptual, with limited empirical research examining its implementation within educational curricula. This indicates an opportunity for future research to develop learning models that systematically integrate AI literacy into educational practices.

The third theme that has gained increasing attention is epistemic cognition, which relates to how individuals understand the nature of knowledge and how it is constructed and validated. The epistemic cognition plays a critical role in learning, as it influences how individuals evaluate information and determine the credibility of knowledge claims. In digital learning environments, this ability becomes increasingly important as learners encounter diverse sources of information with varying levels of credibility (Greene et al., 2021). Research on epistemic cognition shows that individuals with more advanced epistemic understanding tend to evaluate information more critically and are better able to distinguish between credible information and unsupported claims. However, studies explicitly linking epistemic cognition to the use of AI technologies in learning remain limited. Most existing research examines epistemic cognition within traditional or general digital learning contexts, without specifically addressing AI-mediated learning environments.

The fourth theme that has rapidly developed in recent years is generative AI in learning. The generative AI technologies, such as large language models, have transformed how individuals access information and produce knowledge (Dwivedi et al., 2023). These technologies enable users to obtain answers instantly, thereby altering patterns of interaction with information. The use of generative AI in education can provide significant benefits when applied as a supportive learning tool. However, these technologies also introduce epistemic risks, such as over-reliance on AI systems and the potential dissemination of inaccurate information (Kasneci et al., 2023). Further highlights that the use of generative AI may affect academic integrity, as it allows students to generate academic texts automatically without engaging in deep cognitive processes (Cotton et al., 2023).

Based on these four themes, it can be concluded that research on Artificial Intelligence in education has shifted from a focus on technological development toward more complex inquiries into the implications of

technology use on learning processes and knowledge construction. This shift indicates that the use of AI in education cannot be separated from cognitive and epistemic dimensions that influence how individuals understand knowledge.

Table 3. Synthesis of Research Themes in AI in Education

Theme	Research Focus	Research Gap
Artificial Intelligence in Education	Development of AI-based learning systems	Limited exploration of epistemic aspects
AI Literacy	Understanding and use of AI technologies	Limited implementation in educational curricula
Epistemic Cognition	Evaluation of knowledge	Not yet explicitly linked to AI contexts
Generative AI	AI in knowledge production	Epistemic risks remain underexplored

Sources: Leung and Qiao (2021), Kasneci (2023), Kidd and A. Birhane (2023)

Table 3 presents a thematic synthesis of research on Artificial Intelligence in Education, identifying four major research themes, their corresponding research focus, and existing gaps. The first theme, Artificial Intelligence in Education (AIED), primarily focuses on the development of AI-based learning systems. However, the literature reveals a limited exploration of epistemic aspects, particularly regarding how AI influences knowledge construction and evaluation. The second theme, AI literacy, emphasizes understanding and using AI technologies. Despite its importance, the implementation of AI literacy within educational curricula remains limited, indicating a gap between conceptual development and practical application. The third theme, epistemic cognition, addresses how individuals evaluate knowledge and determine its validity. While this theme highlights the importance of critical evaluation, it has not yet been explicitly linked to the use of AI technologies in learning contexts.

The fourth theme, generative AI, focuses on AI's role in knowledge production. Although this area has grown rapidly, epistemic risks such as misinformation, over-reliance on AI, and reduced critical thinking remain underexplored. The table demonstrates a clear shift in research from technical development toward more complex epistemic and pedagogical considerations, while also highlighting significant gaps that require further investigation.

Research Gap

Based on the thematic analysis, several research gaps emerge. First, most studies on Artificial Intelligence in education still focus on the technical aspects of technology development, while the epistemic implications of AI use remain underexplored. Second, although the concept of AI literacy has been widely discussed, its implementation in learning contexts has not been empirically examined to the same extent. Third, research on epistemic cognition is rarely directly linked to the use of Artificial Intelligence in learning, despite its critical importance in navigating complex digital information environments. Fourth, the use of generative AI in learning introduces new challenges related to information reliability and academic integrity; however, research on these issues is still in its early stages and requires further exploration. Based on these identified gaps, this study proposes integrating reflective technology literacy and critical epistemic attitudes as a foundation for developing learner competencies in the era of Artificial Intelligence.

Discussion and Findings

The advancement of Artificial Intelligence in education has not only introduced changes in the technical aspects of learning but has also fundamentally transformed how knowledge is produced, disseminated, and evaluated. This transformation marks a paradigm shift from information-based learning to knowledge–evaluation–based learning. In this context, technology no longer functions merely as a supporting tool but also as an epistemic agent that actively participates in knowledge construction.

Early research on Artificial Intelligence in education primarily focused on the development of adaptive learning systems to improve learning effectiveness (Zawacki-Richter et al., 2019). These systems enable personalised learning tailored to individual needs, thereby enhancing learning efficiency. However, this approach still positions technology as an instrumental tool, without fully considering its broader implications on how learners understand knowledge.

With the emergence of generative AI, the role of technology in learning has become more complex. Generative AI does not merely present information; it also generates new knowledge based on learned data patterns (Dwivedi et al., 2023). This introduces significant epistemic implications, as

the boundary between "human-constructed knowledge" and "machine-generated knowledge" becomes increasingly blurred. Learners are no longer interacting solely with information, but also with algorithmically generated representations of knowledge. The use of large language models in education can help students understand complex concepts, but may also reduce cognitive engagement if used without critical reflection (Kasneci et al., 2023). These technologies enable students to obtain answers instantly, potentially shifting learning processes from exploratory activities to more passive, consumptive behaviours.

These findings indicate that the development of Artificial Intelligence in education produces two opposing implications. On the one hand, AI technologies enhance access to knowledge and support more personalised, adaptive learning processes. On the other hand, they may undermine learning quality if not accompanied by critical and reflective thinking skills. The concept of epistemic cognition becomes increasingly relevant. Epistemic cognition refers to how individuals understand the nature of knowledge and its verification (Greene et al., 2021). Individuals with higher levels of epistemic cognition tend not to accept information passively; instead, they critically evaluate sources, evidence, and the logic underlying knowledge claims.

The literature review indicates that most research on epistemic cognition is still conducted within traditional or general digital learning contexts. Studies specifically examining the relationship between epistemic cognition and the use of AI technologies in learning remain limited. This highlights a significant research gap, particularly in understanding how learners evaluate information generated by Artificial Intelligence systems.

The concept of AI literacy has emerged in response to the increasing use of AI technologies in everyday life. AI literacy includes the ability to understand how AI systems work, use these technologies effectively, and evaluate their social and ethical implications (Ng et al., 2021). In educational contexts, AI literacy is an essential competency that ensures learners are not only capable of using technology but also understand its broader implications.

The literature indicates that current approaches to AI literacy tend to focus on technical and ethical aspects, while the epistemic dimension remains underexplored. Most studies discuss how students understand AI systems or how these technologies impact society, but relatively few examine how

learners evaluate the validity of AI-generated information. These findings reveal a conceptual gap between technological literacy and epistemic capability in learning. AI literacy equips individuals to use technology, but does not necessarily ensure their capacity to evaluate knowledge generated by such technologies critically. Conversely, epistemic cognition provides evaluative capabilities, yet has not been sufficiently integrated into the context of AI-mediated learning.

Conceptual Model of AI Epistemic Competence

This study proposes a conceptual model that integrates two domains that have traditionally developed separately in the literature, namely Artificial Intelligence literacy (AI literacy) and epistemic cognition. The integration of these two domains is essential, as advances in Artificial Intelligence have fundamentally transformed how knowledge is produced, accessed, and validated in learning environments. The literature indicates that AI literacy provides a crucial foundation for individuals to understand how technology operates and its implications. AI literacy encompasses understanding fundamental AI concepts, using technology effectively, and awareness of the ethical and social implications of its use (Ng et al., 2021). However, this form of AI literacy remains primarily focused on the use and understanding of technology, without explicitly developing evaluative capabilities regarding the validity of knowledge generated by AI systems.

The concept of epistemic cognition provides a framework for understanding how individuals evaluate knowledge. Epistemic cognition includes individuals' beliefs about the nature of knowledge, its sources, and its verification. Individuals with more advanced epistemic cognition tend to be more critical in evaluating information and are less likely to accept knowledge claims without sufficient evidence (Greene et al., 2021). The literature review finds that these two concepts are rarely integrated in the context of Artificial Intelligence use in education. Most studies on AI literacy do not address epistemic aspects, while research on epistemic cognition has not specifically examined interactions between individuals and AI systems. This gap highlights the need for a conceptual framework that integrates both domains. This study proposes the concept of AI epistemic competence, defined as the ability of learners to understand, evaluate, and use Artificial Intelligence technologies critically, reflectively, and responsibly in learning processes. This

competence extends beyond technical skills, encompassing cognitive and reflective abilities to evaluate knowledge generated by AI systems.

Structure of the Conceptual Model

The proposed conceptual model consists of five interrelated dimensions:

1. AI Literacy (Understanding AI Technologies)
2. Critical Evaluation of Information
3. Epistemic Cognition
4. Reflective Technology Use
5. Ethical and Responsible AI Use

These five dimensions form an integrated framework that illustrates how learners interact with AI technologies in learning processes.

Table 4. Dimensions of AI Epistemic Competence

Dimension	Description	Theoretical Basis
AI Literacy	Ability to understand how AI works and how to use it	Leung and Qiao (2021)
Critical Evaluation	Ability to evaluate information generated by AI	Greene and Torney-Purta (2023)
Epistemic Cognition	Understanding the nature and validity of knowledge	Greene and Torney-Purta (2023)
Reflective Use	Ability to reflect on the use of technology	Selwyn (2021)
Ethical Awareness	Awareness of ethical aspects in AI use	Floridi (2022)

Sources: Leung and Qiao (2021), Leung and Qiao (2021), Selwyn (2021), Floridi (2022)

Table 4 presents the dimensions of AI epistemic competence, outlining the core components learners need to interact effectively and critically with Artificial Intelligence technologies. The model consists of five interrelated dimensions. The first dimension, AI literacy, refers to the ability to understand how AI systems function and how they can be effectively utilised in learning contexts. The second dimension, critical evaluation, emphasises the ability to assess the accuracy, credibility, and reliability of information generated by AI systems.

The third dimension, epistemic cognition, focuses on learners' understanding of the nature of knowledge and how it is constructed and validated. This dimension underpins the development of critical judgment in evaluating knowledge claims. The fourth dimension, reflective use, highlights

the importance of reflecting on how technology is used and how it influences learning processes. The final dimension, ethical awareness, addresses the ability to consider ethical issues such as fairness, transparency, and accountability in the use of AI technologies. Overall, the table illustrates an integrated framework that combines technical understanding, cognitive evaluation, reflective awareness, and ethical responsibility as essential components of AI epistemic competence.

Relationships Between Dimensions

The model suggests that AI literacy serves as the foundational layer that enables learners to understand how technology operates. However, technical understanding alone is insufficient to ensure meaningful learning. Therefore, this capability must be followed by critical evaluation, which refers to the ability to assess information generated by AI systems. This evaluative ability is further strengthened by epistemic cognition, which enables learners to understand that knowledge is not absolute but is constructed through processes of verification and argumentation. Individuals with more developed epistemic cognition are more critical in evaluating knowledge claims (Greene et al., 2021).

The model also highlights the importance of reflective technology use, defined as the ability to critically reflect on how technology is used in learning processes. The use of technology in education must be accompanied by critical awareness of its impact on learning (Selwyn, 2021). Ultimately, these processes culminate in ethical and responsible AI use, which involves the ability to use technology responsibly. The use of AI technologies must consider ethical principles such as fairness, transparency, and accountability (Floridi et al., 2022).

Integration of the Model in Learning Contexts

The conceptual model of AI epistemic competence demonstrates that the development of learner competencies in the era of Artificial Intelligence cannot rely solely on technological mastery. Education must adopt approaches that integrate technological literacy with critical and reflective thinking skills. In practice, this integration can be implemented through various strategies, including: 1) Critical analysis of AI-generated outputs, 2) Verification of digital information sources, 3). Reflective discussions on the use of technology, 4) Evaluation of bias in AI systems. The use of AI in learning can provide

significant benefits when supported by pedagogical strategies that promote cognitive engagement. Without such approaches, AI technologies risk reducing the quality of learning processes (Kasneci et al., 2023).

Conceptual Contribution

The proposed conceptual model offers several important contributions to the educational literature. First, it integrates two domains that have largely been studied separately, namely AI literacy and epistemic cognition. Second, it provides a conceptual framework for understanding how learners interact with AI technologies during learning. Third, it provides a foundation for developing curricula and instructional strategies to enhance learners' epistemic competence in the era of Artificial Intelligence. Thus, the proposed model contributes not only at the theoretical level but also provides practical implications for the development of education in the digital age.

Pedagogical Implications for Education in the AI Era

The advancement of Artificial Intelligence in education requires not only changes in technology use but also a fundamental transformation in pedagogical approaches. The results of the systematic literature review indicate that AI technologies have reshaped how learners acquire and process information, requiring educational approaches to adapt to the characteristics of a new digital knowledge ecosystem. Pedagogical approaches can no longer focus solely on knowledge transmission but must shift toward developing critical, reflective, and evidence-based thinking skills. In the digital era, the role of education has shifted from delivering information to facilitating learning processes that enable learners to understand how knowledge is constructed and verified. Therefore, integrating reflective technological literacy and critical epistemic attitudes becomes a key element in instructional design (Selwyn, 2021).

Table 5. Implications for Learner Competency Development in the AI Era

Competency Aspect	Condition Without Epistemic and AI Literacy	Expected Competency	Implications for Learners
Information Access	Learners accept information from AI or the internet	Learners can evaluate the credibility of	Learners become more selective in accepting digital information

Competency Aspect	Condition Without Epistemic and AI Literacy	Expected Competency	Implications for Learners
	without verifying sources.	information sources	
AI Usage	Using AI only to obtain instant answers	Using AI as a tool for knowledge exploration	AI is utilised to deepen conceptual understanding
Information Analysis	Information is passively accepted without critical analysis	Learners can compare multiple information sources	Critical and analytical thinking skills improve
Knowledge Verification	No re-checking of information	Verifying information through multiple references	Reduces the risk of misinformation in learning
Decision Making	Decisions are made based on unverified information	Decisions are based on evidence and information analysis	Learners become more rational in problem-solving
Awareness of Technological Bias	Lack of understanding of algorithmic bias in AI systems	Understanding the limitations and biases of AI technologies	Learners become more critical of technological outputs
Reflective Learning	Learning focuses only on outcomes	Learners reflect on thinking processes and technology use	Learning becomes deeper and more meaningful
Ethical Use of AI	Risk of plagiarism or unethical use of AI	Using AI responsibly and transparently	Learners' academic integrity improves

Sources: Developed by the authors based on a synthesis of Leung and Qiao (2021), Leung and Qiao (2021), Selwyn (2021), Floridi (2022), Luckin et al (2016), Holmes and Tuomi (2022), Kidd and A. Birhane (2023)

Table 5 illustrates the implications of learner competency development in the era of Artificial Intelligence by contrasting conditions without epistemic and AI literacy with the expected competencies and their corresponding impacts. The table highlights a clear transformation in how

learners interact with information and technology. Without epistemic and AI literacy, learners tend to accept information passively, rely on AI for instant answers, and make decisions based on unverified data. This condition reflects a surface-level engagement with knowledge, where technology is used primarily for efficiency rather than understanding.

In contrast, the expected competencies emphasise active, critical, and reflective engagement with information. Learners are expected to evaluate the credibility of sources, compare multiple perspectives, verify information through reliable references, and use AI as a tool for exploration rather than mere answer generation. This shift indicates a transition from passive consumption to active knowledge construction. The table also underscores the importance of higher-order cognitive skills, such as critical thinking, analytical reasoning, and evidence-based decision-making. Additionally, it highlights the need for awareness of algorithmic bias and the limitations of AI systems, which are essential for developing informed and critical users of technology.

Furthermore, the inclusion of reflective learning and ethical AI use demonstrates that competency development in the AI era extends beyond cognitive abilities to include metacognitive and ethical dimensions. Learners are expected to reflect on their thinking processes, use technology responsibly, and maintain academic integrity. Overall, the table shows that developing learner competencies in the AI era requires a comprehensive integration of technical, cognitive, reflective, and ethical capabilities, ensuring that learners can engage with AI technologies in a meaningful, critical, and responsible manner.

Integration in Curriculum and Instructional Design

The literature review indicates that one of the main implications of AI development in education is the need to integrate AI literacy and epistemic competence into curricula. That education in the AI era must develop curricula that not only teach technical skills but also foster critical thinking toward digital information (UNESCO, 2023). This integration can be implemented by incorporating topics such as how AI systems work, potential algorithmic biases, and ethical implications of technology use in learning. AI literacy should include understanding fundamental concepts, effective use of technology, and the ability to evaluate its societal impact (Ng et al., 2021). The

primary challenge in implementing such curricula lies in balancing technical and epistemic aspects. A curriculum focused solely on technological mastery risks producing learners who can use technology but are unable to critically evaluate AI-generated knowledge. Conversely, a curriculum that emphasises critical thinking without technological understanding may hinder learners' ability to interact effectively with digital systems.

Reflective and Knowledge-Verification-Based Learning Strategies

Beyond curriculum changes, the advancement of AI also requires a transformation in instructional strategies. Learning in the AI era must be designed to encourage learners to reflect on their learning processes and verify the information they encounter. The ability to evaluate information is a core component of epistemic cognition and directly influences learning quality. Therefore, instructional strategies should encourage students to question information sources, evaluate evidence, and compare multiple perspectives before concluding (Greene et al., 2021).

Approaches such as inquiry-based learning and problem-based learning are particularly relevant, as they promote active engagement in information exploration and evaluation. AI technologies can enhance learning effectiveness when combined with pedagogical strategies that promote cognitive engagement (Kasneci et al., 2023). In practice, educators can use AI as a tool to stimulate critical discussion rather than as a final source of answers. For example, AI-generated outputs can be analysed to evaluate accuracy, identify biases, and be compared with more credible sources.

Development of Process-Oriented Assessment

The advancement of AI also necessitates changes in assessment systems. Traditional evaluation methods that focus on final outputs are no longer sufficient to measure learners' critical and reflective thinking abilities, especially in AI-supported learning contexts. The generative AI introduces new challenges in academic assessment, as students can produce academic texts automatically. In such cases, assessments focusing solely on final products become less relevant, as they cannot distinguish between student work and AI-generated outputs (Cotton et al., 2023). Assessment must therefore be designed to evaluate learners' thinking processes, including their ability to assess information, justify decisions, and reflect on their learning. Examples include reflective portfolios, research-based projects, and analytical

discussions that require students to demonstrate their reasoning processes. Such approaches not only improve evaluation quality but also foster critical and reflective thinking, which are essential components of AI epistemic competence.

Transformation of the Teacher's Role as an Epistemic Facilitator

In the era of Artificial Intelligence, the role of teachers undergoes a significant transformation. Teachers are no longer the sole source of knowledge but act as facilitators who guide learners in understanding how knowledge is constructed and verified in digital environments. Educators must develop pedagogical approaches that help students understand the relationships among technology, information, and knowledge. In this context, teachers serve as epistemic facilitators, supporting learners in evaluating information and developing critical thinking skills.

Teachers must possess adequate technological literacy to understand both the potential and the limitations of AI in learning. The importance of developing educators' digital competencies to ensure effective integration of technology into education (UNESCO, 2023). This transformation highlights that educational change in the AI era involves not only learners but also educators, who must continuously adapt their approaches and competencies.

IMPLICATIONS FOR LEARNER COMPETENCY DEVELOPMENT

The development of Artificial Intelligence in education not only affects learning methods but also reshapes the competency profile required of learners. In an AI-driven knowledge ecosystem, learners are no longer required merely to acquire information but must also be able to evaluate, verify, and use knowledge critically. This transformation suggests that 21st-century competencies must be redefined to incorporate AI's role in learning processes. While traditional competencies include critical thinking, problem-solving, collaboration, and digital literacy, AI introduces a new dimension: the ability to interact critically and reflectively with intelligent systems.

21st-Century Competencies in the Context of AI

In educational literature, 21st-century competencies are often associated with critical thinking, creativity, communication, and collaboration. However, the advancement of AI requires an expanded definition. In the digital era, competencies extend beyond cognitive abilities

to include understanding how technology influences knowledge production and distribution. AI literacy is a crucial component of digital competence, enabling individuals to interact with AI technologies effectively and responsibly.

Strengthening Critical and Reflective Thinking Skills

The literature indicates that critical and reflective thinking skills are increasingly important in the AI era. Individuals with higher epistemic cognition are better able to evaluate information critically and understand that knowledge is dynamic and contextual (Greene et al., 2021). In AI-supported learning environments, learners are exposed to algorithmically generated information. This requires them to actively evaluate the accuracy, relevance, and credibility of such information rather than passively accepting it. Reflective thinking is equally essential, as it allows learners to understand how they use technology and how it shapes their knowledge construction processes. That critical reflection on technology use is a key component of broader digital literacy.

Epistemic Resilience to Digital Misinformation

One of the major challenges in the AI era is the increasing risk of misinformation generated by AI systems. Generative AI can produce highly coherent information at scale, but it does not guarantee factual accuracy. AI systems generate outputs based on data patterns rather than grounded truth (Dwivedi et al., 2023). In this context, learners must develop epistemic vigilance, defined as the ability to evaluate the reliability of information. The epistemic vigilance is a cognitive mechanism that enables individuals to assess whether information can be trusted.

This ability is crucial in digital learning environments where information sources vary in credibility. Without it, learners risk becoming passive consumers of misleading information. The literature also highlights the impact of AI on academic integrity. The generative AI may reduce cognitive engagement and increase plagiarism risks (Cotton et al., 2023). Therefore, learner competencies must include responsible technology use and awareness of its limitations in academic contexts.

Table 6 outlines the key competency implications for learners in the era of Artificial Intelligence, highlighting the essential capabilities required to navigate AI-driven learning environments. The table identifies five core

competency dimensions. The first, AI literacy, emphasises learners' ability to understand and effectively use AI technologies. The second, critical thinking, focuses on evaluating information and avoiding passive acceptance of AI-generated content. The third dimension, epistemic cognition, concerns understanding the nature of knowledge and distinguishing between valid and unreliable information.

Table 6. Learner Competency Implications in the AI Era

Competency Dimension	Description	Implication
AI Literacy	Understanding AI technologies	Ability to use technology effectively
Critical Thinking	Evaluating information	Avoid passive information acceptance
Epistemic Cognition	Understanding knowledge	Ability to distinguish valid information
Reflective Thinking	Reflecting on technology use	Conscious use of technology
Epistemic Vigilance	Resistance to misinformation	Reduced susceptibility to false information

Source: Leung and Qiao (2021), Leung and Qiao (2021), Hofer and Pintrich (1997), Kuhn (1991), Holmes, M. Bialik, and Fadel (2019), Selwyn (2021), Floridi (2022), Cotton, and J. Shipway (2023)

The fourth dimension, reflective thinking, underscores the importance of conscious, critical reflection on how technology is used in learning processes. The final dimension, epistemic vigilance, refers to learners' ability to resist misinformation by critically assessing the credibility of information sources, particularly those generated by AI systems. Overall, the table demonstrates that learner competencies in the AI era extend beyond technical proficiency, requiring a combination of cognitive, reflective, and evaluative abilities to ensure meaningful and responsible engagement with AI technologies.

Future Research Agenda

Research on Artificial Intelligence in education remains dynamic and offers extensive opportunities for further exploration. Although the number of studies in Artificial Intelligence in Education has increased significantly, research still predominantly focuses on technological development, while

epistemic aspects remain underexplored. A key priority for future research is the empirical validation of AI epistemic competence models. Most existing studies, including this one, remain conceptual. Future research should examine the relationship between AI literacy, epistemic cognition, and learning outcomes. Greene et al. (2021) highlight the importance of epistemic cognition in information evaluation, yet its role in AI-mediated learning remains insufficiently studied.

Future research should also explore how generative AI can be effectively integrated into learning without reducing cognitive engagement. Generative AI enhances access to information, but it may also diminish critical thinking if misused. Another important direction is the development of instruments to measure learners' epistemic competence in AI contexts. Existing instruments are largely designed for traditional learning environments and may not adequately capture AI-related evaluation skills (Dwivedi et al., 2023). Research on ethical implications must also be expanded. AI use must consider fairness, transparency, and accountability, which are particularly relevant in educational contexts (Floridi et al., 2022).

Interdisciplinary research is also essential, as AI in education intersects with fields such as psychology, pedagogy, and social sciences. Future studies should also examine contextual differences between developed and developing countries, as Selwyn (2021) notes that social, economic, and cultural factors influence technology adoption. Finally, longitudinal studies are needed to assess the long-term impact of AI on learners' critical and reflective thinking abilities, as most current research focuses on short-term effects. In conclusion, integrating reflective technological literacy and critical epistemic attitudes represents a promising direction for advancing education in the AI era.

REFERENCES

Bond, M., Buntins, K., Bedenlier, S., Zawacki-Richter, O., & Kerres, M. (2021). Mapping research in student engagement and educational technology in higher education: A systematic evidence map. *International Journal of Educational Technology in Higher Education, 18*. https://doi.org/10.1186/s41239-021-00235-6

Chen, L., Chen, P., & Lin, Z. (2022). Artificial intelligence in education: A review. *IEEE Access, 10*, 75264–75278. https://doi.org/10.1109/ACCESS.2022.3196219

Cotton, D. R. E., Cotton, P. A., & Shipway, J. R. (2023). Chatting and cheating: Ensuring academic integrity in the era of ChatGPT. *Innovations in Education and Teaching International*. https://doi.org/10.1080/14703297.2023.2190148

Dwivedi, Y. K., et al. (2023). So what if ChatGPT wrote it? Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI. *International Journal of Information Management, 71*, 102642. https://doi.org/10.1016/j.ijinfomgt.2023.102642

Floridi, L., et al. (2022). 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

Greene, J. A., Azevedo, R., & Torney-Purta, J. (2021). Epistemic cognition and its role in learning and instruction. *Educational Psychologist, 56*(3), 1–17. https://doi.org/10.1080/00461520.2021.1903827

Hofer, B. K., & Pintrich, P. R. (1997). The development of epistemological theories: Beliefs about knowledge and knowing. *Review of Educational Research, 67*(1), 88–140.

Holmes, K., & Tuomi, A. (2022). State of the art in AI and education. *Computers & Education*. https://doi.org/10.1016/j.compedu.2022.104553

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

Hwang, G.-J., Xie, H., Wah, B. W., & Gašević, D. (2020). Vision, challenges, roles and research issues of artificial intelligence in education. *Computers and Education: Artificial Intelligence, 1*, 100001. https://doi.org/10.1016/j.caeai.2020.100001

Kidd, C., & Birhane, A. (2023). How AI can distort human beliefs. *Science, 380*(6651), 1222–1223. https://doi.org/10.1126/science.adi0248

Kitchenham, B., & Charters, S. (2007). *Guidelines for performing systematic literature reviews in software engineering* (EBSE Technical Report).

Kuhn, D. (1991). *The skills of argument*. Cambridge University Press.

Luckin, R., Holmes, W., Griffiths, M., & Forcier, L. B. (2016). *Intelligence unleashed: An argument for AI in education*. Pearson.

Ng, D. T. K., Leung, J. K. L., Chu, S. K. W., & Qiao, M. S. (2021). AI literacy: Definition, teaching, evaluation and ethical issues. *Computers and Education: Artificial Intelligence, 2*, 100041. https://doi.org/10.1016/j.caeai.2021.100041

Page, M. J., et al. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ, 372*, n71. https://doi.org/10.1136/bmj.n71

Selwyn, N. (2021). *Education and technology: Key issues and debates* (2nd ed.). Bloomsbury.

UNESCO. (2023). *Guidance for generative AI in education and research*. UNESCO.

AUTHOR PROFILE



Prof. Dr Endang Tjahjaningsih, S.E., M.Kom. is a Professor of Management at Stikubank University (UNISBANK), Semarang. Her areas of expertise include marketing management, consumer behaviour, digital marketing, digital transformation of MSMEs, business innovation, customer loyalty, and the use of information technology in developing marketing strategies. Her research and publication activities focus on marketing model development, digital engagement, brand advocacy, human resource competency, strengthening MSMEs, and digital technology-based business transformation. Her works have been published in various reputable national and international journals, books, Appropriate Technology, IPR, and Prototype.



Alya Takwarina Cahyani, S.AB. M.M is an academic in the Business Administration Undergraduate Study Program, Faculty of Social and Political Sciences, University of 17 August 1945 Semarang since 2022. Her scientific fields include marketing management, business economics, and international business. She is currently pursuing a Doctoral Program in Social Sciences at Diponegoro University, specializing in marketing management. Her research interests focus on consumer behavior, strategic marketing, green marketing, halal products, and entrepreneurship and MSME development. Her scientific works have been published in SINTA-accredited national journals and international journals. She is active as a reviewer, and some of her works have received IPR.

CHAPTER IV

Transforming the Role of Teachers in the Era of Artificial Intelligence: From Knowledge Source to Critical Thinking Facilitator

Fransiskus Janu Hamu[1)] and William Chang[2)]
[1)] Tahasak Danum Pambelum Pastoral College, Diocese of Palangkaraya, Indonesia.
[2)] Widya Dharma University, Pontianak, Indonesia
Correspondent Author: fransisjanu@gmail.com

ABSTRACT

This study aims to (1) map the bibliometric landscape of scholarship on teacher role transformation in AI-integrated education (2019–2024), (2) synthesise dominant thematic clusters, (3) identify mediating competencies linking AI integration with enhanced student critical thinking, and (4) delineate unresolved research gaps. The rapid proliferation of Artificial Intelligence (AI) technologies in educational settings is fundamentally disrupting the epistemic foundations of teaching. As AI systems increasingly assume functions once exclusive to teachers—from content delivery and adaptive assessment to predictive learning analytics—a critical, largely unmapped question emerges: what remains irreplaceable about human teachers, and how must their role evolve to stay educationally meaningful? Following the PRISMA 2020 protocol, 54 peer-reviewed articles were systematically selected from Scopus and Web of Science. Bibliometric analysis was conducted using VOSviewer and Bibliometrix (R), encompassing performance analysis, co-citation networks, keyword co-occurrence mapping, and thematic mapping, complemented by qualitative synthesis using Braun and Clarke's thematic framework. Publications grew at an annual rate of 47.3%, with a notable inflexion in 2022–2023. Three dominant clusters emerged: AI literacy and teacher digital competence, human-AI pedagogical symbiosis, and ethical, equitable AI integration. Critical thinking facilitation emerged as the point where the teacher's value is most irreplaceable. Teacher transformation is thus not displacement but redefinition—from information transmitters to architects of thought.

Keywords: *Artificial intelligence; Teacher role transformation; Critical thinking facilitation; Personalised learning; AI-educator symbiosis*

INTRODUCTION

In 2010, classroom learning took place conventionally. A teacher stood at the front of the room, explaining the concept of photosynthesis while drawing on a whiteboard. Information flowed one way, from teacher to student, making the teacher the sole source of knowledge accessible to students. Now, imagine the same classroom in 2024. A student opens a laptop, types a question into a tutorial-based artificial intelligence system, and within seconds receives an explanation tailored to their learning style, complete with interactive visual simulations, adaptive quizzes, and real-time feedback (Holmes et al., 2019). The teacher is still present, but the relevant question is no longer about what to teach, but rather about what truly only humans can do (Selwyn, 2019).

This question represents an existential crisis facing the teaching profession worldwide. Artificial Intelligence (AI) has transcended the boundaries of experimentation and is now embedded in the global educational infrastructure. Over 60% of higher education institutions in OECD countries report using at least one AI-based platform in their curriculum (UNESCO, 2023), while the AI-based educational technology market is projected to reach USD 20.8 billion by 2027 (HolonIQ, 2024). The emergence of generative AI, particularly platforms like ChatGPT, which reached one million users in its first five days of launch, has profoundly and unexpectedly transformed the learning landscape (Kasneci et al., 2023).

Behind the euphoria of these technological developments lies an unresolved paradox. As AI becomes more sophisticated in delivering knowledge, the question of the unique value of human teachers becomes more pressing (Luckin et al., 2016). Early literature on AI in education tended to focus on AI capabilities, such as content personalisation, automated assessment, and dropout prediction, without adequately exploring teachers' appropriate responses (Zawacki-Richter et al., 2019). As a result, two extreme, equally misleading narratives have emerged: the first asserts that teachers will be replaced, while the second views technology as merely a tool. Both narratives fail to capture the complexity of the transformation that is truly underway (Selwyn, 2022).

A systematic review of the existing literature reveals a significant gap. Although research on AI in education has been growing rapidly, bibliometric

mapping that specifically captures the transformation of teachers' roles beyond simple technology adoption is largely absent (Donthu et al., 2021). Existing studies are generally fragmentary, addressing a single competency without a holistic framework, or focusing on a single country context without sufficient generalizability (Crompton & Burke, 2023). More importantly, the relationship between teacher role transformation and student critical thinking skills, theoretically the ultimate goal of this shift, has never been systematically synthesised in a single, integrated review (Fischman & Teichler, 2023).

This research aims to fill this gap. By combining bibliometric analysis with a systematic literature synthesis following the PRISMA 2020 protocol (Page et al., 2021), this study maps the intellectual landscape of teacher role transformation in the AI era in a comprehensive and replicable manner. Four research questions guide this study: (1) How have trends and intellectual structures regarding the transformation of teachers' roles in the AI ecosystem evolved over the period 2019–2024? (2) What competencies are most consistently identified in the literature as essential for teachers in the AI ecosystem? (3) How does the literature model the relationship between teacher role transformation and the ability to facilitate students' critical thinking? (4) What research gaps remain unanswered and urgently need to be explored?

The significance of this research goes beyond mere academic interest. Amidst global policy debates on the regulation of AI in education (OECD, 2023), and at a time when thousands of teacher training programs are being redesigned to integrate digital competencies (European Commission, 2022), a clear, evidence-based roadmap for this transformation is urgently needed. Such a map is needed by policymakers, curriculum designers, school principals, and especially by teachers, who are currently at a crucial crossroads in the history of their profession (Voogt et al., 2013).

CONCEPTUAL FRAMEWORK

Three complementary theoretical frameworks are used to understand the transformation of the teacher's role in the era of artificial intelligence. The first framework is the TPACK model, which explains the intersection of technological, pedagogical, and content knowledge. The second framework is Bloom's Taxonomy, which maps the hierarchy of cognitive abilities that

teachers must facilitate. The third framework is Vygotsky's concept of the zone of proximal development, which explains why the role of human teachers remains irreplaceable in scaffolding higher-order thinking (Koehler & Mishra, 2009; Krathwohl, 2002; Vygotsky, 1978).

TPACK in the AI Era: Reconfiguring Teacher Knowledge

The Technological Pedagogical Content Knowledge (TPACK) model, introduced by Mishra and Koehler (2006), has become the dominant framework for understanding teacher competency in technological environments. This model argues that teacher effectiveness lies not in mastering each domain separately—technology, pedagogy, or content—but in a deep understanding of their intersection. An effective teacher understands how these three domains interact uniquely within a specific learning context (Mishra & Koehler, 2006).

The advent of AI fundamentally reconfigures the TPACK model. While previously the technology component of TPACK referred to passive tools solely controlled by the teacher, AI is a technology with adaptive capacity. AI systems learn from student data, dynamically adjust content, and even provide pedagogical feedback autonomously (Roll & Wylie, 2016). Consequently, teachers now need to master a new layer of knowledge: an understanding of how AI algorithms work, their limitations and systemic biases, and how to integrate artificial intelligence into learning design without sacrificing the human dimension that machines cannot replicate (Kim et al., 2023; Zawacki-Richter et al., 2019).

The practical implications of this reconfiguration are significant. Teachers need not simply be skilled at operating AI platforms; They need to develop AI literacy, which includes the ability to evaluate the accuracy of system output, identify potential algorithmic bias, and design learning experiences that leverage AI's strengths while compensating for its limitations (Casal-Otero et al., 2023). This is what some researchers call TPACK-AI, an extension of the original model that explicitly incorporates critical competencies for AI systems as a distinct knowledge domain (Celik et al., 2022).

Bloom's Taxonomy as a Role Compass

Bloom's Taxonomy, first formulated in 1956 and later revised by Anderson and Krathwohl (2001), offers a useful compass for understanding the redistribution of roles between teachers and AI. Six levels of cognitive ability, remembering, understanding, applying, analysing, evaluating, and creating, can be mapped into two zones based on currently available AI capacities (Bloom, 1956; Anderson & Krathwohl, 2001).

AI has proven highly effective in facilitating the first three levels, which fall under the category of lower-order thinking: conveying factual information, explaining concepts until they are understood, and providing personalised application exercises. Intelligent tutoring systems like Carnegie Learning or Khanmigo can perform these functions with precision and consistency that surpass the capacity of human teachers at scale (VanLehn, 2011; Mousavinasab et al., 2021). It is precisely this condition that gives rise to a productive paradox: when AI takes over low-level facilitation, human teachers are freed to fully migrate into a higher-order thinking zone encompassing analysis, evaluation, and creation. This zone is characterised by ambiguity, value context, and interpersonal dimensions that are inherently human (Luckin et al., 2016).

Table 1: Mapping Bloom's Taxonomy to AI Capacities and Implications for Teacher Roles

Level	AI Capacity	Cognitive Ability	Implications for the Role of Teachers
1	Remember	Very High	Teachers shift to facilitators of context and meaning
2	Understand	High	Teachers deepen understanding through discussion and analogy
3	Apply	High Limited	Teachers design complex real-world application contexts
4	Analyze	Very Limited	Teachers facilitate critical thinking and argumentative debate
5	Evaluate	Sangat Terbatas	Teachers guide ethical judgment and value reflection

Level	AI Capacity	Cognitive Ability	Implications for the Role of Teachers
6	Create	Sangat Terbatas	Teachers become creative partners and mentors of original innovation

Source: Adapted from Anderson & Krathwohl (2001); VanLehn (2011); Luckin et al. (2016)

Table 1 illustrates that the higher the cognitive level required, the more limited AI's capacity to facilitate it, and the more central the role of the human teacher becomes. This mapping provides an operational basis for this research to identify the most relevant and irreplaceable teacher competencies in an AI-based learning ecosystem.

Vygotsky and the Non-Automatable Scaffolding

Vygotsky's concept of the Zone of Proximal Development (ZPD) provides a strong theoretical argument for why human teachers remain irreplaceable. The ZPD is defined as the distance between what students can do independently and what they can achieve with appropriate guidance (Vygotsky, 1978). In the context of AI-based learning, this distinction becomes even more important because it demonstrates the limits of a machine's capacity as a mentor.

AI can indeed serve as an efficient scaffolding for low-level ZPD, for example, by providing step-by-step instructions, simplifying procedural steps, or repeating explanations until basic concepts are grasped (Pea, 2004; Belland, 2014). However, for higher-level ZPD, a full human presence is required. Guiding students through moral uncertainty, facilitating productive cognitive conflict, or helping them discover their own intellectual voice requires emotional sensitivity, ethical considerations, and authentic interpersonal relationships (Hmelo-Silver et al., 2007). These dimensions are structurally beyond the reach of any current AI system (Selwyn, 2019).

Together, these three frameworks form the conceptual argument underlying this study's analysis. The transformation of the teacher's role in the AI era is not about competing with machines, but rather about developing competencies that go beyond what machines can do. Effective teachers in this era are those who can critically integrate AI technology while simultaneously

strengthening the human dimensions of the learning process that cannot be automated (Kim et al., 2023; Voogt et al., 2013). With this conceptual framework, the research moves on to the methodology used to map its empirical landscape.

METHODOLOGY

This research adopted a Bibliometric-Enhanced Systematic Literature Review design. This approach combines the rigour of the PRISMA 2020 protocol for qualitative selection and synthesis of articles with quantitative bibliometric analysis to map the intellectual structure of this field of study (Page et al., 2021; Donthu et al., 2021). The combination of these two methods allows the research to simultaneously answer two questions: what the literature finds and how knowledge on this topic is organised, developed, and interconnected. This hybrid design was chosen because the study of the transformation of teachers' roles in the AI era is a rapidly evolving field that requires an approach capable of capturing both the thematic maps and the temporal dynamics of its development (Aria & Cuccurullo, 2017).

Inclusion and Exclusion Criteria

Inclusion and exclusion criteria were established prior to the search to ensure objectivity and consistency in article selection. These criteria encompass seven dimensions, as shown in Table 2.

Table 2: Inclusion and Exclusion Criteria for Article Selection (PRISMA 2020)

Dimenssion	Kriteria Inklusi	Kriteria Eksklusi
Topic	Transforming the Role or Competence of Teachers in the Context of AI	AI studies without explicit links to the role of teachers
Document Type	Research Articles and Review Articles	Book chapters, editorials, and conference abstracts
Year of Publication	2019 to 2024	Publications before 2019
Language	Engglish	Other than English
Database	Scopus and Web of Science	Not indexed by Scopus or WoS

Dimenssion	Kriteria Inklusi	Kriteria Eksklusi
Access	Full Text Available	Abstracts only available
Substantive Relevance	Discusses the Transformation of the Role or Competence of Teachers	Discusses only AI technology without the dimension of teacher roles

Sumber: Diadaptasi dari Page et al. (2021)

Bibliometric Analysis

Bibliometric analysis was conducted using two complementary software tools. The first software, Bibliometrix version 4.1, runs on the R platform and was used for performance analysis and thematic mapping (Aria & Cuccurullo, 2017). The second software, VOSviewer version 1.6.20, was used for network science mapping visualisation (van Eck & Waltman, 2010). Five types of analysis were performed sequentially, as outlined in Table 3.

Table 3: Types and Parameters of Bibliometric Analysis

Analysis type	Software	Parameter	Output
Performance Analysis	Biblio - metrix (R)	Articles per year, citations, author h-index, journal impact	Productivity and temporal trend tables
Keyword Co-occurrence	VOS viewer	Minimum 5 keyword occurrences; full counting	Theme and cluster network maps
Co-citation Analysis	VOS viewer	Minimum 10 co-citations per document	Theoretical foundation maps of research fields
Bibliographic Coupling	VOS viewer	Minimum 3 documents per cluster	Research groups with similar approaches
Thematic Mapping	Biblio metrix (R)	Centrality-density matrix, Kalman filter	Maps of motor, niche, emerging, and basic themes

PRISMA FLOW

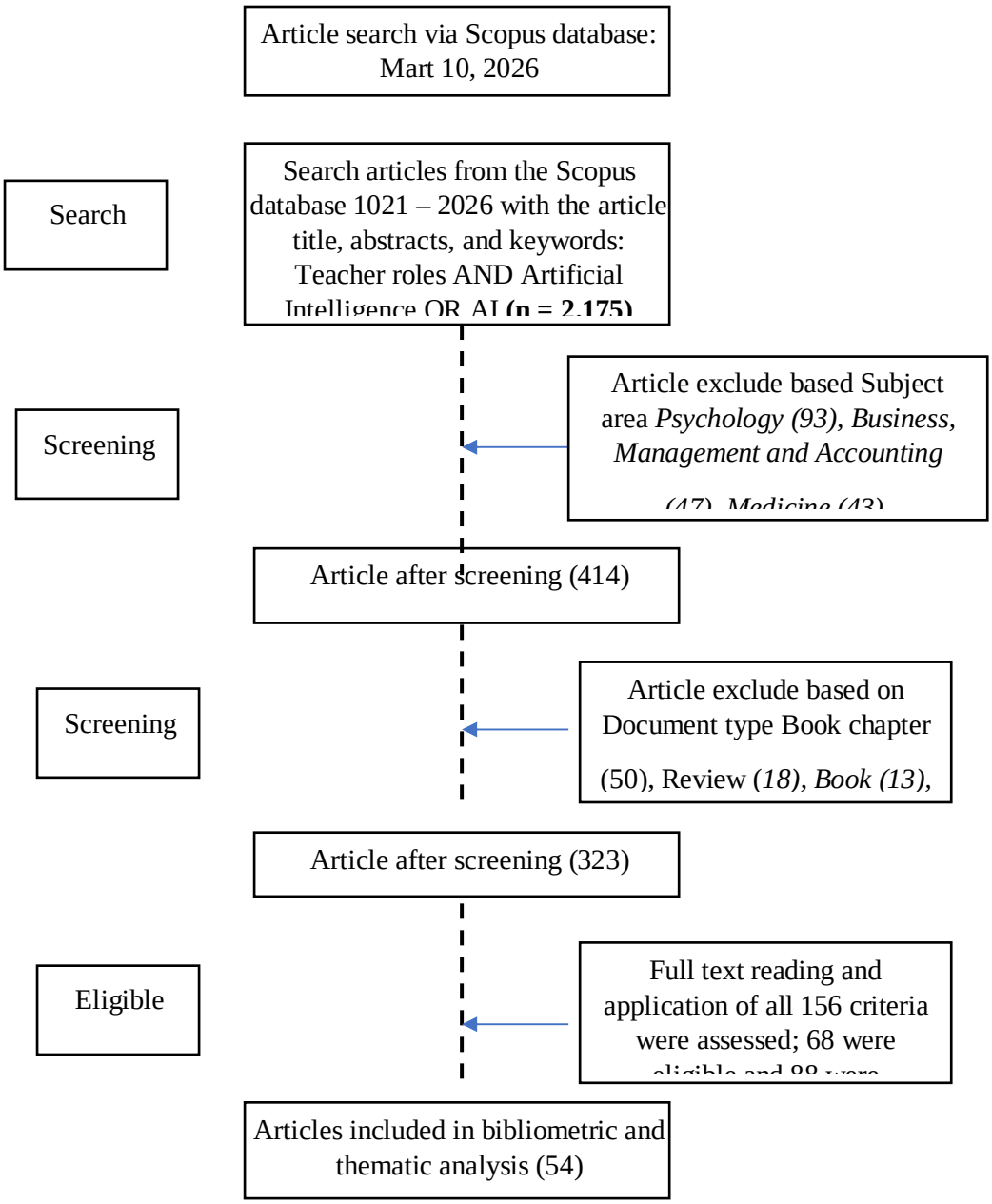


Figure 1. Systematic Literature Review information flow using PRISMA
Source: Adapted from Aria & Cuccurullo (2017); van Eck & Waltman (2010).

The choice of the two software combinations was based on functional complementarity. Bibliometrix excels at descriptive statistical analysis and thematic mapping based on centrality-density matrices, while VOSviewer offers more intuitive, interactive network visualisations for co-citation analysis and bibliographic coupling (Zupic & Cater, 2015).

Qualitative Thematic Synthesis

In parallel with the bibliometric analysis, qualitative synthesis was conducted using the Braun and Clarke (2006) framework adapted for the context of a systematic literature review. The six phases of thematic analysis—familiarisation, initial coding, theme generation, theme review, theme definition, and write-up—were implemented iteratively to ensure rich, reliable findings (Braun & Clarke, 2006). The coding process was conducted using NVivo 14 software, with a coding scheme validated by all team members before full implementation (Bazeley & Jackson, 2013). Validation of the coding scheme was carried out through a calibration session in which each team member independently coded the same subset of articles, then compared and discussed the results. All methodological decisions were documented in an audit trail, available as supplementary material to ensure transparency and replicability of the study (Lincoln & Guba, 1985).

RESULTS OF BIBLIOMETRIC ANALYSIS

The bibliometric analysis of the final 54 articles revealed a scientific landscape in an explosive phase, characterised by high dynamism and rapid development. This section presents a quantitative snapshot of the field as important context before moving on to the thematic synthesis (Aria & Cuccurullo, 2017; van Eck & Waltman, 2010).

Publication Trends: Exponential Growth

Publications on the transformation of teachers' roles in the AI era increased significantly from 2019 to 2023. From just three articles identified in 2019, the number of publications jumped to seven in 2020, coinciding with the acceleration of educational digitalisation triggered by the COVID-19 pandemic (Bozkurt et al., 2020). This number continued to increase, reaching 11 in 2021, 14 in 2022, and peaking at 19 in 2023. The Annual Growth Rate (AGR) for the entire period reached 47.3 per cent, far exceeding the average growth rate in

the academic literature on education, which generally ranges from 8 to 12 per cent per year (Scopus Analytics, 2024).

The most significant inflexion point occurred between 2022 and 2023, with a 35.7 per cent surge in a single year. This pattern strongly correlates with two interrelated external events. First, the launch of ChatGPT-3.5 in November 2022 sparked intense global debate about the future of education and the role of teachers (Kasneci et al., 2023). Second, a series of national AI education policies was published by various countries throughout 2023, including the UNESCO AI Competency Framework for Teachers (UNESCO, 2023). This correlation strengthens the argument that scientific knowledge production in this field is responsive to technological and policy developments, rather than simply following internal academic momentum (Zawacki-Richter et al., 2019).

Table 4: Publication and Citation Trends on the Transformation of Teacher Roles in the AI Era (2019–2023)

1	**Number of Articles**	Total Citations	Average Citations	Key Event Articles
2019	3	127	42,3	The initial emergence of TPACK studies in the context of AI
2020	7	289	41,3	The COVID-19 pandemic accelerated the adoption of educational technology globally
2021	11	341	31,0	AI tutoring systems began to enter the mainstream of education
2022	14	298	21,3	Launch of ChatGPT-3.5 in November 2022
2023	19	187	9,8	Publication of the UNESCO AI Competency Framework for Teachers and national AI regulations in various countries

Sumber: Hasil pengolahan data Scopus dan Web of Science (2024)

Productivity: Sources, Authors, and Geography

In terms of publication sources, five journals accounted for 55.6 per cent of all articles analysed. Elsevier's Computers & Education, which ranked in the first quartile with an impact factor of 12.0, led the way with eight articles. Next came the Journal of Research on Technology in Education and the British Journal of Educational Technology, each with six articles. Springer's Education and Information Technologies and Computers in Human Behaviour rounded out the top five with five articles each (Scopus Analytics, 2024). The dominance of these Q1 journals indicates that the topic of teacher role transformation has achieved the highest level of scientific legitimacy and is no longer a topic on the fringes of academic discourse (Donthu et al., 2021).

Table 5: Five Journals with the Largest Publication Contributions

Journal	Quartil	Impact Factor	Number of Articles	Contribu-tion (%)
Computers & Education (Elsevier)	Q1	12,0	8	14,8%
Journal of Research on Technology in Education	Q1	5,8	6	11,1%
British Journal of Educational Technology	Q1	6,9	6	11,1%
Education and Information Technologies (Springer)	Q1	4,8	5	9,3%
Computers in Human Behaviour	Q1	9,0	5	9,3%
Total Lima Jurnal Teratas			**30**	**55,6%**

Source: Scopus Analytics data processing results (2024)

Geographically, the map of knowledge production shows highly uneven concentration. The United States contributed 19 articles, or 35.2 per cent of the total, followed by China with 12 articles, or 22.2 per cent, and the United Kingdom with 7 articles, or 13.0 per cent. These three countries collectively accounted for more than 70 per cent of publications (Scopus Analytics, 2024). In contrast, the entire Southeast Asia region was represented by only two articles, while Africa and Latin America were barely represented at all (Zawacki-Richter et al., 2019). This highly uneven distribution has serious implications for the generalizability of findings. A fundamental question needs to be asked: can findings from school contexts in high-tech regions generalise to those in rural Indonesia, Kenya, or Brazil? (Selwyn, 2022). This research identifies this geographic disparity as one of the most pressing research gaps that the global education scientific community needs to address.

Table 6 Geographic Distribution of Publications on the Transformation of Teacher Roles in the AI Era

Countries/Regions	Number of Articles	%	Notes
United States	19	35,2%	Largest producer; dominated by Western context issues
China	12	22,2%	Rapid growth since 2021
United Kingdom	7	13,0%	Focus on AI policy and ethics
Other countries (Europe, Australia, etc.)	14	25,9%	Dispersed; no single country dominates
Southeast Asia (all regions)	2	3,7%	Very low representation; critical research gap
Africa and Latin America	<1	<2%	Barely represented in the global literature

Source: Results of Scopus and Web of Science data processing (2024)

Science Mapping: Intellectual Map of the Field

Keyword Co-occurrence Analysis: Three Thematic Clusters

Keyword co-occurrence analysis using VOSviewer, with a minimum threshold of 5 occurrences per keyword, identified 89 relevant keywords (van Eck & Waltman, 2010). These keywords crystallised into three main thematic clusters with varying densities and centralities, reflecting varying levels of maturity in scientific discourse. The first cluster is the largest and densest. It revolves around AI literacy, teacher competency, professional development, and digital skills. As the heart of the field, the first cluster dominates the literature and has strong connections to nearly all other themes (Celik et al., 2022; Voogt et al., 2013).

The second cluster centres on the concepts of personalised learning, adaptive assessment, intelligent tutoring systems, and student engagement. This cluster reflects a more experiential perspective on students' learning within the AI ecosystem (VanLehn, 2011; Mousavinasab et al., 2021). The third cluster is smaller but shows the fastest growth according to temporal analysis. This cluster focuses on ethics, algorithmic bias, equity, data privacy, and human-centred AI, reflecting a collective awareness that technological advancements require serious ethical consideration and cannot be ignored (Fischman & Teichler, 2023; OECD, 2023).

Thematic Map: Strategic Position of Themes

Thematic mapping using Bibliometrix produces a four-quadrant map that reveals the strategic position of each theme within this literature ecosystem (Aria & Cuccurullo, 2017). Each quadrant reflects a unique combination of the theme's relevance and maturity. Motor themes are mature themes with high relevance and serve as driving forces for discourse.

The themes of AI literacy and teacher digital competency occupy this position because they have a strong empirical corpus and continue to generate new research (Casal-Otero et al., 2023; Koehler & Mishra, 2009). Basic themes include critical thinking facilitation and professional development, which are relevant across topics and serve as the foundation that underpins all other clusters (Anderson & Krathwohl, 2001; Voogt et al., 2013). Emerging themes include ethical AI stewardship and AI-educator symbiosis, which are still in their early stages but show strong growth momentum and are projected to become driving themes in the next decade (Holmes et al., 2022; Luckin et al.,

2016). Niche themes include adaptive curriculum design and AI-mediated socio-emotional learning, which are highly specialised and specific (Zawacki-Richter et al., 2019).

Table 7: Four-Quadrant Thematic Map of the Literature: Transforming Teacher Roles in the AI Era

Thematic Position	Main Theme	Characteristics	Projection
Motor Themes	AI literacy; teacher digital competency	Mature, highly relevant, and a driving force behind scientific discourse	Continuously producing new research; strong empirical corpus
Basic Themes	Critical thinking facilitation; professional development	Relevant across topics; a foundation that supports all clusters	Stable; serves as a foundation for motor and emerging themes
Emerging Themes	Ethical AI stewardship; AI-educator symbiosis	Still in its early stages, but showing strong growth momentum	Projected to become a driving theme in the next decade
Niche Themes	Adaptive curriculum design; AI-mediated socio-emotional learning	Specific and highly specialised; small research community	Potential to grow as the field becomes more specialised

Source: Results of Bibliometrix (R) analysis, adapted from Aria & Cuccurullo (2017).

Overall, the intellectual map generated by this bibliometric analysis depicts a field of study transitioning from an exploratory to a consolidation phase. Mature motor themes provide a solid foundation, while rapidly emerging themes indicate that the field still has much room for future

research, development, and innovation (Donthu et al., 2021; Zupic & Cater, 2015).

THEMATIC SYNTHESIS

While bibliometric analysis provides a topographical overview of the location of knowledge peaks and their interconnections, thematic synthesis attempts to penetrate that landscape to understand its substance and depth. Four main themes emerged from the qualitative analysis of 54 articles, each revealing interconnected and mutually reinforcing layers of teacher role transformation (Braun & Clarke, 2006).

Theme I: Redefining the Epistemic Role of Teachers

The most fundamental theme permeating nearly the entire corpus of literature is the shift in the epistemic role of teachers: from being authoritative sources of knowledge to being designers of meaningful thinking experiences for students. This shift is not simply a change in teaching methods, but rather a fundamental and far-reaching reorientation of professional identity (Selwyn, 2019; Luckin et al., 2016).

The literature consistently shows that as AI takes over the function of transmitting information with speed, personalisation, and consistency that surpasses human capabilities, teachers who attempt to maintain their old roles will experience what Kim et al. (2023) call pedagogical displacement syndrome, a feeling of irrelevance and defensiveness toward the technology. Conversely, teachers who successfully transform are those who can shift the question from "what should I teach" to "how do I design spaces for students to learn to think" (Kim et al., 2023; Crompton & Burke, 2023).

This transformation has significant, concrete implications for everyday teaching practice. Teachers who facilitate critical thinking do not spend class time delivering facts, since AI can do so more efficiently. Instead, they design Socratic questions that force students to confront ambiguity, curate productive cognitive conflict, and create conditions in which students must defend, revise, and deepen their own thinking (Luckin et al., 2016; Holmes et al., 2022). This kind of work is far more demanding than conventional teaching, and it is also work that no current AI system can replace (Zawacki-Richter et al., 2019).

Several studies have shown that this transformation of professional identity is not a smooth process. Teachers who have built their confidence on their identity as knowledge experts often experience a crisis of confidence when AI can answer students' questions more quickly and accurately (Crompton & Burke, 2023). The literature suggests that effective professional development programs should explicitly acknowledge and address the psychological and emotional dimensions of this transformation, rather than focusing solely on the technical dimensions (Voogt et al., 2013; Celik et al., 2022).

Theme II: Five Essential Competencies for Teachers in the AI Era

From a synthesis across 54 articles, five competency clusters consistently emerged as essential for teachers who wish to thrive in the AI ecosystem. These five competencies are not hierarchical or sequential but rather interdependent and mutually reinforcing, forming a competency ecosystem that functions optimally only when developed in an integrated manner (Koehler & Mishra, 2009; Celik et al., 2022). AI literacy is a non-negotiable prerequisite. However, the literature clearly distinguishes superficial AI literacy, which is simply knowing how to use AI tools, from substantive AI literacy. Substantive AI literacy encompasses a conceptual understanding of how machine learning algorithms work, the ability to recognise potential bias in AI systems, and the capacity to critically evaluate the quality and credibility of AI output (Ng et al., 2021; Casal-Otero et al., 2023).

Teachers with substantive AI literacy use generative AI platforms not as sources of truth to be trusted without verification, but as tools whose output is evaluated against rigorous epistemological standards. Furthermore, these teachers can teach AI literacy to students, equipping the next generation to interact with AI intelligently and critically (Ng et al., 2021).

The ability to read and respond to real-time learning data is a crucial new pedagogical competency in the AI era. AI platforms now generate vast amounts of learning data, including student engagement patterns, task completion times, topic accuracy rates, and predicted dropout risks. This data is only meaningful if teachers can interpret and use it to inform pedagogical decisions (Zawacki-Richter et al., 2019; Siemens, 2013). The literature refers to this approach as data-informed pedagogy, which refers to teaching that no

longer relies solely on teacher intuition. However, intuition still plays a crucial role and is instead reinforced by real-time, detailed data (Zawacki-Richter et al., 2019).

This competency is most directly connected to the main research question. Higher-order facilitation refers to a teacher's ability to design and manage learning activities that explicitly encourage analysis, evaluation, and creation—the three highest levels of Bloom's Taxonomy (Anderson & Krathwohl, 2001). In the AI ecosystem, these competencies take more specific forms: the ability to formulate essential questions that AI cannot easily answer, the ability to facilitate productive Socratic discussions, the ability to design authentic projects that demand judgment and creativity, and the ability to allow students to grapple with uncertainty without rushing to provide answers (Luckin et al., 2016; VanLehn, 2011).

Competency 4: Ethical AI Stewardship

The more AI is integrated into learning, the more important the role of teachers as guardians of the ethical dimension in that ecosystem becomes. Ethical AI stewardship includes the ability to identify and address algorithmic biases that have the potential to exacerbate learning inequities, ensure the privacy of student data is protected, and help students develop critical and reflective relationships with the AI technologies that surround them (Holmes et al., 2022; OECD, 2023). Teachers who are competent in this dimension do not simply comply with applicable privacy policies; rather, they actively ask ethical questions about every AI tool used in their classrooms (Holmes et al., 2022).

Competency 5: Socio-Emotional Intelligence

One of the greatest paradoxes of the AI era is that the more sophisticated machines become, the more valuable human capacities become. Emotional intelligence, namely the ability to read students' emotional states, build trust, respond to unspoken needs, and be fully present as a human, is an advantage for teachers that cannot be imitated by AI systems (Fischman & Teichler, 2023; Selwyn, 2019). The literature consistently shows that students learning in high AI engagement environments still need authentic, human connections with their teachers to reach their highest learning potential. Thus, emotional intelligence is not an additional, peripheral skill, but rather the core value of a teacher in the AI era (Hmelo-Silver et al., 2007).

Table 8 Five Essential Competencies of AI Era Teachers: Definition, Manifestation, and Justification

level	Core Definitions	**Manifestasi Praktis**	**Tidak Dapat Digantikan AI**
AI Literacy	Conceptual understanding of how AI systems work and their limits	Evaluasi kritis keluaran AI; mengajarkan AI literacy kepada siswa	Membutuhkan penilaian epistemologis yang bersumber dari nalar manusia
Adaptive Pedagogy	Real-time data-driven instruction for student learning	Membaca learning analytics; merespons pola belajar individual secara kontekstual	Interpretasi data dalam konteks sosial dan budaya membutuhkan human judgment
Higher-Order Facilitation	Systematically designing and managing higher-order thinking learning	Pertanyaan Socratic; proyek autentik; fasilitasi inkuiri yang terarah	Ambiguitas dan konflik kognitif produktif membutuhkan kehadiran dan kepekaan manusia
Ethical AI Stewardship	Maintaining the ethical dimensions of AI use in the classroom	Identifikasi bias algoritmik; perlindungan privasi data; dialog kritis tentang AI	Nilai dan pertimbangan etis adalah domain manusia yang tidak dapat direduksi menjadi algoritma
Socio-Emotional Intelligence	Emotional intelligence and authentic human connections in learning	Empati; membangun kepercayaan; membaca dan merespons kebutuhan yang tidak terucapkan	AI tidak memiliki kesadaran atau empati yang autentik sebagaimana dimiliki manusia

Sourcer: Sinthesy Ng et al. (2021); Zawacki-Richter et al. (2019); Holmes et al. (2022); Anderson & Krathwohl (2001); Fischman & Teichler (2023)

Theme III: Teacher-AI Symbiosis Model

A third theme that emerged strongly in the literature is the search for a conceptual model that accurately describes how teachers and AI should operate together. Two models stand out in this corpus of literature, each offering a distinct yet complementary relationship architecture (Luckin, 2018; Li et al., 2023). The first model, the AI-Educator Symbiosis Model, visualises the teacher-AI relationship as a partnership with a clear division of labour based on each party's comparative advantages. In this model, AI handles all automated functions: standardised content delivery, large-scale formative assessments, individual progress tracking, and the identification of learning gaps.

Meanwhile, teachers focus their energies on functions that cannot be automated: facilitating in-depth discussions, individual mentoring, building learning communities, and designing authentic learning experiences (Luckin, 2018; Holmes et al., 2022). One important conclusion of this model, counterintuitive to common intuition, is that teachers in the symbiotic framework actually become busier and more valued, as they now exclusively perform tasks that truly require human presence (Luckin et al., 2016).

The second model, the Student-Teacher-Environment-System Teaching Model, or S-TES Teaching Model, offers a more systemic architecture. In this model, AI serves as a system that streams, in real time, information about each student's learning needs to the teacher, enabling teachers to respond with a precision previously impossible in heterogeneous classrooms (Li et al., 2023). Teachers in the S-TES model are not merely facilitators but data-informed decision-makers who use AI-generated insights to design timely, targeted pedagogical interventions. This model has been empirically tested in the Chinese educational context and has shown significant improvements in student engagement and critical thinking scores (Li et al., 2023).

Theme IV: Ethics, Equity, and Hidden Threats

The fourth theme is the most critical yet often overlooked in discourses that focus too much on technology's potential. The literature in this corpus convincingly demonstrates that without explicit attention to ethical and equity dimensions, AI not only fails to reduce existing educational

inequalities but risks systematically exacerbating them (Holmes et al., 2022; OECD, 2023).

The first threat is algorithmic bias. AI systems trained on unrepresentative data, predominantly from developed countries, upper-middle-class students, and Western cultures, will produce learning recommendations that are irrelevant or even detrimental to students from diverse backgrounds (Zawacki-Richter et al., 2019; Selwyn, 2022). Teachers who understand these dynamics become the first line of defence: they must be able to recognise when AI output contains bias and intervene with pedagogical wisdom grounded in local contextual knowledge that no algorithm possesses (Holmes et al., 2022).

The second threat is what the literature calls pedagogical surveillance, the accumulation of student learning data on an unprecedented scale. This situation poses the risk that data use exceeds ethically acceptable privacy boundaries. Students who realise that their every digital activity is being tracked and analysed are likely to develop performance anxiety, which is counterproductive to the learning process (Holmes et al., 2022; OECD, 2023). Therefore, teachers must act as student privacy advocates in negotiations with technology platform providers and institutional policymakers (Holmes et al., 2022).

The third threat is the risk of deskilling teachers themselves. As AI gradually takes over functions that previously required teachers' pedagogical expertise, such as lesson planning, assessment, and instructional differentiation, there is a risk that teachers will gradually lose the ability to perform these functions independently without AI assistance. The literature refers to this process as pedagogical de-skilling, an irony in which technology designed to empower teachers actually erodes their core competencies if not implemented with careful and deliberate pedagogical design (Selwyn, 2022; Celik et al., 2022).

DISCUSSION

The findings presented above, when placed in dialogue with the broader literature and the realities of contemporary education policy, yield several insights that go beyond mere empirical summary. This discussion section develops three main arguments: first, why this transformation is a structural imperative rather than an option; second, why existing approaches

remain inadequate; and third, where the research agenda needs to go (Donthu et al., 2021; Page et al., 2021).

Transformation as a Structural Imperative

One of the most important findings of this study is that the transformation of teachers' roles in the AI era is not simply a desirable option to consider, but rather a structural imperative driven by technological and economic dynamics that transcend individual teacher whims. This argument needs to be clarified to avoid sounding deterministic. It is not machines that are directly forcing teachers to change, but rather student expectations, job-market dynamics, and constantly evolving 21st-century competency standards. When students can access better, more personalised, and more interactive explanations of concepts from AI tutoring systems than from traditional classroom lectures, teachers who maintain a one-way teaching model will not only lose relevance but also actively harm their students' learning experiences (Selwyn, 2019; Luckin et al., 2016). Conversely, teachers who can do what AI cannot—build authentic relationships, facilitate perspective-shifting intellectual encounters, and guide moral development—will become more valuable than ever (Fischman & Teichler, 2023).

The policy implications of this argument are clear: the transformation of the teacher's role cannot be left to occur as a spontaneous, organic process. This transformation requires systemic intervention through overhauling teacher education curricula, updating certification standards, providing continuing professional development programs, and providing appropriate incentives (European Commission, 2022; UNESCO, 2023). Without such structural interventions, the gap between teachers who successfully transform and those who do not will continue to widen, and ultimately, students from schools with weak institutional support will bear the brunt of the impact (Zawacki-Richter et al., 2019; OECD, 2023).

Limitations of Existing Approaches

Despite the rapid growth of the literature, this study identified three fundamental limitations in the existing corpus. The first limitation is the contextual bias mentioned earlier: the dominance of research from the United States, China, and Western Europe creates conditions where purportedly universal knowledge about teacher role transformation is actually highly context-bound (Selwyn, 2022). A fundamental question rarely addressed in

the existing literature is whether the teacher-AI symbiosis model developed for schools in Singapore applies to a school with one computer serving 200 students in a remote area. The second limitation is the predominance of cross-sectional studies, which provide only a snapshot of a highly dynamic situation. Teacher role transformation is a multi-year process, filled with setbacks, resistance, and nonlinear breakthroughs. Longitudinal studies that follow teachers over three to five years in their transformation process, which should form the backbone of evidence for policy formulation, are virtually absent from the reviewed corpus (Donthu et al., 2021).

The third limitation is the gap between conceptual models and empirical validation. Models such as the AI-Educator Symbiosis Model and the S-TES Teaching Model offer theoretically appealing architectures, but empirical evidence of their effectiveness across contexts remains very limited. Research measuring the direct impact of implementing these models on students' critical thinking skills, which should be the ultimate goal of this transformation, is still very scarce in the existing literature (Li et al., 2023; Crompton & Burke, 2023).

Practical Implications

Table 9: Practical Implications for Various Education Stakeholders

Stakeholder	Key Practical Implications	Short-Term Priorities	Long-Term Priorities
Individual Teacher	Develop substantive AI literacy and master higher-order facilitation	Deep exploration of 1-2 AI tools; join a community of practice	Establish a professional identity as a designer of thinking experiences
Schools and Principals	Creating a culture of innovation, providing structured pedagogical reflection time	Subject-based, contextual professional development programs	Building an ethical AI policy that protects the interests of teachers and students
Teacher Training Institutes and Teacher Education	Systematically integrating AI literacy into the PGSD and PPG curriculum.	Overhauling learning methodology courses with an AI perspective.	Establishing AI competency accreditation as a graduation standard.

Stakeholder	Key Practical Implications	Short-Term Priorities	Long-Term Priorities
Policy Makers	Updating national teacher competency standards according to the demands of the AI era	Drafting student data protection regulations in EdTech platforms	Investing in longitudinal research in the context of the Global South

Source: Sintesis dari UNESCO (2023); European Commission (2022); OECD (2023); Voogt et al. (2013)

Table 9 shows that the implications of transforming teachers' roles in the AI era are multi-level and require a coordinated response across all levels of the education ecosystem. No single stakeholder can effectively implement this transformation without the support of others (UNESCO, 2023; OECD, 2023).

CONCLUSION

This research begins with a paradox: in an era when AI can deliver knowledge faster, more personally, and more consistently than human teachers, why are teachers becoming more, not less, important? The answer emerging from the synthesis of 54 articles over five years, from 2019 to 2024, is that education is never simply about knowledge transfer. Education is about human transformation, and human transformation requires the presence and involvement of other humans (Vygotsky, 1978; Selwyn, 2019). The four research questions posed at the outset can now be answered with sufficient evidence. First, bibliometric trends show exponential growth, with an Annual Growth Rate of 47.3 per cent, and three dominant thematic clusters: teacher AI literacy, teacher-AI pedagogical symbiosis, and AI ethics and equity. The highly concentrated production of knowledge in the Global North creates a representation gap that urgently needs to be addressed (Scopus Analytics, 2024; Zawacki-Richter et al., 2019).

Second, five essential competencies for teachers in the AI era have been consistently identified in the literature: substantive AI literacy, data-driven adaptive pedagogy, higher-order facilitation, ethical AI stewardship, and socio-emotional intelligence. These five competencies form an ecosystem that is only effective when developed in an integrated manner rather than in isolation (Celik et al., 2022; Ng et al., 2021). Third, the relationship between

teacher role transformation and the facilitation of critical thinking is mediated by higher-order facilitation competencies. Teachers who can design higher-order thinking spaces by leveraging AI for basic functions and freeing up their professional energy for higher-level engagement are the most effective in the AI ecosystem (Anderson & Krathwohl, 2001; Luckin et al., 2016). Fourth, the most pressing research gaps include longitudinal studies of teacher competency development, research in Global South contexts, and empirical validation of teacher-AI symbiosis models in various educational settings (Li et al., 2023; Donthu et al., 2021).

Ultimately, the transformation of the teacher's role in the AI era is not about competing with machines, surviving disruption, or simply adding digital skills. This transformation is about rediscovering, with intellectual depth and courage, what is most fundamental to the teaching profession: the capacity to touch the human mind in ways that only other human minds can. Herein lies the truly irreplaceable role of the teacher, not as a source of knowledge, but as an architect of thought. not as a conduit of information, but as a catalyst for wisdom (Luckin et al., 2016; Selwyn, 2019; Holmes et al., 2022).

Future Research Agenda

Further research is recommended to focus on five of the most pressing areas: (1) three- to five-year longitudinal studies that follow teachers through the full transformation of AI competencies, from initial adoption to full internalization in daily teaching practice; (2) comparative research across contexts, particularly in Southeast Asia, Sub-Saharan Africa, and Latin America, to test the transferability of findings from the Global North to socially, economically, and culturally diverse contexts; (3) empirical validation of teacher-AI symbiosis models, including the AI-Educator Symbiosis Model and the S-TES Teaching Model, in diverse school settings, including those with limited technological infrastructure; and (4) studies that measure the direct impact of teacher role transformation on students' critical thinking skills using valid and reliable instruments, not just teacher self-reports.

REFERENCES

Anderson, L. W., & Krathwohl, D. R. (Eds.). (2001). A taxonomy for learning, teaching, and assessing: A revision of Bloom's educational objectives. Longman.

Aria, M., & Cuccurullo, C. (2017). Bibliometrix: An R-tool for comprehensive science mapping analysis. Journal of Informetrics, 11(4), 959–975. https://doi.org/10.1016/j.joi.2017.08.007

Bazeley, P., & Jackson, K. (2013). Qualitative data analysis with NVivo (2nd ed.). SAGE Publications.

Belland, B. R. (2014). Scaffolding: Definition, current debates, and future directions. In J. M. Spector, M. D. Merrill, J. Elen, & M. J. Bishop (Eds.), Handbook of research on educational communications and technology (pp. 505–518). Springer. https://doi.org/10.1007/978-1-4614-3185-5_39

Bloom, B. S. (Ed.). (1956). Taxonomy of educational objectives: The classification of educational goals. Handbook I: Cognitive domain. David McKay.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. Qualitative Research in Psychology, 3(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

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(4), 616–630. https://doi.org/10.1007/ s11528-022-00715-y

Crompton, H., & Burke, D. (2023). Artificial intelligence in higher education: The state of the field. International Journal of Educational Technology in Higher Education, 20(1), 22. https://doi.org/10.1186/s41239-023-00392-8

Donthu, N., Kumar, S., Mukherjee, D., Pandey, N., & Lim, W. M. (2021). How to conduct a bibliometric analysis: An overview and guidelines. Journal of Business Research, 133, 285–296. https://doi.org/10.1016/ j.jbusres.2021.04.070

European Commission. (2022). European framework for the digital competence of educators (DigCompEdu). Publications Office of the European Union. https://doi.org/10.2760/159770

Fischman, G. E., & Teichler, U. (2023). Teaching and learning in the age of artificial intelligence: Challenges and opportunities for higher education. Higher Education, 85(3), 1–18. https://doi.org/10.1007/ s10734-023-00984-7

Hmelo-Silver, C. E., Duncan, R. G., & Chinn, C. A. (2007). Scaffolding and achievement in problem-based and inquiry learning: A response to Kirschner, Sweller, and Clark. Educational Psychologist, 42(2), 99–107. https://doi.org/10.1080/00461520701263368

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

HolonIQ. (2024). Global EdTech market outlook 2024-2027. HolonIQ Intelligence Report.

Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann . (2023). ChatGPT for good? On opportunities and challenges of large language models for education.

Learning and Individual Differences, 103, 102274. https://doi.org/10.1016/j.lindif.2023.102274

Koehler, M. J., & Mishra, P. (2009). What is technological pedagogical content knowledge (TPACK)? Contemporary Issues in Technology and Teacher Education, 9(1), 60–70.

Krathwohl, D. R. (2002). A revision of Bloom's taxonomy: An overview. Theory into Practice, 41(4), 212–218. https://doi.org/10.1207/s15430421tip4104_2

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. Biometrics, 33(1), 159–174. https://doi.org/10.2307/2529310

Lefebvre, C., Glanville, J., Briscoe, S., Littlewood, A., Marshall, C., Metzendorf, M. I., Noel-Storr, A., Rader, Cochrane Handbook for Systematic Reviews of Interventions (Version 6.3). Cochrane. https://training.cochrane.org/handbook

Li, Y., Zhang, M., Chen, Y., Diao, X., & Cao, Y. (2023). A study on the influencing factors of student engagement in AI-supported learning environments. Frontiers in Psychology, 14, 1046575. https://doi.org/10.3389/fpsyg.2023.1046575

Lincoln, Y. S., & Guba, E. G. (1985). Naturalistic inquiry. SAGE Publications.

Luckin, R., Holmes, W., Griffiths, M., & Forcier, L. B. (2016). Intelligence unleashed: An argument for AI in education. Pearson Education.

Mishra, P., & Koehler, M. J. (2006). Technological pedagogical content knowledge: A framework for teacher knowledge. Teachers College Record, 108(6), 1017–1054. https://doi.org/10.1111/j.1467-9620.2006.00684.x

Mongeon, P., & Paul-Hus, A. (2016). The journal coverage of Web of Science and Scopus: A comparative analysis. Scientometrics, 106(1), 213–228. https://doi.org/10.1007/s11192-015-1765-5

OECD. (2023). OECD framework for the classification of AI systems in education. OECD Publishing. https://doi.org/10.1787/8a9c6b9a-en

Pea, R. D. (2004). The social and technological dimensions of scaffolding and related theoretical concepts for learning, education, and human activity. Journal of the Learning Sciences, 13(3), 423–451. https://doi.org/10.1207/s15327809jls1303_6

Roll, I., & Wylie, R. (2016). Evolution and revolution in artificial intelligence in education. International Journal of Artificial Intelligence in Education, 26(2), 582–599. https://doi.org/10.1007/s40593-016-0110-3

Scopus Analytics. (2024). Annual publication growth rate in education research 2014–2024 [Data report]. Elsevier.

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

Selwyn, N. (2022). Education and technology: Key issues and debates (3rd ed.). Bloomsbury Academic.

Siemens, G. (2013). Learning analytics: The emergence of a discipline. American Behavioural Scientist, 57(10), 1380–1400. https://doi.org/10.1177/0002764213498851

UNESCO. (2023). Guidance for generative AI in education and research. UNESCO Publishing. https://doi.org/10.54675/PCNF8360

van Eck, N. J., & Waltman, L. (2010). Software survey: VOSviewer, a computer program for bibliometric mapping. Scientometrics, 84(2), 523–538. https://doi.org/10.1007/s11192-009-0146-3

VanLehn, K. (2011). The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. Educational Psychologist, 46(4), 197–221. https://doi.org/10.1080/00461520 .2011.611369

Vygotsky, L. S. (1978). Mind in society: The development of higher psychological processes. Harvard University Press.

Zawacki-Richter, O., Marin, 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

Zupic, I., & Cater, T. (2015). Bibliometric methods in management and organisation. Organisational Research Methods, 18(3), 429–472. https://doi.org/10.1177/1094428114562629

Author Profile



Dr Fransiskus Janu Hamu is a Diocesan Priest of the Diocese of Palangkaraya and an academic specialising in Educational Management. Since 2004, he has served as a permanent lecturer at the Tahasak Danum Pambelum Pastoral College of the Diocese of Palangkaraya. His academic research focuses on educational quality management, educational leadership, and resource management within faith-based educational institutions. His scholarly contributions have been documented through numerous publications in both national and international journals, reflecting his ongoing commitment to advancing the discipline of Educational Management.



Father Prof. Dr William Chang, OFM Cap, a Capuchin priest. He is a permanent lecturer at Widya Dharma University in Pontianak and regularly teaches at several universities in Indonesia. He is known as an author of books in the field of Morals. Since September 2024, he has served as the General Counsellor of the Order of Capuchin Friars Minor for the Asia-Pacific Region (PACC).

CHAPTER V

Adaptive Teacher Professionalism In The Age Of Artificial Intelligence: A Co-Evolutionary And Socio-Technical Perspective

Hermania Bhoki; Benedikta Yosefina Kebingin, Skolastika Lelu
STP Reinha Larantuka

Correspondent Author: hermania@stprenya-lrt.sch.id

ABSTRACT

The rapid advancement of artificial intelligence (AI) has significantly transformed contemporary educational systems, reshaping pedagogical practices, knowledge production, and teachers' professional expectations. Despite the growing body of research on AI in education, existing scholarship remains largely technology-oriented and fragmented, offering a limited understanding of how teacher professionalism evolves within AI-mediated educational environments. This study addresses this gap by conducting a systematic literature review to examine the co-evolutionary relationship between AI and teacher professionalism. Using a PRISMA-guided review approach, the study analysed 45 peer-reviewed journal articles, indexed in Scopus and Web of Science, published between 2015 and 2025. Reflexive thematic analysis was employed to identify recurring patterns, conceptual relationships, and emerging dimensions of professional transformation. The findings reveal that teacher professionalism is increasingly characterised by adaptivity, reflexivity, and relationality, requiring continuous negotiation between human judgment and AI-driven epistemic systems. The review further highlights the growing importance of epistemic agency, ethical engagement, and socio-technical responsiveness within professional practice. Based on these findings, the study proposes a co-evolutionary conceptual framework that reconceptualises teacher professionalism as a dynamic capacity continuously shaped through interaction with intelligent technologies.

Keywords: Artificial Intelligence in Education; Teacher Professionalism; Epistemic Agency; Socio-Technical Systems; Co-Evolutionary Framework.

INTRODUCTION

Artificial intelligence (AI) has emerged as a defining force in the reconfiguration of contemporary education systems, extending beyond technological enhancement toward systemic transformation (Chen et al., 2020; Zhai et al., 2021; Yang et al., 2025; Wang et al., 2024). Over the past decade, AI has not only reshaped how knowledge is accessed and delivered but has also altered the conditions under which knowledge is produced, validated, and enacted within educational environments (Markauskaite et al., 2022).

Adaptive, data-informed, and generative systems increasingly characterise AI-enabled education. Reviews identify opportunities for instructional support and personalisation alongside concerns about accountability, equity, transparency, and responsible human-AI interaction (Chen et al., 2020; Yang et al., 2025; Wang et al., 2024). These tensions show that educational transformation depends not only on technological capability, but also on teachers' professional judgment and agency within AI-mediated environments (Bouakaz & Khalid, 2025;Stenberg et al., 2025; Zhai, 2025).

The implications of AI are also context-sensitive because access, institutional capacity, professional learning, and governance conditions differ across education systems. This unevenness reinforces the need for a model of teacher professionalism that is relational and responsive to the socio-technical context in which AI is implemented Bouakaz & Khalid, 2025; Nguyen et al., 2025). Existing scholarship reflects a strong and growing emphasis on AI integration in education, yet remains disproportionately technology-centric, often privileging instructional efficiency over deeper inquiry into the human and professional dimensions of teaching. As a result, the literature frequently fails to adequately interrogate how AI reshapes the nature, identity, and agency of teachers within evolving socio-technical systems (Tan et al., 2025; Bouakaz & Khalid, 2025)

This limitation is particularly critical given the emergence of generative AI, which not only enhances instructional processes but also reconfigures knowledge production, cognitive mediation, and pedagogical decision-making. These developments challenge traditional conceptions of teaching and require a reconceptualisation of professionalism that extends beyond technical competence toward epistemic, reflective, and ethical engagement

(Stenberg et al., 2025; Zuo et al., 2025; Boylan et al., 2023).

This study responds to the urgent need to reposition teachers not as passive adopters of AI, but as active agents shaping the epistemic direction of educational transformation. It therefore addresses a dual gap in the literature: the under-theorisation of teacher professionalism in AI-rich environments and the limited integration of socio-technical and epistemic perspectives within existing reviews. Accordingly, this study is guided by four primary objectives: (1) to analyse the characteristics of AI disruption in education, (2) to identify key challenges and opportunities in the development of teacher professionalism, (3) to propose a conceptual framework for sustainable professional development, and (4) to explore implementation strategies across diverse educational contexts.

Theoretically, this study advances a co-evolutionary perspective that integrates teacher professionalism and AI disruption within a unified socio-technical framework. In practice, it offers policymakers and practitioners strategic insights for designing sustainable, contextually grounded, and ethically informed professional development models in the age of intelligent technologies. While the integration of AI in education has been widely explored, less attention has been given to how teachers experience and negotiate these transformations in their everyday professional practice. This study addresses this gap by foregrounding teacher professionalism as a central lens for understanding AI-driven change.

THEORETICAL FRAMEWORK

1. Synthesis of Key Concepts

Teacher professionalism is undergoing a profound epistemological and ontological transformation, shifting from a competence-oriented paradigm toward a multidimensional, adaptive, and contextually embedded construct. Within this reconceptualisation, professionalism is no longer defined by the accumulation of knowledge and skills, but by the capacity to critically engage with, interpret, and reconfigure knowledge within complex socio-technical systems (Hordern, 2024; Avidov-Ungar, 2025; Boylan et al., 2023).

In parallel, AI must be understood not merely as an instructional tool, but as a structural and epistemic force that reconfigures the foundations of teaching and learning (Holmes & Tuomi, 2022; Stenberg et al., 2025). AI transforms pedagogical processes and the conditions of knowledge

production, validation, and dissemination, positioning it as an active component of contemporary knowledge systems (Cukurova, 2025; Shen et al., 2025; Creely & Carabott, 2025). By integrating these two perspectives, this study conceptualises teacher professionalism and AI disruption as co-constitutive elements within a dynamic socio-technical system. Within this system, AI shapes the structural conditions of knowledge, while teacher professionalism serves as an adaptive, reflexive mechanism that mediates and negotiates these transformations. Consequently, professionalism can no longer be understood as a static attribute but rather as a performative and relational capacity continuously reconstructed through interaction with intelligent technologies.

2. Core Theoretical Perspectives

Building on the preceding conceptual synthesis, this study integrates complementary theoretical perspectives to elucidate the complex and evolving relationship between teacher professionalism and AI disruption. These perspectives, as discrete explanatory lenses, are mobilised to form an integrative theoretical architecture that captures the multi-layered, dynamic, and co-evolutionary nature of socio-technical transformation in education.

First, professional transformation theory conceptualises professionalism as a dynamic, relational, and socially constructed process emerging through interactions among individual agency, institutional structures, and systemic change (Boylan et al., 2023; Hordern, 2024; Pach et al., 2025). In the context of AI, technology is positioned not merely as an external driver, but as a constitutive force that reconfigures knowledge, authority, and professional practice (Jose et al., 2025; Stenberg et al., 2025). Consequently, teacher professionalism is understood as an evolving formation in which identity, authority, and practice are continuously negotiated within AI-mediated environments.

Second, digital learning ecology theory conceptualises education as a complex adaptive system in which learning emerges through nonlinear interactions among human actors, technological infrastructures, and socio-institutional contexts (Fahrni et al., 2025; Sato et al., 2025; Costa et al., 2025; Bouakaz & Khalid, 2025). Within this ecology, AI operates as a semi-autonomous epistemic agent that redistributes cognitive processes across human and machine actors (Ali et al., 2024; Cukurova, 2025). This

redistribution gives rise to distributed professionalism, in which pedagogical and cognitive functions are co-produced within socio-technical networks (Fahrni et al., 2025; Sidorkin, 2025).

Third, this study critically interrogates continued reliance on the Technological Pedagogical Content Knowledge (TPACK) framework. Although TPACK has been instrumental in conceptualising the integration of technology, pedagogy, and content knowledge, it remains oriented toward the use of technology as an instructional resource (Jiménez Sierra et al., 2023; Aldemir et al., 2025). In the era of AI, this perspective is insufficient because AI also participates in knowledge production and redistributes epistemic authority (Oved & Alt, 2025; (Tan et al., 2025; Cukurova, 2025). Accordingly, this study advances a shift beyond integrative knowledge models toward an epistemically grounded and reflexive conception of professional capacity.

Fourth, augmented intelligence theory conceptualises AI as a system that can extend human cognitive capacity through collaborative interaction (Markauskaite et al., 2022; Cukurova, 2025; (Giannakos et al., 2025). However, this study problematizes the assumption that AI only augments professional capacity by foregrounding its potential to substitute, reconfigure, or redistribute core pedagogical functions (Sidorkin, 2025; Fu & Weng, 2024). This perspective introduces critical questions concerning professional autonomy, agency, and the boundaries between human and machine cognition (Krakowski, 2025; Zhai et al., 2021).

Finally, complexity and adaptive systems theory reconceptualise education as a nonlinear, emergent, and self-organising system characterised by feedback, interaction, and adaptive change. Within such environments, teacher professionalism is best understood as adaptive professionalism (Cheng et al., 2025), defined by the capacity to navigate uncertainty, respond to emerging conditions, and enact context-sensitive practice within dynamic socio-technical systems (Sato et al., 2025; Diamond & Bulfin, 2025; Mardahl-Hansen & Højholt, 2025; Dahl, 2024; Exley et al., 2025).

Taken together, these theoretical perspectives converge to support a co-evolutionary and socio-technical understanding of teacher professionalism, in which professional practice is continuously shaped through interaction with AI as both a structural and an epistemic force. This integrative framework provides the conceptual foundation for analysing how teacher professionalism

is reconstituted in the context of AI-driven educational transformation.

3. Relationship Between Concepts

The relationship between teacher professionalism and AI disruption is neither linear nor unidirectional; rather, it is dialectical and co-evolutionary, shaped through continuous interaction within dynamic socio-technical systems (Floridi et al., 2018; Krakowski, 2025). This study departs from deterministic perspectives that position technology as an exogenous force acting upon education, and instead conceptualises AI and teacher professionalism as mutually constitutive elements that co-produce the conditions of teaching and learning.

Within this framework, AI is conceptualised as an epistemic structural force that reconfigures established configurations of knowledge, authority, and pedagogical practice (Nguyen et al., 2025). Generative AI, in particular, changes how knowledge is produced, validated, and circulated, thereby challenging teachers' traditional position as primary knowledge holders (Jose et al., 2025; Wang et al., 2024; Bouakaz & Khalid, 2025). Consequently, the locus of expertise shifts from possession of information toward the capacity to critically interpret, validate, and contextualise machine-generated outputs.

In response to this disruption, teacher professionalism evolves as an adaptive and reflexive capacity rather than a fixed set of competencies (Veldhuis et al., 2025). This transformation entails a reorientation of professional identity and practice, in which teachers assume the role of epistemic mediators at the interface between human judgment and algorithmically generated knowledge (El-Soussi, 2025; Juutilainen et al., 2024; Jose et al., 2025). Such mediation requires technical proficiency, critical reflexivity, ethical awareness, and contextual sensitivity.

The interaction between AI and teacher professionalism gives rise to co-evolutionary dynamics that redefine educational practice (Mimoudi & Mokhtari, 2025; Feng et al., 2025). First, an epistemic shift redistributes cognitive authority between human and machine actors (Krakowski, 2025; Jose et al., 2025). Second, professional reconfiguration continuously renegotiates teacher identity, agency, and practice (Li, 2023; Juutilainen et al., 2024). Third, normative negotiation mediates tensions between technological efficiency and ethical-humanistic values (Fu & Weng, 2024).

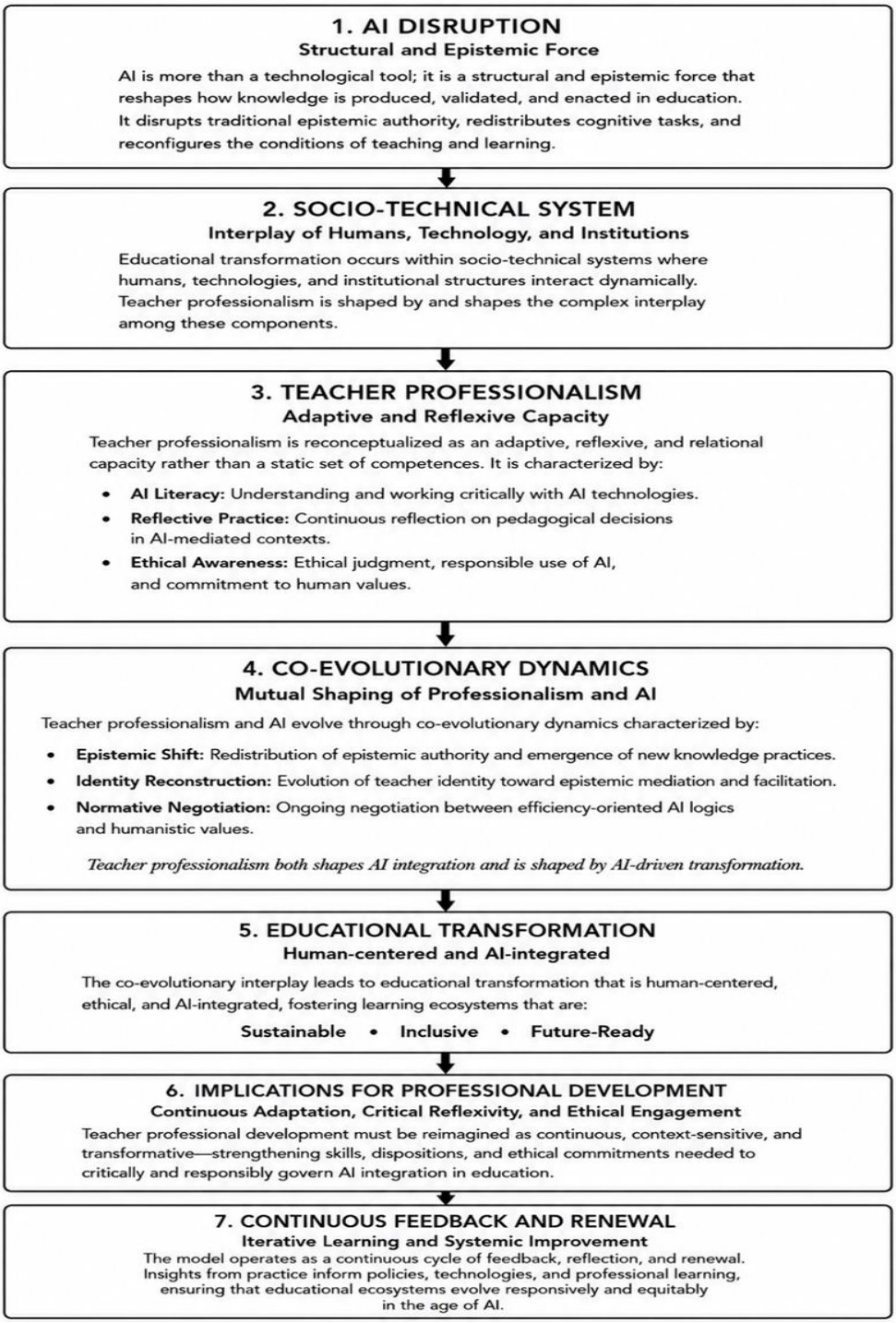
Crucially, these dynamics indicate that teacher professionalism is not merely shaped by AI disruption, but actively shapes the trajectory of technological integration itself. In this sense, professionalism functions as a critical site of governance, determining whether AI is enacted as a tool of instrumental efficiency or as a resource for human-centred educational transformation. Taken together, this relational perspective reinforces the central argument of this study: that the future of education is not defined solely by the presence of artificial intelligence, but by the capacity of teacher professionalism to critically and reflexively mediate its integration into complex socio-technical systems.

4 . Research Positioning

A systematic review of the literature published between 2015 and 2025 reveals a clear, dominant research trajectory characterised by an intensified focus on technology integration and the proliferation of AI applications in education. This body of work has significantly advanced the understanding of instructional innovation, data-driven learning, and adaptive educational systems. However, despite this rapid expansion, the literature remains disproportionately technology-centric, often privileging functional and instrumental perspectives over deeper conceptual inquiry into the human and professional dimensions of educational transformation.

More specifically, existing scholarship tends to frame AI primarily as a tool for enhancing instructional efficiency, while comparatively neglecting its implications for reconfiguring teacher professionalism as a multidimensional and adaptive construct. As a result, the conceptualisation of professionalism remains fragmented, with studies frequently isolating dimensions such as competence, pedagogy, and technology rather than examining their systemic interdependence within socio-technical contexts.

This fragmentation is further exacerbated by the limited integration of interdisciplinary theoretical perspectives, leading to the absence of a comprehensive conceptual framework capable of explaining how teacher professionalism evolves in response to AI as an epistemic-based structural force. Consequently, current research provides only a partial understanding of the dynamic interplay between technological transformation and *Figure 1: Co-x*


1. AI DISRUPTION
Structural and Epistemic Force
AI is more than a technological tool; it is a structural and epistemic force that reshapes how knowledge is produced, validated, and enacted in education. It disrupts traditional epistemic authority, redistributes cognitive tasks, and reconfigures the conditions of teaching and learning.
2. SOCIO-TECHNICAL SYSTEM
Interplay of Humans, Technology, and Institutions
Educational transformation occurs within socio-technical systems where humans, technologies, and institutional structures interact dynamically. Teacher professionalism is shaped by and shapes the complex interplay among these components.
3. TEACHER PROFESSIONALISM
Adaptive and Reflexive Capacity
Teacher professionalism is reconceptualized as an adaptive, reflexive, and relational capacity rather than a static set of competences. It is characterized by:
AI Literacy: Understanding and working critically with AI technologies.
Reflective Practice: Continuous reflection on pedagogical decisions in AI-mediated contexts.
Ethical Awareness: Ethical judgment, responsible use of AI, and commitment to human values.
4. CO-EVOLUTIONARY DYNAMICS
Mutual Shaping of Professionalism and AI
Teacher professionalism and AI evolve through co-evolutionary dynamics characterized by:
Epistemic Shift: Redistribution of epistemic authority and emergence of new knowledge practices.
Identity Reconstruction: Evolution of teacher identity toward epistemic mediation and facilitation.
Normative Negotiation: Ongoing negotiation between efficiency-oriented AI logics and humanistic values.
Teacher professionalism both shapes AI integration and is shaped by AI-driven transformation.
5. EDUCATIONAL TRANSFORMATION
Human-centered and AI-integrated
The co-evolutionary interplay leads to educational transformation that is human-centered, ethical, and AI-integrated, fostering learning ecosystems that are:
Sustainable • Inclusive • Future-Ready
6. IMPLICATIONS FOR PROFESSIONAL DEVELOPMENT
Continuous Adaptation, Critical Reflexivity, and Ethical Engagement
Teacher professional development must be reimagined as continuous, context-sensitive, and transformative—strengthening skills, dispositions, and ethical commitments needed to critically and responsibly govern AI integration in education.
7. CONTINUOUS FEEDBACK AND RENEWAL
Iterative Learning and Systemic Improvement
The model operates as a continuous cycle of feedback, reflection, and renewal. Insights from practice inform policies, technologies, and professional learning, ensuring that educational ecosystems evolve responsively and equitably in the age of AI.

In response to these limitations, this study advances a conceptually integrative and theoretically grounded framework that reconceptualises teacher professionalism as an adaptive, reflexive, and relational construct embedded within AI-driven socio-technical systems. Moving beyond competency-based and instrumental perspectives, this framework foregrounds epistemic agency, ethical judgment, and contextual responsiveness as core dimensions of professional practice.

This study makes a distinct theoretical contribution by shifting the analytical focus from knowledge integration toward a co-evolutionary understanding of teacher professionalism and AI disruption. In doing so, it conceptualises professionalism not as an individual attribute but as a dynamic capacity co-constituted through interaction with AI as a generative system, thereby extending existing theoretical models in the field. Accordingly, the primary contribution of this study lies in the development of a co-evolutionary conceptual framework that explains how teacher professionalism both shapes and is shaped by AI-driven transformation.

This framework not only advances theoretical understanding but also provides a foundation for rethinking sustainable professional development in the context of intelligent technologies, emphasising the need for continuous adaptation, critical reflexivity, and ethical engagement.

To synthesise the study's theoretical positioning and thematic findings, Figure 1 presents a coevolutionary framework that illustrates how AI-driven disruption reshapes teacher professionalism within socio-technical educational systems. The framework highlights the dynamic interplay between epistemic transformation, professional adaptation, and ethical engagement in AI-mediated education.

As illustrated in Figure 1, teacher professionalism is conceptualised not as a fixed set of competencies but as an adaptive, reflexive, and relational capacity that continuously evolves through interaction with AI-driven educational transformation. The framework, therefore, provides an integrative foundation for understanding sustainable, human-centred professional development in the context of intelligent technologies.

METHODOLOGY

This study employs a systematic literature review (SLR) to provide a rigorous, transparent, and replicable synthesis of research on teacher

professionalism in the context of AI. The SLR approach is particularly appropriate for integrating fragmented and interdisciplinary scholarship while enabling the development of theoretically grounded insights. To ensure methodological rigour, the review follows the PRISMA 2020 guidelines, which provide a structured framework for identification, screening, eligibility assessment, and inclusion of studies.

1 Data Sources and Search Strategy

This study employed a systematic literature review (SLR) to examine the evolving relationship between teacher professionalism and artificial intelligence (AI) within contemporary educational contexts. To ensure methodological rigour, transparency, and replicability, the review process followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA 2020) framework.

The literature search was conducted using two internationally recognised academic databases, namely Scopus and Web of Science, selected for their extensive coverage of high-impact, peer-reviewed scholarly publications across the fields of education, educational technology, and artificial intelligence. The search scope was restricted to studies published between 2015 to 2025 to capture the most recent developments in AI-driven educational transformation, particularly the emergence of generative AI and its implications for teacher professionalism.

To maximise both analytical relevance and interdisciplinary breadth, the search strategy utilised combinations of keywords and Boolean operators. The search strings incorporated terms related to teacher professionalism (e.g., “teacher professionalism,” “teacher identity,” and “teacher competence”), artificial intelligence (e.g., “artificial intelligence,” “AI,” “machine learning,” and “generative AI”), and educational contexts (e.g., “education,” “teaching,” and “learning”). The search process was conducted iteratively, allowing refinement of keyword combinations and screening procedures to ensure conceptual alignment with the objectives of the review. This iterative strategy enabled the identification of studies that not only addressed AI integration in education but also critically engaged with issues of professional identity, adaptive professionalism, epistemic agency, pedagogical transformation, and socio-technical change within educational systems.

2 Study Selection and Inclusion Criteria

The study selection process followed a PRISMA-guided procedure to ensure systematic identification, screening, eligibility assessment, and inclusion of relevant literature. The initial database search yielded 320 records from Scopus and Web of Science. After 10 duplicate entries were removed, 310 studies remained for screening. During title and abstract screening, 200 studies were excluded due to limited relevance. Subsequently, 110 full-text articles were assessed for eligibility, resulting in the exclusion of 65 studies that did not sufficiently address teacher professionalism in AI-mediated educational contexts or lacked substantial conceptual contribution. Ultimately, 45 peer-reviewed journal articles were retained for the final qualitative synthesis.

The inclusion criteria for the primary analytical corpus were defined as follows: (1) peer-reviewed journal articles; (2) publications indexed in Scopus and/or Web of Science; (3) studies published between 2015 - 2025; (4) studies focusing on artificial intelligence in educational settings; and (5) explicit engagement with teacher professionalism, professional identity, professional development, adaptive professionalism, or related pedagogical transformation. Studies were excluded if they focused solely on technical system development without educational or professional relevance, lacked substantive engagement with teacher-related dimensions, or provided insufficient theoretical or conceptual contribution to the thematic synthesis.

3 Study Selection Process (PRISMA Flow)

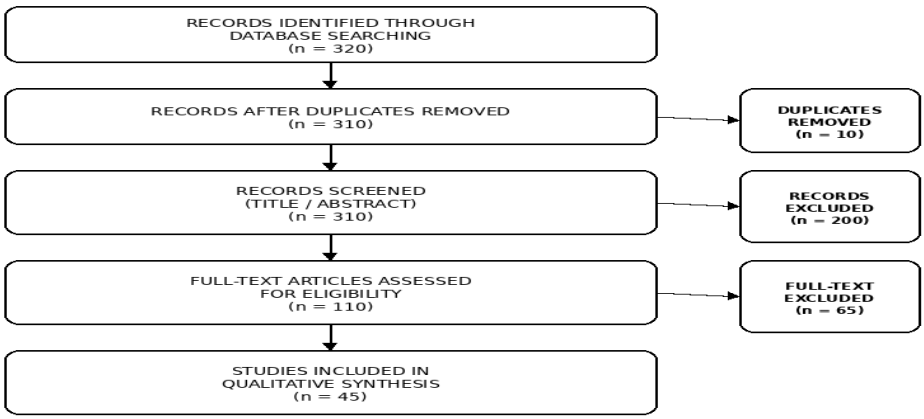


Figure 2. PRISMA flow diagram

Only the 45 sources meeting all inclusion criteria were retained. Policy

reports, general technology studies, methodological references, and other sources not directly aligned with teacher professionalism in AI-mediated education were excluded from the analytical corpus, the body citations, and the reference list.

The study selection process followed the PRISMA framework and is summarised in Figure 2. In the identification stage, 320 records were retrieved from the selected databases. After removing 10 duplicate records, 310 studies remained for screening. During title and abstract screening, 200 studies were excluded because they did not align with the review's focus.

Subsequently, 110 full-text articles were assessed for eligibility. Of these, 65 studies were excluded for reasons including insufficient relevance to teacher professionalism, lack of focus on AI in education, and limited conceptual contribution. Ultimately, 45 peer-reviewed journal articles were retained for the final qualitative synthesis. These 45 studies constitute the complete review corpus and are listed in Appendix A.

As illustrated in Figure 2, the study selection process followed a systematic and transparent procedure, thereby strengthening the rigour and reproducibility of the review.

4 Corpus Definition

To strengthen analytical rigour and conceptual clarity, the corpus was defined as a single set of 45 peer-reviewed journal articles indexed in Scopus and/or Web of Science. These studies constitute the empirical and conceptual foundation of the thematic synthesis and directly informed the coding structure, thematic categories, and co-evolutionary framework presented in this review.

The corpus was constructed through the PRISMA-guided selection process and included only studies with substantial relevance to teacher professionalism and AI-mediated educational transformation. The retained studies provided direct empirical, theoretical, or systematic insights into adaptive professionalism, epistemic agency, pedagogical transformation, ethical engagement, and socio-technical change. No separate supporting literature category was retained in the final manuscript. Sources with indirect, contextual, or merely technological relevance were removed to preserve alignment between the abstract, the analytical corpus, the in-text citations, the reference list, and Appendix A.

This single-corpus approach strengthens validity, coherence, and traceability by ensuring that each cited source directly contributes to the research questions and the thematic synthesis.

5 Quality Appraisal of Included Studies

Following the study selection process, a systematic quality appraisal was conducted to ensure the methodological rigour, analytical reliability, and conceptual relevance of the included studies before thematic synthesis. The appraisal process was designed not as a mechanism for excluding studies, but as a means of informing the interpretive weighting and analytical integration of findings within the review.

The quality assessment primarily employed the Mixed Methods Appraisal Tool (MMAT), widely recognised for evaluating the methodological quality of heterogeneous research designs in systematic reviews. Given the interdisciplinary nature of the reviewed literature, the MMAT framework was applied flexibly across empirical, conceptual, and review-based studies to maintain consistency while accommodating methodological diversity. Each study was evaluated according to five analytical dimensions: (1) clarity and relevance of the research focus; (2) appropriateness of the methodological design; (3) adequacy and credibility of evidence or data sources; (4) rigour and coherence of analytical procedures; and (5) consistency between findings, interpretation, and conclusions.

Based on these dimensions, studies were categorised into three interpretive quality levels: high, medium, and low. The appraisal results indicate that the majority of studies within the primary analytical corpus demonstrated high methodological and conceptual quality, thereby strengthening the robustness, credibility, and analytical reliability of the thematic synthesis. All 45 included studies were appraised using the same five analytical dimensions, with appropriate interpretation across empirical, conceptual, and review-based designs. A summary of the appraisal framework is presented in Table 1.

Importantly, no studies were excluded solely on the basis of quality appraisal. Instead, the assessment informed the analytical weighting, interpretive emphasis, and conceptual integration of findings throughout the thematic synthesis process. A detailed source-level appraisal of the full review corpus is provided in Appendix A.

Table 1. MMAT-based Quality Appraisal Framework Applied in the Review

Criterion	Guiding Appraisal Question	Interpretive Purpose in This Review
Clarity of research focus	Does the study clearly articulate a well-defined research problem, objective, or analytical focus?	Ensures conceptual alignment with the core focus of this review on teacher professionalism and AI in education.
Methodological appropriateness	Is the selected research design appropriate and well-justified for addressing the stated research aims?	Establishes confidence in the validity and credibility of the study's methodological approach.
Adequacy of evidence	Are the data sources, materials, or empirical inputs sufficiently described, relevant, and appropriate?	Enables differentiation between empirically grounded studies and those with limited evidentiary support.
Analytical rigor	Is the data analysis process systematic, transparent, and logically developed?	Informs the weighting of findings during thematic synthesis and supports analytical reliability.
Coherence of conclusions	Do the interpretations and conclusions logically follow from the presented evidence?	Ensures that conceptual insights used in framework development are well-supported and credible.

6 Data Analysis

The study employed reflexive thematic analysis to examine patterns, conceptual relationships, and recurring interpretive dimensions across the selected literature. This analytical approach was considered appropriate because the reviewed studies encompassed diverse methodological orientations, conceptual perspectives, and disciplinary traditions related to teacher professionalism and AI. The coding process followed an iterative and reflexive analytical procedure consisting of multiple stages of open coding, category refinement, thematic abstraction, and conceptual integration. Initial coding focused on identifying recurring concepts associated with AI-mediated educational transformation, teacher identity, adaptive professionalism,

epistemic agency, ethical engagement, and socio-technical change. These codes were subsequently compared, reorganised, and consolidated into higher-order analytical categories and themes through continuous interpretive refinement. Throughout the analysis, iterative memoing and reflexive comparison were employed to strengthen conceptual coherence and analytical consistency across coding cycles. This process enabled the continuous re-evaluation of emerging interpretations, relationships between categories, and theoretical alignment between empirical findings and the broader conceptual framework of the study.

The resulting themes were synthesised into a structured coding framework (Table 2) that systematically connects initial codes, analytical categories, thematic dimensions, conceptual interpretations, and illustrative evidence from the primary analytical corpus. These themes were further integrated into a co-evolutionary conceptual framework (Figure 3), illustrating the dynamic interplay between AI-driven epistemic transformation and adaptive teacher professionalism within socio-technical educational systems. The thematic synthesis was based exclusively on the 45 studies listed in Appendix A, ensuring direct correspondence among the analysed corpus, the evidence reported in Table 2, and the final reference list.

3.7 Trustworthiness and Research Limitations

To strengthen the review's trustworthiness, several strategies were employed throughout the analytical process. First, methodological transparency was maintained through explicit documentation of the search strategy, inclusion criteria, corpus definition, coding procedures, and thematic development process. Second, analytical consistency was enhanced through iterative coding cycles, reflexive memoing, and continuous comparison between data, categories, and conceptual interpretations. Third, conceptual validation was supported through the systematic alignment between the primary analytical corpus, thematic synthesis (Table 2), and the co-evolutionary conceptual framework (Figure 3).

Maintaining a single 45-study corpus also strengthened analytical rigour by ensuring that contextual framing, theoretical development, thematic interpretation, and conclusions were all grounded in the same set of directly relevant peer-reviewed sources.

Nevertheless, several limitations should be acknowledged. First, the

review relied primarily on Scopus and Web of Science, which may have introduced database selection bias by potentially excluding relevant studies indexed elsewhere. Second, the focus on English-language peer-reviewed publications may have limited the inclusion of regionally specific or non-English scholarship, particularly from underrepresented educational contexts. Third, as the study employed a qualitative and interpretive synthesis, the findings remain dependent on the conceptual scope and reporting quality of the included literature rather than on causal or statistical generalisation. Despite these limitations, the study provides a theoretically grounded and methodologically rigorous synthesis that contributes to the emerging discourse on teacher professionalism and AI-mediated educational transformation.

FINDINGS (Thematic Synthesis)

1 Overview of the Thematic Structure

The findings of the thematic analysis reveal a systematic pattern of transformation in the conceptualisation of teacher professionalism amid AI-driven disruption. Based on the analysed literature, three interrelated core themes were identified, namely: (1) adaptive professionalism, (2) AI as a structural and epistemic force, and (3) the dialectical relationship between humans and AI. These themes do not operate in isolation; they form an integrated analytical framework that reflects the dynamic interplay between teacher professionalism and AI-driven educational transformation. This thematic structure serves as the foundation for the study's co-evolutionary framework (Feng et al., 2025). To provide a systematic representation of the thematic synthesis process, Table 2 presents the relationships between initial codes, categories, themes, conceptual interpretations, and supporting literature.

Table 2. Coding Structure & Thematic Development from the *Literature*

Initial Codes	Categories	Themes	Conceptual Interpretation	Illustrative Evidence (Scopus Only)
AI transforma-tion	Technological disruption	AI as a structural and epistemic force	AI reshapes knowledge production, authority, and pedagogical	Holmes & Tuomi (2022); Wang et al. (2024); Giannakos et al.

Initial Codes	Categories	Themes	Conceptual Interpretation	Illustrative Evidence (Scopus Only)
			systems	(2025)
Human–AI co-creation	Epistemic transformation	Redistribution of epistemic authority	Knowledge is increasingly distributed across human and technological actors	Zhai et al. (2021); Chen et al. (2020); Markauskaite et al. (2022)
Teacher interpreting AI outputs	Cognitive restructuring	Epistemic mediation	Teachers act as mediators who interpret and contextualise AI-generated knowledge	Cukurova (2025); Jose et al. (2025); Costa et al. (2025)
Teacher role shift	Professional identity	Professional reconfiguration	Teacher identity evolves toward epistemic mediation and facilitation	El-Soussi (2025); Dahl (2024); Juutilainen et al. (2024)
AI integration in pedagogy	Pedagogical transformation	Reflexive professionalism	Professional practice requires critical and reflective engagement with AI	Fahrni et al. (2025); Song et al. (2025); Tan et al. (2025)
Ethics, bias, and governance	Ethical dimension	Normative negotiation	Teachers navigate tensions between efficiency and humanistic values	Floridi (2025); Ali et al. (2024); Fu & Weng (2024)
AI-supported ecosystems	Learning ecology	Distributed professionalism	Professional practice is co-produced within socio-technical systems	Veldhuis et al. (2025); Sidorkin (2025); Zuo et al. (2025)
Continuous learning and adaptation	Professional development	Adaptive professionalism	Professionalism is dynamic and requires ongoing adaptation	Avidov-Ungar (2025); Boylan et al. (2023); Exley et al. (2025)

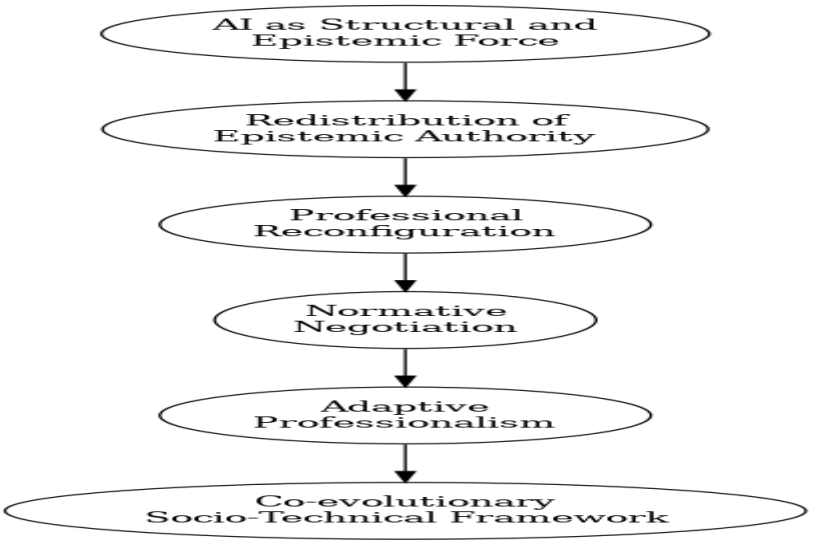


Figure 3 illustrates the progression from initial coding categories toward higher-order thematic synthesis, highlighting the co-evolutionary relationship between AI-driven epistemic transformation and adaptive teacher professionalism.

2 Adaptive Professionalism

The analysis demonstrates a fundamental shift from traditional, competency-based models of teacher professionalism toward a more adaptive, reflexive, and multidimensional construct. Rather than being defined by the accumulation of pedagogical and content knowledge, professionalism is increasingly understood as the capacity to navigate complexity, uncertainty, and continuous transformation within AI-mediated environments.

Three core dimensions characterise this emerging form of professionalism. First, epistemic agency, which refers to teachers' capacity to critically evaluate, interpret, and contextualise knowledge, particularly in relation to AI-generated outputs. Second, reflexive capacity, which enables teachers to continuously reassess their practices, assumptions, and pedagogical strategies in response to evolving technological conditions. Third, ethical engagement, which involves addressing issues such as algorithmic bias, data privacy, and the broader human implications of AI integration. Taken

together, these dimensions indicate that professionalism is no longer a stable attribute, but an adaptive capacity that evolves through interaction with socio-technical systems.

3 AI as Structural and Epistemic Disruption

The findings further reveal that AI operates not merely as a technological tool, but as a structural and epistemic force that fundamentally reconfigures the foundations of education (Holmes & Tuomi, 2022). This disruption manifests in three interrelated ways, grounded in complexity theory and socio-technical systems thinking. First, AI transforms the conditions of knowledge production by enabling algorithmic content generation that challenges traditional hierarchies of expertise (Giannakos et al., 2025; Wang et al., 2024). Second, it reshapes cognitive mediation as learning increasingly occurs through interaction with intelligent systems (Cukurova, 2025; Markauskaite et al., 2022). Third, AI redefines pedagogical practice by shifting teachers from information providers toward facilitators and epistemic mediators (Schmidt et al., 2025; Song et al., 2025; Stenberg et al., 2025).

Collectively, these transformations destabilise established configurations of authority, knowledge, and pedagogical practice, thereby necessitating a reconceptualisation of teacher professionalism that extends beyond technical competence.

4 Dialectical Human–AI Relationship

The interaction between teacher professionalism and AI disruption is best understood as a dialectical, co-evolutionary relationship rather than a unidirectional process of technological impact. The findings indicate that teachers and AI systems are engaged in a dynamic interplay in which each continuously shapes and is shaped by the other. Within this relationship, teachers are repositioned as epistemic mediators, responsible for interpreting, validating, and contextualising AI-generated knowledge. This role requires a shift from authority based on knowledge possession to authority grounded in critical judgment and contextual understanding.

At the same time, AI systems influence the structure of pedagogical practice, introducing new forms of distributed cognition and decision-making. This interaction gives rise to a hybrid knowledge environment, in

which knowledge is co-produced through human–machine collaboration. The thematic analysis culminates in a conceptual synthesis that integrates the relationships between AI disruption and teacher professionalism, as illustrated in Figure 3. This framework extends existing literature by providing a co-evolutionary perspective that captures the dynamic interplay between teacher professionalism and AI disruption.

4.5 Co-Evolutionary Dynamics

Further analysis identifies coevolutionary dynamics that shape the relationship between teacher professionalism and AI-driven disruption. First, an epistemic shift emerges, characterised by the redistribution of cognitive authority between human and machine actors. Second, a process of professional reconfiguration takes place, in which teacher identity, practice, and agency are continuously negotiated within AI-mediated environments. Third, a space of normative negotiation arises, reflecting the tensions between technological efficiency and the humanistic values of education. These dynamics indicate that teacher professionalism does not merely respond to technological change but actively participates in shaping the direction and meaning of educational transformation.

4.6 Conceptual Integration

The thematic synthesis in this study yields a conceptual integration that connects AI-driven disruption and teacher professionalism within a unified socio-technical framework. As presented in Figure 3, AI operates as a structural and epistemic force that shapes the conditions of education, while teacher professionalism functions as an adaptive capacity that mediates this transformation. The interaction between these elements generates co-evolutionary dynamics that ultimately determine the trajectory of educational transformation.

Taken together, Table 2 and Figure 3 provide analytically and conceptually complementary representations of the findings. Table 2 presents the analytical structure of the thematic synthesis process, whereas Figure 3 depicts the conceptual integration as a co-evolutionary framework. This approach moves beyond descriptive reporting toward critical conceptual synthesis, thereby offering a more comprehensive understanding of how teacher professionalism evolves in response to AI disruption in education.

DISCUSSION

1. Reconceptualizing Teacher Professionalism as an Adaptive Construct

The findings point to a fundamental reconceptualisation of teacher professionalism, shifting it from a static, competency-based paradigm toward a more adaptive, reflexive, and relational construct (Avidov-Ungar, 2025; Hordern, 2024). This aligns with scholarship that conceptualises professionalism as a dynamic and socially constructed process. The transformation extends beyond incremental professional development and is evident in everyday teaching practice, where teachers must respond to rapidly evolving technological conditions (Boylan et al., 2023).

Within this context, professionalism is no longer defined by knowledge alone, but by the capacity to critically interpret, evaluate, and reconfigure knowledge under conditions of uncertainty and algorithmic mediation. This shift highlights the growing importance of epistemic agency as a defining feature of professional expertise. This perspective is consistent with socio-technical theory, which emphasises that professional practices are continuously shaped through interactions between human actors and technological systems.

These findings suggest that teachers should be understood not as passive recipients of AI-driven change, but as active agents who continuously negotiate their professional roles in response to emerging technological conditions. This implies that teacher professionalism is inherently dynamic, shaped through ongoing interaction with socio-technical environments.

2. AI as Structural and Epistemic Disruption

The findings demonstrate that AI operates as a structural and epistemic force that reconfigures the conditions under which knowledge is produced, validated, and enacted (Holmes & Tuomi, 2022; Cukurova, 2025; Jose et al., 2025. This interpretation aligns with contemporary perspectives that conceptualise AI as an epistemic infrastructure shaping how knowledge is generated and validated.

From this perspective, knowledge production is no longer exclusively anchored in human expertise but increasingly co-constructed through algorithmic processes, data-driven systems, and machine learning models. This also resonates with contemporary perspectives that position AI as an

epistemic infrastructure shaping how knowledge is generated and validated. As a result, epistemic authority is redistributed across human and technological actors, introducing new forms of agency while raising critical questions about accountability, transparency, and legitimacy.

Within AI-mediated environments, teachers are no longer positioned as sole knowledge authorities but as epistemic mediators who critically interpret, evaluate, and contextualise AI-generated outputs. This shifts the analytical focus beyond technological adoption toward a deeper examination of how AI reshapes epistemic authority and professional practice. These findings further indicate that teachers act as active epistemic agents who engage with and negotiate the implications of AI-generated knowledge.

Beyond TPACK: Toward Epistemic and Reflexive Professionalism

The findings challenge the sufficiency of competency-based frameworks such as Technological Pedagogical Content Knowledge (TPACK) in capturing the complexity of AI-mediated educational environments (Jiménez Sierra et al., 2023). While TPACK provides a valuable foundation for understanding the integration of technology, pedagogy, and content, it remains insufficient for explaining the epistemic implications of generative AI (Aldemir et al., 2025; Oved & Alt, 2025).

This critique aligns with emerging scholarship that questions the adequacy of integrative knowledge frameworks for addressing the epistemic implications of AI-driven environments. Unlike conventional digital tools, AI operates as a generative, semi-autonomous system capable of producing and mediating knowledge beyond traditional pedagogical boundaries.

Accordingly, the findings indicate a shift toward epistemic and reflexive engagement with AI, in which teachers are required not only to use technological tools but also to critically evaluate, interpret, and govern their outputs. This suggests that teachers actively redefine their professional roles through reflexive engagement with AI, rather than merely applying technology within predefined pedagogical frameworks.

Co-Evolutionary Dynamics of Teacher Professionalism and AI

The findings reveal that the relationship between teacher professionalism and AI disruption is inherently co-evolutionary rather than deterministic. This co-evolutionary perspective is grounded in complexity

theory and socio-technical systems thinking, which emphasise mutual adaptation between human agency and technological structures (Feng et al., 2025).

Three interrelated dynamics characterise this relationship. First, an epistemic shift occurs, marked by the redistribution of cognitive authority between human and machine actors. Second, a process of professional reconfiguration emerges, in which teacher identity, agency, and practice are continuously renegotiated within AI-mediated environments. Third, the interaction generates a space of normative negotiation, where tensions between technological efficiency and humanistic values must be actively mediated. These dynamics indicate that teachers are not passive recipients of technological transformation, but key actors who actively shape its trajectory within educational practice. Such dynamics reflect the recursive relationship between structure and agency, as emphasized in contemporary socio-technical frameworks.

Integrative Synthesis

Taken together, the findings indicate that teacher professionalism should be understood not as a residual outcome of technological change, but as a constitutive force that shapes the direction and meaning of AI integration in education. This interpretation aligns with integrative socio-technical frameworks that conceptualise transformation as an emergent and co-constructed process.

Through this lens, teachers are positioned as epistemic mediators who interpret, negotiate, and govern the integration of AI within pedagogical practice. Teacher professionalism thus functions as an active and generative force within AI-mediated educational transformation. Importantly, these findings have global relevance, as education systems worldwide are increasingly confronted with similar challenges related to AI integration, professional adaptation, and ethical governance. This synthesis provides a coherent foundation for understanding the dynamic interplay between human agency and technological transformation and sets the stage for the theoretical and practical implications that follow.

IMPLICATIONS

6.1 Theoretical Implications

This study advances the theoretical understanding of teacher professionalism by repositioning it from a competency-based framework toward an adaptive, multidimensional, and relational construct embedded within AI-driven socio-technical systems. It demonstrates that professionalism cannot be adequately captured through static models of knowledge and skills alone. By incorporating AI as a structural and epistemic force, this study extends professional transformation theory beyond its traditional scope. Technology is not treated as an external driver, but as a constitutive element that reshapes the conditions under which professionalism is enacted.

The findings also challenge dominant frameworks such as TPACK, which remain limited in addressing the epistemic implications of AI. Instead, this study highlights the need for a shift toward epistemic and reflexive professional capacity, integrating cognitive, ethical, and socio-technical dimensions of practice. Furthermore, the study contributes to the emerging discourse on digital learning ecology by advancing the concept of distributed professionalism, in which professional practice is co-produced through interactions between human actors and intelligent technologies. Taken together, these contributions offer a co-evolutionary and socio-technical framework that provides a more robust foundation for understanding teacher professionalism in the context of AI-driven transformation.

6.2 Practical Implications

From a practical perspective, the findings highlight that professional development must move beyond technical skill acquisition to cultivate epistemic agency, critical reflexivity, and ethical judgment.

Teachers are required not only to use AI tools but to interpret, validate, and contextualise AI-generated knowledge within their practice. For educational institutions, this implies the need for systemic and continuous professional development strategies that integrate AI literacy, pedagogical innovation, and ethical awareness. Short-term, skill-based training is insufficient to address the complexity of AI-mediated environments.

For policymakers, the findings emphasise the importance of adopting a

holistic approach that aligns technological innovation with sustainable teacher development. Policies must support advanced digital literacy, ethical governance of AI, and equitable access to professional learning opportunities. More broadly, the findings suggest that the success of educational transformation in the age of AI is determined not by technological adoption alone, but by teachers' and institutions' capacity to engage with technology critically, contextually, and ethically.

6.3 Limitations of the Review

Several limitations should be acknowledged. First, the review was restricted to English-language publications in Scopus and Web of Science, and to peer-reviewed journal articles published between 2015 and 2025. While these criteria ensured quality and focus, they may have excluded relevant regional and non-English studies. Second, the study employs a qualitative, interpretive synthesis. Its contribution lies in conceptual integration rather than in estimating causal relationships. As such, the findings remain dependent on the scope and quality of reporting in the included literature. These limitations highlight the need for broader empirical validation and more diverse contextual exploration in future research.

FUTURE RESEARCH DIRECTIONS

The findings of this study reveal that, despite the rapid expansion of research on AI in education, the integration of teacher professionalism as a multidimensional and adaptive construct remains conceptually underdeveloped and empirically limited. Accordingly, future research must move beyond descriptive and technology-centric approaches toward more critical, integrative, and theoretically grounded investigations of the evolving relationship between teacher professionalism and AI-driven transformation.

First, there is a pressing need to advance research that examines teacher professionalism from a deeper epistemological perspective. Existing studies predominantly focus on competencies and technical skills, with limited attention to how AI reshapes the ways teachers construct, validate, and negotiate knowledge. Future research should therefore conceptualise professionalism as an epistemic construct, particularly in relation to the interaction between human cognition and generative AI systems. Such an approach is essential for understanding teachers' roles as epistemic mediators

in increasingly complex learning environments.

Second, future studies must address the persistent conceptual fragmentation within the literature. Current research often isolates dimensions such as pedagogy, competence, and technology, thereby failing to capture their systemic interdependence. There is a need for the development of integrative conceptual models that synthesise these dimensions within a unified analytical framework, potentially drawing on perspectives such as complex adaptive systems and socio-technical systems theory (Sato et al., 2025).

Third, the field requires more robust empirical validation of conceptual frameworks. Much of the existing research remains theoretical or limited in scope, constraining its ability to capture the dynamic and long-term evolution of teacher professionalism. Future research should employ longitudinal designs and mixed-method approaches to examine how professionalism develops over time in response to AI integration across diverse educational contexts.

Fourth, future research should address the significant imbalance in the literature regarding geographical and contextual factors. The majority of existing studies are concentrated in technologically advanced contexts, leaving a critical gap in understanding how AI disruption unfolds in developing regions. Research focusing on contexts such as Indonesia is particularly important, as it can reveal how factors such as digital inequality, institutional capacity, and policy environments shape the trajectory of professional transformation.

Fifth, there is a need to deepen inquiry into the ethical dimensions of teacher professionalism in AI-mediated education. While issues such as algorithmic bias, data privacy, and the potential dehumanisation of learning have been widely acknowledged, the literature lacks operational and context-sensitive ethical frameworks. Future studies should investigate how teachers develop ethical competencies and how humanistic values can be systematically integrated into AI-driven pedagogical practice.

Sixth, future research should critically examine the assumptions underlying the augmented intelligence paradigm, which positions AI primarily as a tool for enhancing human capacity. The findings of this study suggest that AI not only augments but may also substitute or reconfigure core

pedagogical functions. This raises important questions regarding professional autonomy, identity, and the boundaries between human and machine cognition, which require deeper theoretical and empirical exploration.

Finally, future research must move toward a more transformative and reflexive research agenda, shifting the focus from "how AI is used in education" to more fundamental questions concerning how AI reshapes the meaning of teaching, learning, and professionalism itself. In doing so, research should not only analyse technological change but also examine how teachers can actively shape, negotiate, and govern the direction of AI-driven transformation.

CONCLUSION

This study demonstrates that AI represents a structural transformation that reconfigures how knowledge is produced and enacted within educational systems. The findings affirm that teacher professionalism and AI are dynamically intertwined, characterised by shifts in epistemic authority, professional identity, and normative practice. These transformations reposition teachers as active epistemic mediators who interpret, validate, and contextualise knowledge within increasingly complex educational environments.

Ultimately, this study demonstrates that the future of education will not be determined solely by the sophistication of technological systems, but by teachers' capacity to govern their integration into pedagogical practice critically and ethically. In this sense, the future of education will depend not on whether AI becomes more intelligent, but on whether educational systems remain sufficiently human to govern it wisely. This reconceptualisation fundamentally reshapes how teacher professionalism is understood in the age of intelligent technologies.

REFERENCES

Aldemir, T., Bicer, A., Kilinc, S., Moon, J., & Kwok, M. (2025). Exploring emergent AI-TPACK competencies in a two-week AI literacy module for preservice teachers. *Teaching and Teacher Education*, *168*(September), 105231. https://doi.org/10.1016/j.tate.2025.105231

Ali, O., Murray, P. A., Momin, M., Dwivedi, Y. K., & Malik, T. (2024). Technological Forecasting & Social Change: The effects of artificial intelligence applications in educational settings : Challenges and strategies. *Technological Forecasting & Social Change*, *199*(June 2023),

123076. https://doi.org/10.1016/j.techfore.2023.123076

Avidov-Ungar, O. (2025). Empowering Teacher Professionalism Through Personalised Continuing Professional Learning: A Systematic Literature Review Using a Multidimensional Approach to Self-Assessment and Growth. *Education Sciences*, *15*(12), 158–173. https://doi.org/10.3390/educsci15121686

Bouakaz, L., & Khalid, S. (2025). AI in education: a sociological exploration of technology in learning environments. *Frontiers in Education*, *10*(November), 1–9. https://doi.org/10.3389/feduc.2025.1700876

Boylan, M., Adams, G., Perry, E., Booth, J., Boylan, M., Adams, G., Perry, E., Re-, J. B., Boylan, M., Adams, G., Perry, E., & Booth, J. (2023). Professional Development in Education Re-imagining transformative professional learning for critical teacher professionalism : a conceptual review teacher professionalism : a conceptual review. *Professional Development in Education*, *49*(4), 651–669. https://doi.org/10.1080/19415257.2022.2162566

Chen, X., Xie, H., Zou, D., & Hwang, G. (2020). Computers and Education : Artificial Intelligence Application and Theory Gaps during the Rise of Artificial Intelligence in Education. *Computers and Education: Artificial Intelligence*, *1*(July), 100002. https://doi.org/10.1016/j.caeai.2020.100002

Cheng, Y., Fan, Y., Li, X., Chen, G., Gašević, D., & Swiecki, Z. (2025). Asking the right questions of generative artificial intelligence improves writing performance. *Computers and Education: Artificial Intelligence*, *8*(October 2024), 100374. https://doi.org/10.1016/j.caeai.2025.100374

Costa, C., Murphy, M., & Costa, C. (2025). Generative artificial intelligence in education : (what) are we thinking ? *Learning, Media and Technology*, *9884*, 1–12. https://doi.org/10.1080/17439884.2025.2518258

Creely, E., & Carabott, K. (2025). Teaching and learning with AI: an Integrated AI-Oriented Pedagogical Model. *Australian Educational Researcher*, *52*(6), 4633–4654. https://doi.org/10.1007/s13384-025-00913-6

Cukurova, M. (2025). The interplay of learning, analytics and artificial intelligence in education: A vision for hybrid intelligence. *British Journal of Educational Technology*, *56*(2), 469–488. https://doi.org/10.1111/bjet.13514

Dahl, K. K. B. (2024). 'Doubly situated teacher professionalism': School culture, personal narrations and becoming a teacher in Danish schools. *International Journal of Educational Development*, *107*(March), 103047. https://doi.org/10.1016/j.ijedudev.2024.103047

Diamond, F., & Bulfin, S. (2025). Care of the profession: teacher professionalism and learning beyond performance and compliance. *Pedagogy, Culture and Society*, *33*(2), 503–521. https://doi.org/10.1080/

14681366.2023.2239820

El-Soussi, A. (2025). Teacher identity continuum: A framework for online shifts in teacher identity. *International Journal of Educational Research Open*, *8*(October 2024), 100411. https://doi.org/10.1016/j.ijedro.2024.100411

Exley, B., Hoyte, F., & Singh, P. (2025). When the curriculum demands personalisation : adaptive professionalism, pre - packaged plans, and the teaching of phonics and spelling. *The Australian Journal of Language and Literacy*, *48*(2), 161–174. https://doi.org/10.1007/s44020-025-00083-z

Fahrni, D. D. D., Iten, G., Prasse, D., & Hascher, T. (2025). Teachers' practices in the use of digital technology to promote students' self-regulated learning and metacognition: A systematic review. *Teaching and Teacher Education*, *165*(June). https://doi.org/10.1016/j.tate.2025.105150

Feng, S., Zhang, H., & Gašević, D. (2025). Mapping the evolution of AI in education: Toward a co-adaptive and human-centred paradigm. *Computers and Education: Artificial Intelligence*, *9*(November). https://doi.org/10.1016/j.caeai.2025.100513

Fu, Y., & Weng, Z. (2024). Computers and Education : Artificial Intelligence Navigating the ethical terrain of AI in education : A systematic review on framing responsible human-centred AI practices. *Computers and Education: Artificial Intelligence*, *7*(September), 100306. https://doi.org/10.1016/j.caeai.2024.100306

Holmes, W., & Tuomi, I. (2022). State-of-the-art and practice in AI in education. *Wileyonlinelibrary.Com/Journal/Ejed*, 1–29. https://doi. org/10.1111 / ejed.12533

Hordern, J. (2024). Teacher professionalism, expertise and the jurisdictional struggle. *Education Inquiry*, *17*(1), 1–16. https://doi.org/10.1080/20004508 .2024.2308945

Jiménez Sierra, Á. A., Ortega Iglesias, J. M., Cabero-Almenara, J., & Palacios-Rodríguez, A. (2023). Development of the teacher's technological pedagogical content knowledge (TPACK) from the Lesson Study: A systematic review. *Frontiers in Education*, *8*(February). https://doi.org/10.3389/feduc.2023.1078913

Jose, B., Cleetus, A., Joseph, B., Joseph, L., Jose, B., & John, A. K. (2025). Epistemic authority and generative AI in learning spaces: rethinking knowledge in the algorithmic age. *Frontiers in Education*, *10*. https://doi.org/10.3389/feduc.2025.1647687

Juutilainen, M., Metsäpelto, R., & Mäensivu, M. (2024). Beginning student teachers ' agency in teacher identity negotiations. *Teacher Development*, *28*(2), 178–194. https://doi.org/10.1080/13664530.2023.2299683

Krakowski, S. (2025). Human-AI agency in the age of generative AI.

Information and Organisation, *35*(1), 100560. https://doi.org/10.1016/j.infoandorg.2025.100560

Li, X. (2023). A theoretical review on the interplay among EFL teachers' professional identity, agency, and positioning. *Heliyon*, *9*(4), e15510. https://doi.org/10.1016/j.heliyon.2023.e15510

Mardahl-Hansen, T. L., & Højholt, C. (2025). Situated Teacher Professionalism: Understanding the Exploratory and Collaborative Aspects of Teacher's Professionalism. *Integrative Psychological and Behavioural Science*, *59*(2), 1–19. https://doi.org/10.1007/s1224-025-09914-7

Mimoudi, A., & Mokhtari, K. (2025). ' AIA-PCEK ': A new framework for teaching with AI. *Cogent Education*, *12*(1). https://doi.org/10.1080/2331186X.2025.2563171

Nguyen, A., Kishore, S., Hong, Y., Qutab, S., & Dang, B. (2025). Generative Artificial Intelligence (AI) in education: from organising visions to official guidelines. *Information Technology and People*, *38*(8), 172–199. https://doi.org/10.1108/ITP-08-2024-1026

Oved, O., & Alt, D. (2025). Teachers' technological, pedagogical, and content knowledge (TPACK) as a precursor to their perceived adoption of educational AI tools for teaching. *Education and Information Technologies*, *30*(10), 14095–14121. https://doi.org/10.1007/s10639-025-13371-5

Sato, M., Ma, T., & Abbiss, J. (2025). Using complexity theory to understand teaching : Re-framing perspectives from preservice teachers. *Teaching and Teacher Education*, *156*(January), 104905. https://doi.org/10.1016/j.tate.2024.104905

Schmidt, D. A., Alboloushi, B., Thomas, A., & Magalhaes, R. (2025). Integrating artificial intelligence in higher education: perceptions, challenges, and strategies for academic innovation. *Computers and Education Open*, *9*(May), 100274. https://doi.org/10.1016/j.caeo.2025.100274

Shen, G., Wu, W., & Huang, W. (2025). Tackling the AI-Creativity paradox: How educational innovation and knowledge unlock positive-sum dynamics? *Journal of Innovation and Knowledge*, *10*(6), 100816. https://doi.org/10.1016/j.jik.2025.100816

Sidorkin, A. M. (2025). Extended executive cognition, a learning outcome for the AI age. *Computers and Education Open*, *9*(March), 100294. https://doi.org/10.1016/j.caeo.2025.100294

Song, Z., Qin, J., Jin, F., Cheung, W. M., & Lin, C. H. (2025). A case study of teachers' processes for integrating generative artificial intelligence and the factors influencing them. *Teaching and Teacher Education*, *165*(October 2024), 105157. https://doi.org/10.1016/j.tate.2025.105157

Stenberg, C., Åkerfeldt, A., Stenliden, L., & McGrath, C. (2025). Teaching with the algorithm : anticipations of AI in teacher education. *European Journal of Teacher Education*, *48*(5), 1056–1076. https://doi.org/10.1080/02619768.2025.2540793

Veldhuis, A., Lo, P. Y., Kenny, S., & Antle, A. N. (2025). International Journal of Child-Computer Interaction Critical Artificial Intelligence literacy : A scoping review and framework synthesis. *International Journal of Child-Computer Interaction*, *43*(October 2024), 100708. https://doi.org/10.1016/j.ijcci.2024.100708

Wang, S., Wang, F., Zhu, Z., Wang, J., Tran, T., & Du, Z. (2024). Artificial intelligence in education: A systematic literature review. *Expert Systems with Applications*, *252*(PA), 124167. https://doi.org/10.1016/j.eswa.2024.124167

Yang, Y., Kikas, E., Raija, H., Ga, D., Tommi, K., & Azevedo, R. (2025). *Computers and Education : Artificial Intelligence Towards responsible AI for education : Hybrid human-AI to confront the elephant in the room*. *9*(April). https://doi.org/10.1016/j.caeai.2025.100524

Zhai, X., Chu, X., Chai, C. S., Siu, M., Jong, Y., Istenic, A., Spector, M., Liu, J., Yuan, J., & Li, Y. (2021). A Review of Artificial Intelligence (AI) in Education from 2010 to 2020. *Hindawi Complexity*, *2021*. https://doi.org/10.1155/2021/8812542

Zuo, Z., Luo, Y., Yan, S., & Jiang, L. (2025). From Perception to Practice: Artificial Intelligence as a Pathway to Enhancing Digital Literacy in Higher Education Teaching. *Systems*, *13*(8), 1–22. https://doi.org/10.3390/systems13080664

AUTHOR PROFILE



Dr Hermania Bhoki completed her bachelor's degree in Catholic Philosophy of Religion at STFK Ledalero (1999–2003) and served as a CIJ Postulant and Novice advisor (2003–2005), later becoming deputy head of LITBANG at the Pastoral Centre of the Archdiocese of Ende (2005–2010). A regular author of the book IKUT JESUS since 2009, she began lecturing at Reinha Larantuka Pastoral College in 2011, earning a Master's in Education Management from Jakarta State University (015) and a Doctorate from Semarang State University (022). She became Principal of Reinha Larantuka Pastoral College in 2024 and published in journals including REINHA

and Sinta.

Sr. Skolastika Beding, CIJ, was born in Lamalera, Wulandoni District, 
Lembata Regency. After joining the Congregation of the Imitation of Jesus, she earned a bachelor's degree in Religious Education and a Master's (Licentiate) in Canon Law at the Pontifical Urbaniana University in Rome (2012–2018), then completed the Postulator Course at the Dicastery for the Causes of Saints in Rome (2018–2019). She joined Reinha Larantuka School of Pastoral Studies (STP Reinha Larantuka) as a lecturer, a role she continues today, and began doctoral studies at Widya Sasana Institute of Philosophy and Theology (STFT Widya Sasana) in 2023.

Benedikta Yosefina Kebingin was born in Lewoleba, Lembata 
Regency. After joining the Congregation of the Imitation of Jesus, she studied at STFK Pradnyawidya Yogyakarta (1991–1995), earned a Master's in Missiology at Pontifical Urbaniana University, Rome (1999–2002), and a Doctorate in Theology from STFT Widya Sasana Malang (2025). Her ministry includes serving as a CIJ formator, lecturer, Head of the Diploma Program in Catechetics at IPI Malang's Larantuka branch, CIJ General Council member, Deputy Superior General of CIJ, Chairperson of YAPERTHEL, Vice-Rector II of IKTL, and Chairperson of STP Reinha Larantuka. Elected Superior General of CIJ (2019–2025), she has authored three **books.**

CHAPTER VI

The Integrity Crisis of Academic Assessment Systems in the Era of Generative Artificial Intelligence: A Systematic Review of the Literature on the Validity and Credibility of Learning Evaluation

Sulistiyani and Parju
University of 17 Agustus 1945 Semarang, Indonesia

ABSTRAK

The purpose of this study is to examine the development of literature on the integrity crisis of academic assessment systems in the era of generative artificial intelligence (generative AI/GenAI), with a focus on threats to the construct validity and credibility of learning evaluations in higher education. This systematic review of the literature is structured according to PRISMA guidelines. It searches academic sources across Scopus, Web of Science, Crossref, and Google Scholar, using the keywords generative AI, large language model, academic integrity, assessment validity, and higher education, covering publications from 2013 to 2026. The initial search identified over 120 candidate articles across 10 research themes, which were then filtered based on title, abstract, theme relevance, and digital object identifier verification, resulting in 37 articles and core policy documents for in-depth synthesis. The synthesis results show that GenAI is dual in nature, expanding access to learning and personalization, but blurring the lines between legitimate learning assistance and cognitive substitution. Automated detection approaches have proven empirically inadequate, leading the literature to shift toward authentic assessment redesign, strengthening AI literacy, disclosure policies, and trust-based governance as more sustainable validity restoration strategies for higher education institutions in the future.

Keywords: generative artificial intelligence; academic integrity; assessment validity; large language model; higher education governance

INTRODUCTION

Since the public launch of a large language model (LLM) based on generative artificial intelligence in late 2022, higher education systems

worldwide have faced structural disruption in learning assessment practices. GenAI technology enables the production of academic essays, policy analyses, research reports, and literature syntheses in seconds with a level of coherence that resembles human writing (Kasneci et al., 2023). This change is not merely a technical one, but a fundamental shift in how academic knowledge is produced, assessed, and accounted for.

Conventional assessment systems, such as take-home essays, online assignments, and analytical reports, assume that academic products reflect students' individual cognitive processes. This assumption becomes fragile when GenAI can produce high-quality academic texts instantly, raising fundamental questions about whether assessments still measure competency mastery or instead measure students' ability to use technology (Kofinas et al., 2025; Swiecki et al., 2022). This situation gives rise to what can be called an epistemological crisis: uncertainty about the authenticity of knowledge production and the validity of the learning evaluation methods currently used in higher education.

Recent literature suggests that GenAI has an ambivalent role in higher education. On the one hand, this technology has the potential to expand access to learning, enhance educational personalization, and support broader exploration of knowledge (Farrelly & Baker, 2023; Tlili et al., 2023). A recent experimental meta-analysis even suggests that integrating ChatGPT purposefully into instructional design can positively affect student learning outcomes (Deng et al., 2025). However, GenAI's ability to automatically produce academic content also poses serious threats to academic integrity, evaluation fairness, and the authenticity of student work (Cotton et al., 2024; Eke, 2023). The initial response of some universities, which prohibited the use of AI in academic assignments, quickly shifted to a regulatory and integrative approach emphasizing AI literacy, transparency in technology use, and redesigned assessment methods (Perkins, 2023; Sullivan et al., 2023).

Despite this, the available literature remains fragmented. Some studies focus on the effectiveness and limitations of AI detection tools as a defense against academic integrity violations (Weber-Wulff et al., 2023). Another group of studies examines the redesign of authentic assessments through frameworks such as the AI Assessment Scale (AIAS), which openly and in stages places AI's involvement in task design (Furze et al., 2024; Perkins et al.,

2024, 2025). Meanwhile, a third group of literature highlights institutional governance, AI disclosure policies, and AI literacy as prerequisites for successful assessment reform (Chan, 2023; Gonsalves, 2024; Moorhouse et al., 2023; Ng et al., 2021, 2024). This fragmentation suggests that the literature is rich in topics, but has not been widely synthesised through an integrated lens of assessment validity and credibility.

To date, existing systematic reviews have tended to focus on academic integrity in general or on descriptive assessment transformations (Bittle & El Gayar, 2025; Xia et al., 2024) but have not explicitly integrated argument-based construct validity perspectives (Kane, 2013; Messick, 1995) with contemporary empirical findings on the failures of detection approaches and the successes of redesign approaches. To date, no systematic review has been found that specifically synthesises the empirical, conceptual, and policy literature on the assessment integrity crisis using the argument validity framework as the primary lens and explicitly following the PRISMA procedure. This gap makes this systematic review important.

This article's primary contributions lie in three areas. First, it synthesises the scattered literature on AI detection, assessment redesign, and governance into a more systematic and coherent synthesis. Second, it proposes an integrative conceptual model that situates the assessment validity crisis within the context of the antecedents, mediators, moderators, and consequences recurrently identified in the literature. Third, it argues that the future of academic integrity will not be determined solely by the sophistication of detection technologies, but rather by institutions' capacity to design assessment ecosystems that are authentic, transparent, and adaptive to human-AI collaboration. To achieve these goals, this article is structured into nine sections: introduction, research questions, methods, results, discussion, integrative conceptual model, implications, limitations, a further research agenda, and conclusion.

RESEARCH QUESTIONS

This study aims to answer five interconnected research questions. First, how have research trends on GenAI and academic assessment integrity evolved since the public launch of the LLM in 2022 through 2026? Second, what factors consistently emerge as antecedents, mediators, moderators, and

consequences in the relationship between GenAI use and academic assessment validity? Third, to what extent have automated detection-based approaches been empirically proven adequate as mechanisms for maintaining academic integrity? Fourth, what assessment redesign strategies and institutional governance are most widely recommended in the literature to address this validity crisis? Fifth, what research gaps remain, and what agendas should be prioritized in future research.

To formulate research questions systematically, this study used the SPIDER (Sample, Phenomenon of Interest, Design, Evaluation, Research type) framework, which is more appropriate for qualitative topics and education policy than the PICO framework commonly used in the health sector. Details of the SPIDER framework's application to this study are presented in Table 1.

Table 1. SPIDER Framework in Formulating Research Questions

Componen	Definition	Application in This Study
S, Sample	The group or context being studied	Students, lecturers, and higher education institutions
PI, Phenomenon of Interest	The phenomenon studied	The use of generative AI in learning assessment and evaluation
D, Design	Types of studies included	Empirical, conceptual studies, reviews, and official policy documents
E, Evaluation	Aspects evaluated	Construct validity, reliability, integrity, and governance of assessments
R, Research type	Type of research	Qualitative, quantitative, mixed, conceptual, and policy documents

METHOD

This study employed a systematic literature review approach, following the PRISMA reporting framework, to synthesize articles and policy documents relevant to academic assessment integrity in the GenAI era. The search protocol was not formally registered with a registry such as OSF Registries or

PROSPERO prior to the search, and this is explicitly stated as a methodological limitation.

The search was conducted thematically from June to July 2026 by searching Scopus, Web of Science, Crossref, and Google Scholar, accompanied by verification of digital object identifier links in each candidate article through direct checks on official publisher sites such as Elsevier, Springer, Taylor and Francis, Wiley, MDPI, and SAGE. Keywords were arranged into ten theme clusters, including generative AI and academic integrity, assessment validity and reliability, authentic assessment redesign, effectiveness of AI detection tools, institutional policies and governance, AI literacy, contractual plagiarism, trust-based governance, epistemological limits of learning assistance and cognitive substitution, and existing reviews of GenAI in higher education. Examples of Boolean strings used were generative AI OR ChatGPT OR large language model AND academic integrity OR assessment validity AND higher education, with the syntax adjusted for each indexing source.

The publication year range included was 2013 to 2026, given that some classical validity theory references were published before the GenAI era but remain conceptually relevant. While the majority of empirical and policy articles date from 2022 to 2026, since ChatGPT's public launch. Language selection was limited to English-language articles that had undergone peer review or were official policy documents from credible institutions such as UNESCO and TEQSA Australia.

Initial identification yielded approximately 128 candidate articles. After deduplication and metadata verification, 96 unique articles were obtained. Screening of titles and abstracts yielded fifty-eight relevant articles. Full-text eligibility assessment excluded articles with unverified DOIs, irrelevant topics, or preprints that had not yet been peer-reviewed. Thirty-seven articles and a core policy document served as the final synthesis corpus.

Inclusion criteria included articles discussing GenAI in relation to assessment or academic integrity in higher education, whether empirically, conceptually, or in official policy, and that had a verified DOI or an official link. Exclusion criteria included articles focusing on the technical aspects of the algorithm without an educational context, educational levels outside of higher education, non-academic opinions without substantive arguments, and unverified metadata.

Due to the heterogeneity of the corpus, quality assessment was conducted qualitatively based on publisher credibility, peer review status, and consistency of arguments with referenced evidence, in accordance with AACODS principles for grey literature, such as UNESCO and TEQSA publications. Data extraction used a synthesis matrix that included title, publisher, year, study type, methods, main findings, and antecedent, mediator, moderator, and consequence variables. This process resulted in descriptive and narrative thematic analyses that formulated a conceptual framework for the crisis and the restoration of validity in learning assessments.

RESULTS

1. Study Characteristics

Of the thirty-seven articles and core policy documents synthesised, the majority were published between 2023 and 2026, aligning with the pattern of disruption that emerged immediately following the public launch of ChatGPT in late 2022. A small number of references date from before 2022, particularly classic psychometric validity theory literature such as Messick (1995) and Kane (2013), which remains relevant as the conceptual foundation for this study's arguments. In terms of study type, the corpus comprises conceptual and theoretical studies, qualitative and quantitative empirical studies, systematic reviews and scoping reviews, and official policy documents from institutions such as UNESCO and TEQSA Australia.

The distribution of themes shows that four major clusters dominated the discussion: the validity and reliability of assessments in the AI era, the effectiveness and limitations of detection approaches, the redesign of authentic assessments, and institutional governance and AI literacy. This pattern indicates that academic discourse on GenAI and assessment integrity has moved beyond the binary debate over the permissibility of AI use toward a more technical and systemic discussion of how assessment design, policy, and literacy can be reconstructed to remain valid and credible amid the availability of generative technologies.

2. Findings Based on Theme

2.1 Assessment Construct Validity Crisis

The construct validity crisis is the most fundamental theme in this corpus of literature. According to the argument-based validity framework,

validity is not simply the nature of the instrument, but rather the strength of the argument linking assessment scores to claimed competency interpretations (Kane, 2013). This is an epistemological shift pioneered by Messick (1995), who asserted that validity should be assessed by the consequences of using the scores, not just by the accuracy of the measurement itself. Brennan (2013) adds a critical note that Kane's argumentative approach requires clarity regarding the claims of validated interpretations. This clarity becomes problematic when assessments confront the possibility of undetected AI involvement.

Swiecki et al. (2022) was one of the first studies to explicitly link the discourse of artificial intelligence to assessment theory, asserting that AI challenges the fundamental assumption that assessment products directly reflect individual student abilities. This argument is reinforced by the empirical findings of Kofinas et al. (2025), which show that lecturers across various study programs struggle to determine the extent of GenAI's involvement in students' academic output, rendering the validity of authentic assessments highly vulnerable to technological influences that are not readily apparent. A recent study by Ali and Maroulis (2026) reinforces this argument by concluding that GenAI's primary threat to validity lies not in cheating itself, but in construct-irrelevant variance, that is, the variation in scores that arises from differences in access to and skills in using AI, rather than differences in mastery of the competencies actually being measured.

These findings imply that the assessment integrity crisis in the GenAI era is, in fact, a validity crisis in the argumentative sense, not merely a moral crisis regarding student honesty. When an essay assignment no longer ensures that the resulting score reflects conceptual mastery, the chain of validity arguments from the score to the graduation decision or competency certification is broken, regardless of whether or not the student intended to cheat.

2.2 The Fragility of Detection-Based Approaches

A second prominent theme is the empirical fragility of automated detection-based approaches as the primary mechanism for safeguarding academic integrity. A comprehensive study by Weber Wulff et al. (2023) tested 14 AI-based text detection tools, including 12 public tools and Turnitin and PlagiarismCheck, and found that none achieved 80 per cent accuracy, while

only 5 exceeded 70 per cent. The study also revealed systematic bias in some detection tools, which tended to classify certain human writing, particularly that of non-native speakers, as AI-generated, potentially leading to erroneous cheating accusations and unfairly harming students.
Kirim masukan

These findings align with the arguments of Eke (2023) and Perkins (2023), who assert that an excessive focus on AI detection risks creating an adversarial relationship between institutions and students, without truly addressing the root causes of assessment validity issues. Cotton et al. (2024) also emphasise that detection approaches tend to be reactive and easily outdated as AI technology advances much faster than the adaptability of detection tools, making this strategy structurally unsuitable as a long-term solution to the academic integrity crisis.

This pattern suggests that arguments for shifting from a detection logic to a design logic are increasingly supported by robust empirical evidence rather than simply normative preferences. When detection instruments prove unreliable and potentially biased, institutions' reliance on them as the primary safeguard of academic integrity becomes an epistemologically fragile policy choice.

2.3 Redesigning Authentic Assessment and the AIAS Framework

In response to the fragility of detection approaches, the literature indicates a strong shift toward redesigning authentic assessments that openly accommodate AI. The most widely cited framework in this corpus is the Artificial Intelligence Assessment Scale (AIAS) developed by Perkins et al. (2024) and refined by Perkins et al. (2025), which proposes a tiered scale ranging from a complete ban on the use of AI to full collaboration between students and AI, with each tier having different implications for task design and assessment criteria. Furze et al. (2024) report a pilot implementation of this framework in a real-life learning context and find that assessment designs that explicitly acknowledge the involvement of AI help refocus assessment on the process of thinking and reflection, rather than solely on the final product.

Bearman et al. (2024) expand on this argument through the concept of evaluative judgment, namely the capacity of students to assess the quality of work, both their own and AI-generated, as a core competency that needs to be

deliberately developed amidst the availability of generative tools. This argument emphasises that AI should not be viewed as a threat to learning, but rather can be an object of learning itself, provided that assessments are designed to assess students' evaluative and reflective abilities regarding AI output. Corbin et al. (2025) add that higher education institutions still struggle to define clear boundaries regarding acceptable uses of AI, thus proposing a framework for classifying AI use based on learning objectives and stages of the academic process, rather than simply a binary prohibition or permission.k.

Lodge (2025) provides a key critique of the tendency of some institutions to implement blanket bans or unconditional permission policies for AI use, asserting that such approaches ignore the complexities of different pedagogical contexts and disciplines. This argument reinforces the conclusion that effective assessment redesign requires a contextual, multifaceted approach rather than a single, uniform policy across all study programs.

2.4 Governance, AI Literacy, and Disclosure Policies

The fourth theme highlights the importance of institutional governance, AI literacy, and disclosure policies as prerequisites for successful assessment reform. Moorhouse et al. (2023) analysed the official guidelines of several top-tier universities worldwide. They found significant variation in the clarity and comprehensiveness of GenAI policies, ranging from explicit prohibitions to full integration into the curriculum. Chan (2023) proposed a comprehensive AI policy framework encompassing pedagogical, governance, ethical, and operational dimensions as a model for institutions that lack formal policies.

At the level of trust and transparency, Gonsalves' (2024) study at a business school found that 74% of students failed to declare the use of AI despite the mandatory requirement. This finding indicates that administrative compliance mechanisms alone are insufficient to build a culture of integrity and need to be accompanied by strengthening trust and clarifying policy consequences. Similarly, Lancaster et al. (2025) highlighted that academic integrity policies formulated for conventional contractual plagiarism cases need to be fundamentally updated to be relevant to new modes of GenAI-based academic cheating, which are much more difficult to track than traditional assignment-taking practices.

In terms of literacy, Ng et al. (2021) conceptualise AI literacy as a multidimensional competency encompassing cognitive, affective, and ethical aspects, and Ng et al. (2024) developed a validated instrument to empirically measure it. Almatrafi et al.'s (2024) systematic review of AI literacy literature from 2019 to 2023 confirms that strengthening AI literacy in students and lecturers is a crucial prerequisite for effective disclosure policies and assessment redesign, as both rely on a shared understanding of when and how AI use is considered ethical.

At the global policy level, Miao and Holmes' (2023) UNESCO document provides the first international guidance on GenAI in education and research, emphasising the importance of a human-centred approach and rights-based governance in the adoption of generative technologies in the education sector. Miao and Cukurova's (2024) follow-up framework builds on this guidance by offering an AI competency framework for teachers and lecturers. At the national level, TEQSA Australia (2023, 2025) published two consecutive policy documents explicitly promoting assessment reform in the era of artificial intelligence, from the problem-diagnosis stage to practical implementation at the study-program level.

DISCUSSION

The findings of this systematic review indicate that the integrity crisis of academic assessment in the GenAI era is developing amidst two major, simultaneous currents: the push toward efficiency and personalisation of learning, and the pressing need to maintain the validity and credibility of learning evaluations. On the one hand, GenAI is seen as an instrument capable of expanding access to learning and supporting students' more personalised exploration of knowledge (Farrelly & Baker, 2023; Tlili et al., 2023), and has even been shown to have a positive effect on learning outcomes under certain conditions (Deng et al., 2025).

On the other hand, the greater the generative capacity of this technology, the more pressing the need to reaffirm assessment's role as a valid inferential instrument for academic decision-making (Kane, 2013; Kofinas et al., 2025). Therefore, the key issue is not simply whether GenAI is beneficial for learning, but rather how it is positioned within an assessment architecture

that maintains the chain of validity arguments from scores to competency interpretation.

In terms of theoretical stability, this study found that the literature on assessment validity in the AI era still relies heavily on Kane's (2013) argument validity framework, which builds on Messick's (1995) conceptual foundations. However, this framework requires expansion to account for the dynamic and difficult-to-observe context of human-AI collaboration. Brennan (2013) has cautioned that clarity in interpretative claims is a prerequisite for validating arguments, while the conditions of GenAI create new ambiguities about which claims are actually validated when some cognitive processes are delegated to AI systems. Therefore, an expansion of conventional psychometric validity theory is needed towards an assessment validity framework that explicitly incorporates variables such as AI involvement, process transparency, and evaluative judgment as core components of validity arguments.

The empirical conflict in this corpus of literature requires careful reading, as it is not homogeneous. On the one hand, there is optimism about GenAI's potential to enrich learning and personalisation (Deng et al., 2025; Farrelly & Baker, 2023), while on the other hand, there is strong evidence of the failure of the detection approach as an integrity defence mechanism (Weber Wulff et al., 2023). This contradiction does not reflect a lack of evidence on either side but rather reflects that GenAI's impact on assessment validity and integrity depends heavily on how the technology is integrated into the learning design. When GenAI is used as an exploratory tool within an authentically redesigned assessment framework, the impact tends to be positive (Bearman et al., 2024; Furze et al., 2024). However, when GenAI is used as a substitute for cognitive processes in assessments that still rely on the logic of detection and prohibition, the impact tends to undermine validity and trigger adversarial relations between lecturers and students (Eke, 2023; Lodge, 2025; Perkins, 2023).

It is at this point that the argument for a shift from a logic of detection to a logic of design becomes substantively relevant. Redesigning assessments does not mean rejecting regulation or abandoning the risks of fraud, but rather a recognition that the validity of assessments in the AI era can only be restored through designs that explicitly account for the presence of generative technologies, rather than the illusion that they can be fully detected and

prevented. The AIAS framework (Perkins et al., 2024, 2025) and the concept of evaluative judgment (Bearman et al., 2024) demonstrate that the literature has moved toward more pedagogically sound solutions. However, their implementation in practice still faces significant contextual challenges, as Corbin et al. (2025) note regarding the difficulty institutions face in formulating clear boundaries for AI use, and Lodge (2025) notes the risks of uniform policies that ignore disciplinary context.

Furthermore, findings on governance and trust suggest that assessment reform cannot stand alone without the support of clear institutional policies. The wide variation in GenAI policies across universities (Moorhouse et al., 2023) and the low compliance rates of AI declarations, despite their mandatory nature (Gonsalves, 2024), indicate that the assessment integrity crisis is systemic, not simply a problem of individual students. Effective policies require a combination of clear rules, strengthening AI literacy among faculty and students (Almatrafi et al., 2024; Ng et al., 2021, 2024), and a culture of trust that encourages voluntary disclosure rather than compliance based on fear of sanctions. UNESCO global guidelines (Miao & Cukurova, 2024; Miao & Holmes, 2023) and TEQSA national policies (2023, 2025) demonstrate that authoritative institutions at the macro level are beginning to provide a framework of reference. However, implementation at the institutional and study program levels remains highly variable and requires careful contextual adjustment.

From a methodological perspective, this review indicates that the literature on this topic remains dominated by conceptual studies and small-scale case studies, while large-scale, longitudinal experimental studies remain limited, as outlined in Lodge et al.'s (2023) proposed research agenda and Crompton and Burke's (2023) field review. The lack of robust research designs limits the literature's ability to ensure the generalizability of findings regarding the effectiveness of assessment redesign and disclosure policies across different disciplines and cultural contexts. Therefore, a future methodological agenda should focus on experimental and quasi-experimental studies that more objectively assess the long-term impact of frameworks like the AIAS on assessment validity and learning outcomes.

Ultimately, this discussion confirms that the integrity crisis of academic assessment in the GenAI era will not be resolved through a technological race

between generative and detection tools, as the faster-developing generative side structurally always wins such a race. Instead, restoring the validity and credibility of assessment requires a paradigm shift toward evaluation systems consciously designed to accommodate human-AI collaboration, supported by literacy, transparency, and sound institutional governance.

6. Integrative Conceptual Model

Based on the relationship patterns revealed in the previous discussion, this study proposes an integrative conceptual model that positions assessment validity and integrity as consequences of the interaction among antecedents, the design process, mediators, and institutional moderators. At the antecedent level, student and faculty AI literacy, clarity of institutional policies, and the availability of technical guidelines are the main drivers of the assessment ecosystem's readiness for GenAI (Almatrafi et al., 2024; Chan, 2023; Ng et al., 2021, 2024).

At the process level, the use of GenAI in assessment contexts is understood not as a threat to be eliminated, but rather as a design variable that must be explicitly managed through principles of transparency, openness about the level of AI involvement, and process-based assessment. The AIAS framework (Perkins et al., 2024, 2025) and the concept of evaluative judgment (Bearman et al., 2024) serve as key design mechanisms that bridge the antecedent with the outcome. Through these principles, AI engagement can be directed toward strengthening students' reflection and evaluative abilities, rather than simply expediting the production of answers.

The relationship between assessment design and validity outcomes is mediated by the level of trust between faculty and students and adherence to AI disclosure policies (Gonsalves, 2024; Lancaster et al., 2025). When trust is low, even mandatory disclosure policies are less likely to be adhered to, leaving assessment validity vulnerable despite the presence of a design framework. Furthermore, this relationship is moderated by the maturity of institutional governance and the clarity of acceptable boundaries for AI use in each disciplinary context (Corbin et al., 2025; Moorhouse et al., 2023). Institutions with mature governance and contextual policies tend to be better able to translate assessment design principles into consistent practice than those relying on uniform policies or blanket prohibitions (Lodge, 2025).

Thus, the proposed integrative model demonstrates that restoring the validity and credibility of assessments in the GenAI era can only be achieved when all four layers—literacy and policy antecedents, transparent assessment design processes, trust mediators, and institutional governance moderators—work in an integrated manner. Failure in any one layer, such as low AI literacy or weak governance, could potentially undermine the effectiveness of the other layers, even if the assessment design framework used is conceptually sound.

This study theoretically demonstrates the need to transition from conventional psychometric validity theory to argument validity theory, which explicitly incorporates AI engagement, process transparency, and evaluative judgment as core components. The frameworks of Kane (2013) and Messick (1995) remain relevant as foundations, but require conceptual expansion to better account for the dynamic context of human-AI collaboration. Future studies should integrate perspectives on argument validity with the literature on AI literacy, authentic assessment design, and institutional governance in a more integrated manner.

Practically, this study finds an urgent need for higher education institutions to develop implementation strategies at three levels. At the macro level, national policymakers and quality assurance bodies need clear guidance on minimum standards for assessment reform in the AI era, as exemplified by UNESCO and TEQSA documents. At the meso level, institutions need to provide AI literacy training, technical support, and disclosure policies grounded in a culture of trust rather than merely sanctions. At the micro level, lecturers need to redesign assignments and assessment criteria using frameworks such as AIAS to accommodate AI involvement openly and proportionately, aligned with the learning objectives of each course.

LIMITATIONS AND FURTHER RESEARCH

This study has several limitations that readers should be aware of. First, the search protocol was not formally registered with a registry such as OSF Registries or PROSPERO before the search began. Second, the screening and eligibility assessment process was conducted by a single researcher, without calculating interrater agreement measures such as Cohen's Kappa, potentially introducing subjective bias into inclusion and exclusion decisions. Third, the search was limited to English-language articles and policy documents that

could be verified online through official databases and publisher websites, potentially leaving literature in other languages or grey literature that is not digitally indexed.

Fourth, because this topic is evolving rapidly, some of the synthesized articles were published online first, potentially allowing the publisher to adjust the final volume and page numbers after the review was compiled. Fifth, the quality of the studies was assessed qualitatively based on the publishers' credibility and the consistency of their arguments, rather than using a standardized quality assessment instrument such as the Mixed Methods Appraisal Tool. Therefore, comparisons of quality between studies are narrative in nature and cannot be reduced to a single score.

Future research should focus on several priorities. First, develop and empirically test argument validity models that incorporate AI engagement variables as core components, for example, through experimental studies comparing assessment results from conventional designs with those from the AIAS framework. Second, expand longitudinal studies that assess the long-term impact of assessment reforms on the quality of critical thinking and graduate competencies. Third, expand cross-cultural and cross-national studies to understand how different regulatory contexts and academic norms shape variations in the acceptance of AI disclosure policies. Fourth, develop more standardized study quality assessment instruments for interdisciplinary topics such as GenAI and assessment, given the high methodological heterogeneity in the current literature.

CONCLUSION

This systematic review demonstrates that the integrity crisis of academic assessment systems in the GenAI era is fundamentally a validity crisis in the argumentative sense, not simply a moral issue regarding student honesty. Based on a synthesis of 37 articles and core policy documents, this review finds that automated detection-based approaches have been empirically shown to be inadequate for maintaining academic integrity in the long term. The literature consistently calls for systemic reforms that include redesigning authentic assessments using frameworks such as AIAS, strengthening AI literacy, implementing trust-based disclosure policies, and robust institutional governance.

This synthesis also confirms that the effectiveness of assessment reform depends heavily on the interplay between literacy, design, trust, and governance, and that success cannot be achieved through a single intervention at any one level. The tension between optimism about GenAI's pedagogical potential and concerns about threats to assessment validity should not be understood as an unresolvable contradiction, but rather as a sign that the higher education system is in a transition phase toward a new assessment paradigm. Therefore, the future of academic integrity will not be determined by the sophistication of detection tools, but rather by institutions' capacity to design learning ecosystems grounded in evidence that are authentic, transparent, and adaptive to human collaboration with generative artificial intelligence.

REFERENCES

Ali, R., & Maroulis, J. (2026). Assessment validity in the age of generative artificial intelligence: A critical review. Assessment & Evaluation in Higher Education. https://doi.org/10.1080/02602938.2026.2683046

Almatrafi, O., Johri, A., & Lee, H. (2024). A systematic review of AI literacy conceptualization, constructs, and implementation and assessment efforts (2019 to 2023). Computers and Education Open, 6, 100173. https://doi.org/10.1016/j.caeo.2024.100173

Bearman, M., Tai, J., Dawson, P., Boud, D., & Ajjawi, R. (2024). Developing evaluative judgment for a time of generative artificial intelligence. Assessment & Evaluation in Higher Education, 49(6), 893 to 905. https://doi.org/10.1080/02602938.2024.2335321

Bittle, K., & El Gayar, O. (2025). Generative AI and academic integrity in higher education: A systematic review and research agenda. Information, 16(4), 296. https://doi.org/10.3390/info16040296

Brennan, R. L. (2013). Commentary on Validating the interpretations and uses of test scores. Journal of Educational Measurement, 50(1), 74 to 83. https://doi.org/10.1111/jedm.12001

Chan, C. K. Y. (2023). A comprehensive AI policy education framework for university teaching and learning. International Journal of Educational Technology in Higher Education, 20(1), 38. https://doi.org/10.1186/s41239-023-00408-3

Corbin, T., Dawson, P., Nicola, K., & Partridge, H. (2025). Where is the line? It is an absurd line: Towards a framework for acceptable uses of AI in assessment. Assessment & Evaluation in Higher Education, 50(5), 705 to 717. https://doi.org/10.1080/02602938.2025.2456207

Cotton, D. R. E., Cotton, P. A., & Shipway, J. R. (2024). Chatting and cheating: Ensuring academic integrity in the era of ChatGPT. Innovations in Education and Teaching International, 61(2), 228 to 239. https://doi.org/10.1080/14703297.2023.2190148

Crompton, H., & Burke, D. (2023). Artificial intelligence in higher education: The state of the field. International Journal of Educational Technology in Higher Education, 20(1), 22. https://doi.org/10.1186/s41239-023-00392-8

Deng, R., Jiang, M., Yu, X., Lu, Y., & Liu, S. (2025). Does ChatGPT enhance student learning? A systematic review and meta-analysis of experimental studies. Computers & Education, 227, 105224. https://doi.org/10.1016/j.compedu.2024.105224

Eke, D. O. (2023). ChatGPT and the rise of generative AI: Threat to academic integrity? Journal of Responsible Technology, 13, 100060. https://doi.org/10.1016/j.jrt.2023.100060

Farrelly, T., & Baker, N. (2023). Generative artificial intelligence: Implications and considerations for higher education practice. Education Sciences, 13(11), 1109. https://doi.org/10.3390/educsci13111109

Furze, L., Perkins, M., Roe, J., & MacVaugh, J. (2024). The AI Assessment Scale (AIAS) in action: A pilot implementation of GenAI-supported assessment. Australasian Journal of Educational Technology, 40(4), 38–55. https://doi.org/10.14742/ajet.9434

Gonsalves, C. (2024). Addressing student non-compliance in AI use declarations: Implications for academic integrity and assessment in higher education. Assessment & Evaluation in Higher Education. https://doi.org/10.1080/02602938.2024.2415654

Kane, M. T. (2013). Validating the interpretations and uses of test scores. Journal of Educational Measurement, 50(1), 1 to 73. https://doi.org/10.1111/jedm.12000

Kasneci, E., Sessler, K., Kuchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Gunnemann, S., Hullermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., Stadler, M., Weller, J., Kuhn, J., & Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. Learning and Individual Differences, 103, 102274. https://doi.org/10.1016/j.lindif.2023.102274

Kofinas, A. K., Tsay, C. H. H., & Pike, D. (2025). The impact of generative AI on the academic integrity of authentic assessments in higher education. British Journal of Educational Technology, 56(6), 2522–2549. https://doi.org/10.1111/bjet.13585

Lancaster, T., Glendinning, I., Dann, S., Crockett, R., & Draper, M. (2025). Developing policies to address historic contract cheating and misuse of generative artificial intelligence. Journal of Academic Writing, 15(S1), 1 to 13. https://doi.org/10.18552/joaw.v15iS1.1057

Lodge, J. M., Thompson, K., & Corrin, L. (2023). Mapping out a research agenda for generative artificial intelligence in tertiary education. Australasian Journal of Educational Technology, 39(1), 1 to 8. https://doi.org/10.14742/ajet.8695

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as a scientific inquiry into score meaning. American Psychologist, 50(9), 741 to 749. https://doi.org/10.1037/0003-066X.50.9.741

Miao, F., & Cukurova, M. (2024). AI competency framework for teachers. UNESCO. https://doi.org/10.54675/ZJTE2084

Moorhouse, B. L., Yeo, M. A., & Wan, Y. (2023). Generative AI tools and assessment: Guidelines of the world's top-ranking universities. Computers and Education Open, 5, 100151. https://doi.org/10.1016/j.caeo.2023.100151

Ng, D. T. K., Luo, W., Chan, H. M. Y., & Chu, S. K. W. (2024). Design and validation of the AI literacy questionnaire: The affective, behavioral, cognitive, and ethical approach. British Journal of Educational Technology, 55(3), 1082–1104. https://doi.org/10.1111/bjet.13411

Perkins, M. (2023). Academic integrity considerations of AI large language models in the post-pandemic era: ChatGPT and beyond. Journal of University Teaching & Learning Practice, 20(2), Article 7. https://doi.org/10.53761/1.20.02.07

Perkins, M., Roe, J., & Furze, L. (2025). Reimagining the Artificial Intelligence Assessment Scale (AIAS): A refined framework for educational assessment. Journal of University Teaching and Learning Practice, 22(7). https://doi.org/10.53761/rrm4y757

Sullivan, M., Kelly, A., & McLaughlan, P. (2023). ChatGPT in higher education: Considerations for academic integrity and student learning. Journal of Applied Learning & Teaching, 6(1), 31 to 40. https://doi.org/10.37074/jalt.2023.6.1.17

Swiecki, Z., Khosravi, H., Chen, G., Martinez Maldonado, R., Lodge, J. M., Milligan, S., Selwyn, N., & Gasevic, D. (2022). Assessment in the age of artificial intelligence. Computers and Education: Artificial Intelligence, 3, 100075. https://doi.org/10.1016/j.caeai.2022.100075

TEQSA. (2025). Enacting assessment reform in a time of artificial intelligence. Tertiary Education Quality and Standards Agency.

https://www.teqsa.gov.au/guides-resources/resources/corporate-publications/enacting-assessment-reform-time-artificial-intelligence

Tlili, A., Shehata, B., Adarkwah, M. A., Bozkurt, A., Hickey, D. T., Huang, R., & Agyemang, B. (2023). What if the devil is my guardian angel: ChatGPT as a case study of using chatbots in education. Smart Learning Environments, 10(1), 15. https://doi.org/10.1186/s40561-023-00237-x

Weber Wulff, D., Anohina Naumeca, A., Bjelobaba, S., Foltynek, T., Guerrero Dib, J., Popoola, O., Sigut, P., & Waddington, L. (2023). Testing of detection tools for AI-generated text. International Journal for Educational Integrity, 19(1), 26. https://doi.org/10.1007/s40979-023-00146-z

Xia, Q., Weng, X., Ouyang, F., Lin, T. J., & Chiu, T. K. F. (2024). A scoping review on how generative artificial intelligence transforms assessment in higher education. International Journal of Educational Technology in Higher Education, 21, 40. https://doi.org/10.1186/s41239-024-00468-z

AUTHOR PROFILE

Dr. Dra. Sulistiyani, M.M., is a lecturer in the Master of Management Study Program at the Faculty of Economics and Business, University of 17 August 1945 (UNTAG) Semarang. She completed her Bachelor of Management (S1) at the University of 17 August 1945 Semarang in 1991, Master of Management (S2) at Diponegoro University in 2000, and earned her Doctorate of Economics with a concentration in Management from Diponegoro University in 2012. Her doctoral dissertation focused on developing marketing performance through culture-based generative learning and acculturation innovation. In addition to being an active lecturer, she has served as Secretary of the Master of Management Program, Head of the Master of Management Program, Head of the Quality Assurance Unit, and BKD Assessor. Her areas of expertise include marketing management, research methodology, market behaviour, marketing seminars, and human resource management. She is also active in conducting research, publishing scientific papers, and writing books that focus on MSME marketing performance, innovation, organisational learning, and the development of business competitiveness based on local culture.



Drs. Parju SE, Msi. Lecturer, Accounting Study Program, UNTAG Semarang.



The author was born in Sragen on October 21, 1963. He is a permanent lecturer at the Faculty of Economics and Business, UNTAG Semarang. He completed his Bachelor's degree in Management from UNTAG Semarang in 1989, his Bachelor's degree in Accounting from UNTAG Semarang in 2002, and his Master's degree in Financial Management from Hasanuddin University.

CHAPTER VII

Rehumanization of Higher Education in Digital Transformation: A Systematic Review of the Literature

Pilifus Junianto and Brigida Endah Nuraeni
STIE Bentara Persada
Correspondent author: pilifusj@gmail.com

Abstract

This study examines the development of research on GenAI (generative AI/GenAI) in higher education, focusing on the rehumanization of the education system, namely the repositioning of human values as the primary orientation amidst global automation. A Systematic Literature Review (SLR) was conducted on 1,565 Scopus articles for the period 2020–2026. The review followed PRISMA guidelines. Based on a review of titles, abstracts, keywords, and source information, 486 articles were deemed directly or partially relevant to the theme. 74 core articles were selected for in-depth synthesis because they explicitly linked GenAI to ethics, academic integrity, pedagogy, assessment, governance, or a human-centred orientation. The synthesis results indicate that the literature is dominated by empirical studies, centred on the domains of teaching/learning and policy/ethics/governance, and views GenAI as a technology with a dual nature: expanding personalisation, efficiency, and learning support, but simultaneously posing risks to academic integrity, assessment validity, bias, and human agency. Conceptually, the key variables in the corpus form a before-use-after pattern, with AI literacy, self-efficacy, institutional support, and policies serving as antecedents; engagement, motivation, self-regulated learning, and metacognitive support serving as mediators; and assessment design and governance serving as moderators. This study proposes that the future of higher education is not determined solely by the level of AI adoption, but by the institution's capacity to build a human-centred framework that maintains agency, fairness, care, transparency, and academic integrity at the macro, meso, and micro levels.

Keywords: GenAI; higher education; rehumanization; human-centered education; academic integrity

INTRODUCTION

The development of generative artificial intelligence in recent years has rapidly transformed the landscape of higher education, particularly with the emergence of large language models (LLMs) increasingly accessible to lecturers and students (Chan & Hu, 2023; Lim et al., 2023; Stone, 2025). On the one hand, GenAI offers new opportunities for personalised learning, automated feedback, academic writing assistance, and enriching students' interdisciplinary learning experiences (Chan & Hu, 2023; Lim et al., 2023; Stone, 2025). Numerous studies in this corpus confirm that this technology can serve as a pedagogical tool that expands learning support and enhances the learning experience when used purposefully (Dabbagh & Zgheib, 2026; Jing & Mouhong, 2025). However, serious concerns also arise regarding academic integrity, cognitive dependency, algorithmic bias, data privacy, and the diminishment of human roles in pedagogical relationships (Kofinas et al., 2025; Nguyen, 2025).

In this context, discussions about GenAI cannot be limited to issues of technology acceptance or academic efficiency. Higher education normatively functions not only to transfer knowledge but also to shape critical reasoning, ethical responsibility, agency, and reflective capacity in students (Kofinas et al., 2025; Nguyen, 2025). Therefore, a crucial question in the current literature is not simply whether GenAI is beneficial, but rather how this technology should be designed, regulated, and integrated so as not to shift the humanistic goals of education to merely faster output production. Several articles have begun to emphasise that the use of GenAI must be understood through the principles of ethics, pedagogy, and human-centred design so that technology remains in the service of humans, not the other way around (Dabbagh & Zgheib, 2026; Hou, 2026; Jing & Mouhong, 2025; Kofinas et al., 2025; Nguyen, 2025).

Nevertheless, the available literature remains fragmented. Some studies focus on technology adoption using the Technology Acceptance Model (TAM) and the Theory of Acceptance and Use of Technology (UTAUT), with perceived usefulness, social norms, and self-efficacy as the primary determinants of intention to use GenAI (Ren et al., 2026; Ursavaş et al., 2025). Others examine the effects of AI on learning, including engagement, self-regulated learning, and the quality of learning experiences in AI-based

environments (Liu et al., 2026; Xu et al., 2025). Meanwhile, another group of studies highlights the challenges of assessment, academic integrity, and the need for more authentic and adaptive evaluation designs in the GenAI era (Furze et al., 2024; Su et al., 2026). At the institutional level, there is also growing attention to governance and the need for clear AI policies to prevent the wild, unequal, or unethical implementation of AI (Cudjoe & Adebayo, 2025; Tsao, 2025).

This fragmentation indicates that the literature is rich in topics, but has not yet been fully integrated into a single conceptual framework that centres human values as its primary orientation. However, from a rehumanization perspective, higher education in the era of global automation requires a repositioning of human agency, justice, care, transparency, and academic integrity as the foundation for the use of GenAI. This gap makes this SLR important. Using metadata containing 1,565 articles from 2020–2026, this study seeks to synthesise research trends, patterns of variable relationships, the most dominant human values, emerging empirical conflicts, and implementation strategies at the macro, meso, and micro levels.

This article's main contribution lies in three aspects. First, it brings together scattered literature into a more systematic, structured synthesis map. Second, this article proposes a conceptual model that positions GenAI in relation to the antecedents, mediators, moderators, and consequences found repeatedly in the corpus. Third, this article develops the argument that the future of higher education is not determined solely by the speed of AI adoption, but by institutions' capacity to develop human-centred implementation strategies. Therefore, this paper is intended to serve as a conceptual basis for developing further research and formulating higher education policies that remain centred on human values.

RESEARCH QUESTIONS

This study aims to answer five interconnected research questions. First, what are the trends in research development on GenAI in higher education during the 2020–2026 period, specifically those related to rehumanization, human-centred education, ethics, academic integrity, assessment, and institutional governance? Second, what factors consistently emerge as antecedents, mediators, moderators, and consequences of GenAI

use in higher education, and how do the relationships between these factors shape patterns of impact on learning, academic experience, and educational quality? Third, what human values appear most dominantly in the literature, such as agency, justice, care, transparency, integrity, and participation, and how these values are constructed in the discourse of GenAI-based higher education.

Fourth, what empirical risks, issues, and conflicts are most frequently raised in the research, especially those related to plagiarism, bias, dependency, assessment validity, and policy gaps, and to what extent these findings indicate theoretical stability or empirical dissonance. Fifth, what implementation strategies are recommended in the literature at the macro, meso, and micro levels to ensure that GenAI integration remains centred on human values?

METHOD

This research uses the SLR approach to structurally synthesise a collection of articles on the rehumanization of the higher education system in the GenAI era. The main data source comes from Scopus metadata, searched March 19 2016, obtained 1,565 articles spanning 2020–2026 with the query: (TITLE-ABS-KEY (("generative AI" OR "genAI" OR "large language model*" OR LLM* OR ChatGPT OR GPT-3 OR GPT-4 OR Claude OR Gemini OR "Bard" OR "language model*") AND ("higher education" OR universit* OR college* OR "tertiary education" OR "postsecondary education") AND (rehuman* OR "human-centered" OR "human centered" OR humanistic OR "values-based" OR "ethic*" OR "responsible AI" OR "AI ethics" OR dignity OR justice OR fairness OR equity OR "ethics of care" OR agency OR autonomy OR "academic integrity")) AND (LIMIT-TO (SUBJAREA, "SOCI") OR LIMIT-TO (SUBJAREA, "ARTS") OR LIMIT-TO (SUBJAREA, "BUSI") OR LIMIT-TO (SUBJAREA, "PSYC")) AND (LIMIT-TO (DOCTYPE, "ar") OR LIMIT-TO (DOCTYPE, "ch")) AND (LIMIT-TO (LANGUAGE, "English"))).

All of these articles were used as an initial corpus for the identification, screening, and synthesis process. The first stage involves reviewing the metadata for each article, including title, abstract, author keywords, index keywords, journal source, year of publication, and number of citations. This stage aims to map the general character of the literature, including the study

domain, study type, research methods, GenAI focus, and initial relevance to the higher education context.

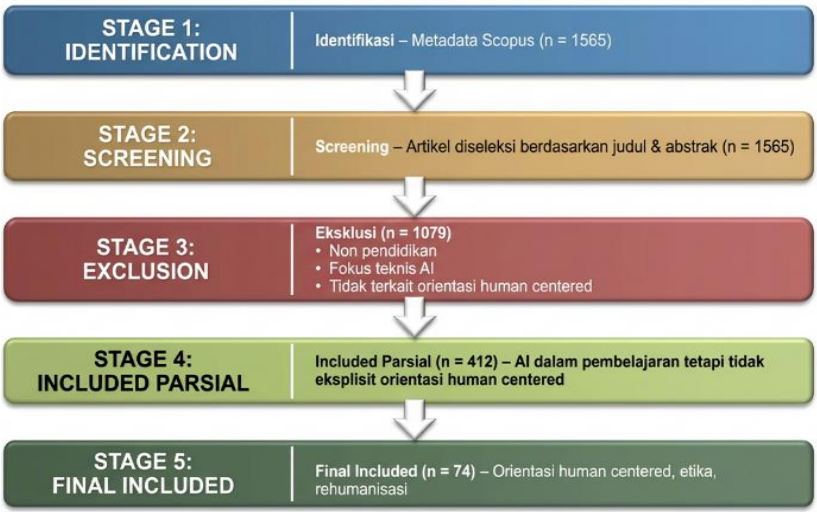
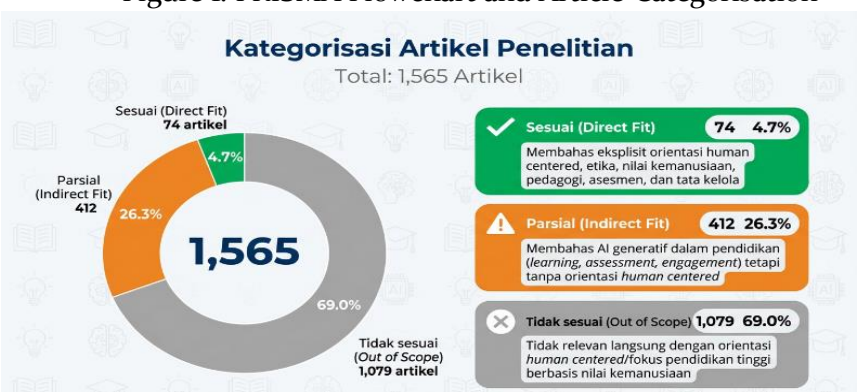


Figure 1. PRISMA Flowchart and Article Categorisation



(Source: Scispace Coding Results, 2026)

The screening process was conducted using inclusion and exclusion criteria developed in accordance with the study's objectives. Articles were included if they discussed GenAI in the context of higher education, explicitly or implicitly, related to learning, assessment, ethics, academic integrity, policy, human-centred education, or humanitarian values. Articles were excluded if they focused solely on the technical aspects of the algorithm without an educational context, predominantly discussed non-educational fields, or had no connection to the theme of rehumanization and the implementation of human-centeredness in higher education. Based on this

process, articles were then classified as strong suitability, partial suitability, and non-suitability.

For in-depth synthesis, 74 core articles were selected as most relevant to the theme because they strongly address issues of ethics, pedagogy, assessment, governance, and a human-centred orientation in the use of GenAI. During the data extraction stage, each article was analyzed using a synthesis matrix that included the article title, journal title, number of citations, country or study context, research type, method, sample or object of study, discussion domain, GenAI focus, study objectives, key findings, antecedent variables, consequences, mediators, moderators, emerging humanitarian values, risks addressed, design principles, macro-, meso-, and micro-level implementation strategies, as well as recommendations and research gaps (Tomczyk et al., 2024). This process resulted in two forms of analysis: descriptive bibliographic analysis and thematic-conceptual analysis.

Descriptive analysis was used to map the distribution of studies by year, research type, method, and domain. In contrast, thematic analysis was used to identify patterns of variable relationships, theoretical stability, empirical conflicts, and methodological gaps. The final stage, narrative and conceptual synthesis, integrated the findings from the core articles into a coherent discussion framework. In this stage, research results were compared to identify general trends, conflicting findings, and under-researched gaps. The synthesis then focuses on formulating a conceptual framework for the rehumanization of the higher education system, situating GenAI within human agency, academic integrity, fairness, transparency, and pedagogical quality.

RESULTS

The extraction results indicate that research articles are dominated by empirical studies (1,045), followed by conceptual studies (416) and reviews (104). This distribution indicates that GenAI has moved beyond the speculative stage into field testing. In terms of methods, quantitative studies (413 articles) were the most common, followed by mixed methods (311 articles) and qualitative (289 articles). The domain distribution shows that teaching and learning (639 articles) are the largest discussion area, followed by policy, ethics, and governance, then research/academic writing, and

assessment/feedback. These findings indicate that GenAI in higher education is treated not merely as a technical issue, but as a pedagogical and institutional issue that touches on the quality of learning, governance, and validity of academic practices.

At the human-centred orientation level, the coding results showed that most of the identified articles (approximately 1,079, ±69.0 %) were categorised as inappropriate (out of scope), as they were not directly relevant to the human-centred orientation and the focus of higher education based on humanitarian values. Meanwhile, approximately 412 articles (±26.3%) fell into the partial (indirect fit) category, discussing GenAI in education (learning, assessment, engagement) but without a human-centred orientation. Seventy-four articles (±4.7%) were categorised as a direct fit, explicitly addressing issues related to higher education with a human-centred orientation, ethics, human values, pedagogy, assessment, and governance.

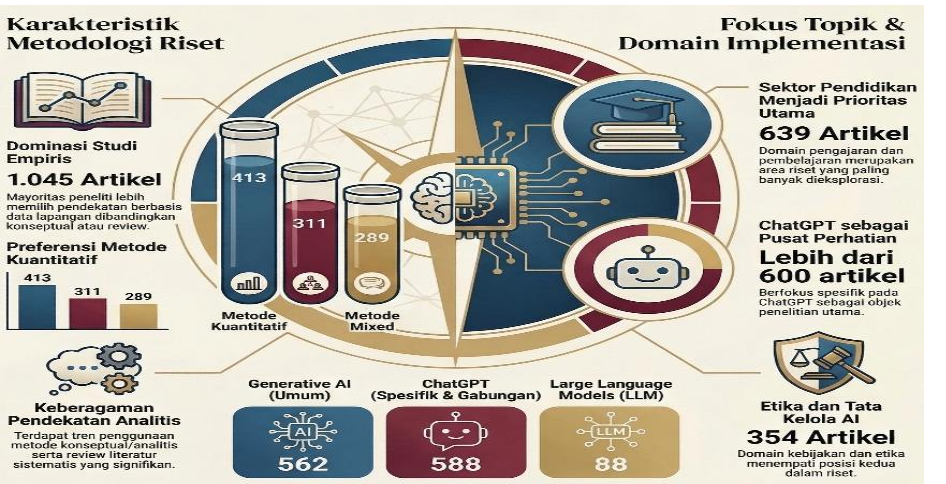
Based on the variable structure, the most frequently emerging pattern was the relationship among the factors driving GenAI use, the mediating psychopedagogical mechanisms, and the resulting learning outcomes or academic consequences. AI literacy, self-efficacy, perceived usefulness, social norms, institutional support, and policies were the most consistent antecedents (Ren et al., 2026; Ursavaş et al., 2025). Meanwhile, the relationship between GenAI use and learning outcomes was not direct but mediated by psychopedagogical variables such as engagement, motivation, self-regulated learning, and metacognitive support.

These studies indicate that improvements in learning quality, satisfaction, and academic performance only occur when AI is used to enhance reflection and learning strategies, rather than as a substitute for students' cognitive activities. Furthermore, moderating variables such as assessment design, institutional policies, and ethical awareness have been shown to significantly strengthen or weaken the relationship between AI use and learning outcomes. In this context, the issues of academic integrity and assessment validity are crucial, as the effectiveness of AI depends heavily on the institution's readiness to design policies and evaluation systems that adapt to technological developments (Kofinas et al., 2025; Su et al., 2026). Thus, the identified variable patterns underscore that the implementation of GenAI in higher education must be understood as an integrated system, where success

is determined not only by the technology itself but also by the interactions among individual factors, pedagogy, and institutional governance.

Among the core articles, issues of ethics and academic integrity were the most prominent. Kofinas et al. (2025) demonstrated that GenAI challenges the integrity of authentic assessments because lecturers cannot always discern the involvement of AI in students' academic output. Furze et al. (2024) demonstrated that assessment design needs to move from a detection logic to a design logic that openly and structurally acknowledges the presence of AI. Similarly, Su et al. (2026) emphasised the importance of new authentic assessment designs that assess process, reflection, and accountability, rather

Figure 2. Article Profile



(Source: NotebookLM Coding Results, 2026)

Meanwhile, on the learning aspect, Liu et al. (2026) demonstrated that acceptance of GenAI influences student engagement through self-efficacy and motivation, while Xu et al. (2025) demonstrated that metacognitive support plays a crucial role in strengthening self-regulated learning in GenAI-based environments. In terms of adoption, Ren et al. (2026) and Ursavaş et al. (2025) showed that perceived usefulness, social norms, and psychological and institutional factors shape the intention and actual use of GenAI. At the policy level, weak or uneven AI governance results in inconsistent, subjective practices that are prone to academic fraud (Cudjoe & Adebayo, 2025; Hayes et al., 2025; Tsao, 2025).

DISCUSSION

The findings of this SLR indicate that GenAI in higher education is developing amidst two major, simultaneous trends: the drive for academic efficiency and the need to maintain human values in the educational process. On the one hand, GenAI is seen as an instrument capable of accelerating cognitive work, supporting personalised learning, and expanding access to student academic support (Chan & Hu, 2023; Lim et al., 2023; Stone, 2025). On the other hand, the greater the level of automation, the more pressing the need to reaffirm higher education's role as a space for the formation of reason, ethics, responsibility, and interpersonal relationships in students (Chan & Hu, 2023; Lim et al., 2023; Stone, 2025). Therefore, the key issue is not simply whether GenAI is beneficial, but rather how it is positioned within a human-centred pedagogical architecture (Hou, 2026; Nguyen, 2025).

Furthermore, regarding theoretical stability, it was found that this aspect still relies heavily on technology acceptance frameworks, particularly TAM and UTAUT, to explain intentions and actual use of GenAI. These models have been shown to help explain why perceived usefulness, perceived ease of use, social norms, and self-efficacy are important determinants of technology adoption by students and lecturers (Ren et al., 2026; Ursavaş et al., 2025). However, this study's findings also indicate that the technology acceptance framework alone is insufficient to explain the complex impact of GenAI on higher education.

The use of AI is not simply a matter of adoption but also of transforming the learning process, changing pedagogical relationships, and redistributing agency between humans and systems. Therefore, expanding the technology adoption model toward a human-centred model of educational AI that incorporates engagement, self-regulated learning, academic integrity, fairness, and assessment design as core components is needed.

The presence of a mediator strengthens this argument. The impact of GenAI on learning outcomes rarely appears as a simple linear relationship. Instead, its influence tends to operate through motivation, self-efficacy, engagement, metacognition, and self-regulation. Liu et al. (2026) showed that acceptance of GenAI influences engagement through self-efficacy and motivation, while Xu et al. (2025) demonstrated that metacognitive support plays a critical role in strengthening self-regulated learning. This means that

GenAI should not be understood as a direct cause of improved performance, but rather as a trigger for certain learning conditions that only produce positive impacts when supported by appropriate pedagogical design.

It is at this point that the concept of rehumanization becomes substantively relevant. Rehumanization in higher education does not mean rejecting AI; rather, it is an effort to place technology under the broader goal of education: the formation of reflective, responsible, and empowered individuals. Core literature shows that values such as agency, fairness, care, integrity, and transparency repeatedly emerge in response to concerns about excessive automation (Hayes et al., 2025; Hou, 2026; Liu et al., 2026; Nguyen, 2025; Xu et al., 2025). However, in many studies, these values remain implicit and have not yet been developed into a complete conceptual framework. This explains why only a small number of articles explicitly adopt a human-centred lens, while most remain focused on practical ethics or risk management.

The empirical conflicts identified need to be read carefully because they are not homogeneous. On the issue of academic integrity and assessment, for example, contradictions between studies do not necessarily indicate weak evidence, but rather reflect the heterogeneity of implementation contexts. On the one hand, GenAI can accelerate feedback and expand personalised learning, even enabling more consistent scoring automation at scale (AlGhamdi et al., 2026; Jing & Mouhong, 2025). Some studies argue that GenAI increases the risk of plagiarism and blurs the boundaries of intellectual property. However, others show that when assessments are redesigned in an authentic, process-based manner, AI can become a legitimate tool for reflection and learning support.

The findings of Kofinas et al. (2025) demonstrate the serious challenges to academic integrity in authentic assessments, while Su et al. (2026) emphasise the need for new assessment designs that are better adapted to the GenAI era. Thus, the empirical conflicts are rooted in pedagogical and governance unpreparedness, not solely in the technology. Another conflict arises over ethics, bias, and privacy. While bias and discrimination have not yet dominated the number of articles, unlike academic integrity, the issue remains highly significant normatively due to its implications for fair access, representation, and the quality of educational decisions. Li et al. (2025) demonstrate that social bias in ChatGPT in higher education has been

meaningfully discussed, although the research landscape is still developing. At the same time, concerns about privacy, trust, and data protection are important factors in students' and academic staff's acceptance of AI (Weis et al., 2026). This means that rehumanization must be built not only to prevent plagiarism but also to maintain fairness, security, and user rights.

Furthermore, at the implementation level, this study demonstrates that rehumanization of AI in higher education requires a multi-layered strategy at the macro, meso, and micro levels. At the macro level, policy and governance are key prerequisites for preventing the development of AI from being wildly uneven. Evidence from Sub-Saharan Africa shows that the absence of policies or weak institutional regulatory frameworks creates a highly variable space for interpretation and is prone to inequitable practices (Cudjoe & Adebayo, 2025). At the meso-level, institutions need to provide training, AI literacy, technical support, and clear guidelines for lecturers and students. At the micro-level, courses and assessments need to be redesigned to emphasise process, reflection, dialogue, and accountability in AI use. The post-digital perspective of Hayes et al. (2025) is important here because it emphasises that AI governance must be sensitive to the social, historical, and relational contexts of higher education practices.

Methodologically, this SLR demonstrates a strong need to improve the quality of research design in this area. The dominance of cross-sectional studies and perception surveys has led many studies to stop at the level of intention to use, satisfaction, or perceived usefulness, without being able to track changes in learning behaviour and academic quality over the long term (Hayes et al., 2025; Hou, 2026; Liu et al., 2026; Nguyen, 2025; Xu et al., 2025).

The lack of longitudinal designs, robust experiments, and mixed methods also limits researchers' ability to understand how GenAI interacts with academic culture, disciplines, and institutional policies (Chan, 2023; Cotton et al., 2024; Hamdan, 2025; Lund et al., 2023; Sah et al., 2025). Furthermore, the limited use of mediators and moderators in many research models makes results appear simple, even though the relationships are actually complex. Therefore, the future methodological agenda should be directed towards richer structural models, cross-context studies, and approaches that link behavioural data, user experiences, and more objective learning outcomes.

Finally, this discussion emphasises that the future of higher education in the GenAI era is not determined by how quickly institutions adopt technology, but by how maturely they design a human-centred framework to guide this technology. Rehumanization, in this context, is not a normative slogan but rather a conceptual and practical strategy to ensure that automation remains subordinate to meaningful educational goals. Therefore, the research proposes a shift from a technology-centric to an education-centric paradigm, treating GenAI as a tool that strengthens human agency rather than replacing it. Although there are more than 1,500 studies on GenAI, only a small number explicitly integrate a human-centred perspective, indicating a significant gap in the literature linking technology to human values in higher education.

1. Integrative Conceptual Model

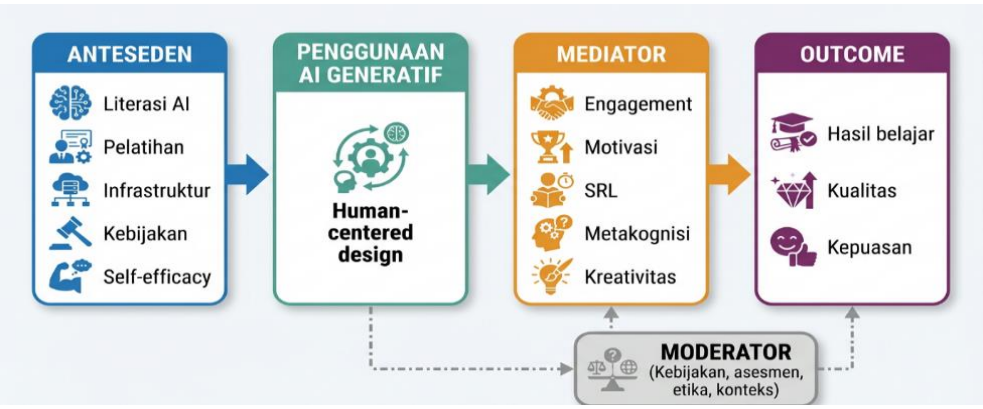
Based on the relationship patterns discussed previously, this article proposes an integrative conceptual model that positions the use of GenAI as a process shaped by antecedent factors, mediated by psychopedagogical mechanisms, moderated by institutional design, and producing both learning consequences and academic risks. In this model, AI literacy, self-efficacy, perceived usefulness, social norms, institutional support, training, infrastructure, and policies serve as key antecedents. These variables influence the intensity and quality of GenAI use by lecturers and students (Ren et al., 2026; Ursavaş et al., 2025).

At the process layer, the use of GenAI is understood not as a neutral technical action but as a pedagogical practice whose quality of impact is largely determined by human-centred principles. These principles include transparency, agency, fairness, care, accountability, and ethical orientation. Through these principles, GenAI can be directed to strengthen reflection, exploration, and engaged learning, not just to speed up the production of answers. The influence of AI use on learning outcomes is then explained through mediators such as engagement, motivation, self-regulated learning, and metacognitive support (Liu et al., 2026; Xu et al., 2025).

However, this relationship is not always linear because it is moderated by assessment design, institutional policy, governance readiness, and academic culture. Assessment designs that still rely on the end product tend to increase the risk of AI misuse. In contrast, assessments based on process,

reflection and accountability are more likely to allow AI to function as a legitimate assistive tool. Likewise, institutions that have clear policies and effective implementation support tend to be better able to minimise bias, uncertainty, and practical conflicts than institutions that lack a mature governance framework (Cudjoe & Adebayo, 2025; Kofinas et al., 2025; Su et al., 2026). Thus, the proposed integrative model shows that human-centred educational outcomes can be achieved only when GenAI is placed within the right pedagogical and institutional ecosystem. The following is a proposed model image:

Figure 3. Integrative Conceptual Model of Human-Centred GenAI



(Source: Scispace Coding Results, 2026)

2. Implications

This study theoretically demonstrates the need for a transition from technology adoption theory to a theory of AI rehumanization in higher education. Frameworks such as TAM and UTAUT remain useful for explaining the initial stages of adoption but are inadequate for capturing the changing pedagogical relationships and normative consequences of GenAI use. Therefore, future studies should integrate a technological perspective with self-regulated learning, metacognition, academic integrity, and human-centred design.

The practical implications of this study also reveal the urgency for higher education institutions to develop implementation strategies at three levels. At the macro level, universities and policymakers need clear regulations regarding the use of GenAI. At the meso level, institutions need to prepare training, technical support, AI literacy, and contextual governance

mechanisms. At the micro level, lecturers need to redesign classroom activities and assessments to be more process-based, reflective, collaborative, and accountable for AI use. This strategy will enable AI integration to align with human values, rather than the other way around.

4. Limitations and Further Research Agenda

This study has several limitations. The initial screening process of 1,565 articles relied on metadata, particularly titles, abstracts, and keywords. Therefore, some conceptual insights that were only apparent in the overall text may not have been fully captured. Although 74 core articles were selected for in-depth analysis and justified by the researchers, the selection process was still guided by a thematic coding framework and therefore requires further validation in light of subsequent scientific developments. Several categories, such as study type, method, domain, and human-centred orientation, were derived from metadata-based coding; therefore, they are best-effort and should be understood as a synthesis map rather than a substitute for a full review of each article.

Future research should address several priorities. First, developing and testing a human-centred generative AI model that explicitly incorporates agency, fairness, care, transparency, and academic integrity as core theoretical components. Second, expanding longitudinal studies and experiments that assess the long-term impact of GenAI on the learning habits, critical thinking skills, and quality of academic production of lecturers and students. Third, expanding cross-cultural and cross-institutional studies to elucidate how policies, social norms, and infrastructure readiness shape variations in empirical findings. Fourth, deepen the study of assessment designs compatible with the GenAI era so that the thinking process, reflection, and responsibility of AI use can be measured more authentically.

CONCLUSION

This SLR demonstrates that GenAI has become a transformational force in higher education. However, its impact should not be understood solely through the lens of technological efficiency or ease of use. Based on the analysis of the Scopus literature obtained, the use of GenAI is always intertwined with issues of agency, academic integrity, fairness, transparency, and the quality of pedagogical relationships. Therefore, the key issue in higher

education is no longer simply accepting or rejecting GenAI, but rather designing conditions for this technology to work in harmony with the humanistic goals of education.

The results of this review substantively demonstrate that the benefits of GenAI tend to emerge when its use is mediated by motivation, engagement, self-regulated learning, and metacognitive support, and moderated by assessment design, institutional policies, and ethical governance. These findings also emphasise that the tension between innovation optimism and ethical concerns should not be seen as an irresolvable contradiction, but rather as a sign that the higher education system is in a transitional phase. Therefore, the rehumanization of the higher education system in the era of global automation must be understood as a scientific, pedagogical, and institutional project that positions GenAI to serve more equitable, reflective, and meaningful learning.

REFERENCES

AlGhamdi, E. M., Li, Y., Gašević, D., & Chen, G. (2026). Leveraging prompt-based LLMs for automated scoring and feedback generation in higher education. *Computers and Education*, *243*. https://doi.org/10.1016/j.compedu.2025.105511

Chan, C. K. Y. (2023). A comprehensive AI policy education framework for university teaching and learning. *International Journal of Educational Technology in Higher Education*, *20*(1). https://doi.org/10.1186/s41239-023-00408-3

Chan, C. K. Y., & Hu, W. (2023). Students' voices on generative AI: perceptions, benefits, and challenges in higher education. *International Journal of Educational Technology in Higher Education*, *20*(1). https://doi.org/10.1186/s41239-023-00411-8

Cotton, D. R. E., Cotton, P. A., & Shipway, J. R. (2024). Chatting and cheating: Ensuring academic integrity in the era of ChatGPT. *Innovations in Education and Teaching International*, *61*(2), 228–239. https://doi.org/10.1080/14703297.2023.2190148

Cudjoe, E. K., & Adebayo, J. O. (2025). Investigating the Existence of AI Policy in Higher Education Institutions in Sub-Saharan Africa. *Proceedings of the Association for Information Science and Technology*, *62*(1), 1414–1417. https://doi.org/10.1002/pra2.1421

Dabbagh, N., & Zgheib, G. (2026). Using Generative AI to Personalise Teaching and Learning in Higher Education Contexts. In *Higher Education in the*

Arab World: Artificial Intelligence (pp. 15–43). https://doi.org/10.1007/978-3-031-99068-7_2

Furze, L., Perkins, M., Roe, J., & MacVaugh, J. (2024). The AI Assessment Scale (AIAS) in action: A pilot implementation of GenAI-supported assessment. *Australasian Journal of Educational Technology*, *40*(4), 38–55. https://doi.org/10.14742/ajet.9434

Hamdan, A. (2025). The double-edged sword of AI-integrated education: an investigation into personalised and inclusive learning in higher education. In *Recent Research in Management, Accounting and Economics (RRMAE): A Case Study on Recent Research in Management, Accounting and Economics* (pp. 381–391). https://doi.org/10.4324/9781003606642-91

Hou, Y. (2026). Natural intelligence, not artificial: a Confucian reframing of generative AI in higher education. *Ethics and Education*. https://doi.org/10.1080/17449642.2026.2629430

Jing, Z., & Mouhong, Z. (2025). Exploring the integration of AI and national quality courses in China: a study on teaching practices in nursing smart education. *BMC Medical Education*, *25*(1). https://doi.org/10.1186/s12909-025-08289-y

Kofinas, A. K., Tsay, C. H.-H., & Pike, D. (2025). The impact of generative AI on the academic integrity of authentic assessments in higher education. *British Journal of Educational Technology*, *56*(6), 2522–2549. https://doi.org/10.1111/bjet.13585

Li, M., Enkhtur, A., Yamamoto, B. A., Cheng, F., & Chen, L. (2025). Potential Societal Biases of ChatGPT in Higher Education: A Scoping Review. *Open Praxis*, *17*(1), 79–94. https://doi.org/10.55982/openpraxis.17.1.750

Lim, W. M., Gunasekara, A., Pallant, J. L., Pallant, J. I., & Pechenkina, E. (2023). Generative AI and the future of education: Ragnarök or reformation? A paradoxical perspective from management educators. *International Journal of Management Education*, *21*(2). https://doi.org/10.1016/j.ijme.2023.100790

Liu, J., Liu, M., Yao, Y., & Li, D. (2026). Perceptions and Acceptance of Generative Artificial Intelligence Influencing Chinese EFL Learners' Engagement in Informal Digital Learning of English: Mediating Roles of Self-Efficacy and Motivation. *International Journal of Applied Linguistics (United Kingdom)*. https://doi.org/10.1111/ijal.70122

Lund, B. D., Wang, T., Mannuru, N. R., Nie, B., Shimray, S., & Wang, Z. (2023). ChatGPT and a new academic reality: Artificial Intelligence-written research papers and the ethics of the large language models in scholarly publishing. *Journal of the Association for Information Science and Technology*, *74*(5), 570–581. https://doi.org/10.1002/asi.24750

Nguyen, K. V. (2025). The Use of Generative AI Tools in Higher Education: Ethical and Pedagogical Principles. *Journal of Academic Ethics*, *23*(3), 1435–1455. https://doi.org/10.1007/s10805-025-09607-1

Ren, J., Shahzad, M. F., Abbas, J., Al-Sulaiti, K. I., & Pilař, L. (2026). Factors influencing the adoption and use of generative artificial intelligence (e.g., ChatGPT) in higher education: an extended UTAUT2 perspective. *Journal of Further and Higher Education*. https://doi.org/10.1080/0309877X.2025.2573026

Sah, R., Hagemaster, C., Adhikari, A., Lee, A., & Sun, N. (2025). Generative AI in higher education: student and faculty perspectives on use, ethics, and impact. *Issues in Information Systems*, *26*(2), 373–386. https://doi.org/10.48009/2_iis_129

Stone, B. W. (2025). Generative AI in Higher Education: Uncertain Students, Ambiguous Use Cases, and Mercenary Perspectives. *Teaching of Psychology*, *52*(3), 347–356. https://doi.org/10.1177/00986283241305398

Su, F., O'Dea, M., Wood, M., & Seidl, E. (2026). Embracing a new world: authentic assessment designs in an age of generative artificial intelligence (GenAI). *Perspectives: Policy and Practice in Higher Education*. https://doi.org/10.1080/13603108.2025.2601741

Tomczyk, P., Brüggemann, P., & Paul, J. (2024). Variable science mapping as a literature review method. *Journal of Marketing Analytics*, *12*(4), 829–841. https://doi.org/10.1057/s41270-024-00336-9

Tsao, J. (2025). Trajectories of AI policy in higher education: Interpretations, discourses, and enactments of students and teachers. *Computers and Education: Artificial Intelligence*, *9*. https://doi.org/10.1016/j.caeai.2025.100496

Ursavaş, Ö. F., Yalçın, Y., İslamoğlu, H., Bakır-Yalçın, E., & Cukurova, M. (2025). Rethinking the importance of social norms in generative AI adoption: investigating the acceptance and use of generative AI among higher education students. *International Journal of Educational Technology in Higher Education*, *22*(1). https://doi.org/10.1186/s41239-025-00535-z

Weis, L., Bele, J. L., & Erčulj, V. (2026). Acceptance and Use of Generative Artificial Intelligence in Higher Education: A UTAUT-Based Model Integrating Trust and Privacy. *Education Sciences*, *16*(2). https://doi.org/10.3390/educsci16020173

Xu, X., Qiao, L., Cheng, N., Liu, H., & Zhao, W. (2025). Enhancing self-regulated learning and learning experience in generative AI environments: The critical role of metacognitive support. *British Journal of Educational Technology*, *56*(5), 1842–1863. https://doi.org/10.1111/bjet.13599

PROFIL AUTHOR

Pilifus Junianto, S.S., M.M. was born in Lubuk Linggau on June 30, 1982. He



earned a Bachelor's in Philosophy from Parahyangan University, Bandung (2008), a Bachelor of Arts in Philosophy from Wedhabakti USD Yogyakarta (2012), and a Master's in Management from Parahyangan University (2017), and is now pursuing a doctorate at Tarumanagara University, Jakarta. His certifications include Qualified Service Excellence Professional (2022), BNSP Digital Marketing (2023), and BNSP Data Science (2025). He has published in reference books, anthologies, and journals, received a Dikti Grant (2024), and reviewed for Emerald Group Publishing. He has taught marketing and research methods at STIE Bentara Persada Batam since 2017.

Brigida Endah Nuraeni was born in Grobogan on February 15, 1968. She



holds a bachelor's degree in Education from Sarjanawiyata Tamansiswa University (UST) and a master's in Human Resource Management from Mercu Buana University, Jakarta. For over seven years, she has researched Management, focusing on Human Resources, while lecturing in the Management Study Program at Bentara Persada College of Economics (STIE BP). She has published in national and international journals and spoken at management events. Her earlier books cover Leadership Management in the Digital Business World, Technology-Based Research Methodology, and Innovation Through Knowledge Management. Contact: brigidaendah15@gmail.com

CHAPTER VIII

Reorienting Educational Goals in the Age of Artificial Intelligence: A Systematic Literature Review of Critical Thinking and Intellectual Autonomy

Benediktus Denar[1] and Theresia Yovita Cendana Sari[2]
[1]STIPAS St. Sirilus Ruteng
[2]Sekolah Tinggi Agama Katolik Negeri Pontianak
Correspondent Author: benydenar@gmail.com

ABSTRACT
This study aims to systematically examine the reorientation of educational goals in the era of artificial intelligence (AI), with particular emphasis on the importance of critical thinking and intellectual autonomy as essential competencies for students. The study employed a systematic literature review (SLR) using the PRISMA framework. Literature searches were conducted in Scopus, ERIC, and Google Scholar using the keywords "artificial intelligence," "education," "critical thinking," and "intellectual autonomy." Following identification, screening, and eligibility assessment, 25 publications met the inclusion criteria and were analyzed using thematic synthesis. The findings reveal that while AI-driven access to information has the potential to enhance learning effectiveness, it may also contribute to technological dependency, intellectual passivity, and the weakening of reflective capacities when not accompanied by appropriate pedagogical frameworks. The study further indicates that the conventional educational paradigm centered on knowledge transmission is increasingly inadequate in the context of cognitive automation. Therefore, educational goals need to be reoriented from merely acquiring and mastering knowledge toward cultivating individuals who are critical, autonomous, reflective, and ethically responsible. In this way, education can continue to fulfill its role as a process of humanization amid the rapid development of artificial intelligence technologies.

Keywords: artificial intelligence; ed1ucational goals; critical thinking; intellectual autonomy; humanization.1

INTRODUCTION

The development of artificial intelligence (AI) in recent years has brought significant changes to various areas of life, including education. The integration of AI into education has shown rapid growth over the past decade and has the potential to revolutionize the education system as a whole (Prahani et al., 2022; Holmes et al., 2019; Zawacki-Richter et al., 2019; Luckin, 2018). Beyond technological innovation, AI challenges education to reconsider its fundamental purpose: not merely preparing learners to acquire knowledge, but cultivating human capacities that cannot be reduced to computational processes.

The presence of this technology no longer functions merely as a tool but is increasingly developing into a cognitive partner that shapes how humans learn (Luckin, 2018; Selwyn, 2019). AI is beginning to influence how humans access, process, and even construct knowledge. Information that was previously acquired through lengthy processes is now available almost instantaneously, while answers to a wide range of questions can be generated within seconds (Kasneci et al., 2023). In this context, the learning process is undergoing a fundamental transformation: from primarily seeking knowledge toward developing the ability to select, evaluate, manage, and interpret available information (Holmes et al., 2019).

However, amid this transformation, formal education in many contexts continues to rely on the traditional paradigm of knowledge transmission (Biesta, 2015). Students are often positioned as recipients of information, while teachers remain the primary authorities of knowledge. An orientation toward cognitive achievement, measured primarily through examinations and academic scores, continues to serve as the dominant indicator of educational success. This paradigm becomes increasingly problematic as AI systems are now capable of performing certain cognitive functions more rapidly and efficiently than humans (Kasneci et al., 2023). In this context, recent studies indicate that AI integration in learning can enhance access to information, student engagement, and understanding; however, it also introduces the risk of technological dependency when not accompanied by appropriate pedagogical frameworks (Shukor & Osman, 2025).

This situation reveals an increasingly evident gap. On the one hand, contemporary society has entered an era that requires higher-order, reflective, and context-sensitive thinking skills. On the other hand, educational practices have not yet fully equipped students with the capacity to develop critical thinking and intellectual autonomy. Recent studies indicate that AI use has ambivalent implications: while it can support the development of critical thinking, it may also weaken self-evaluation and judgment capacities when learners engage with it passively (Kasneci et al., 2023; Ninghardjanti et al., 2025). Indeed, systematic studies suggest that uncritical dependence on AI can diminish cognitive engagement and undermine learners' intellectual autonomy (Holmes et al., 2019).

In such situations, students risk becoming passive users of technology, relying heavily on AI-generated outputs without developing the capacity to independently evaluate, question, and critically process the information provided. Such dependence may not only weaken reflective capacity but also undermine the intellectual autonomy that lies at the heart of the educational process itself (Biesta, 2015; Kasneci et al., 2023).

Based on this situation, several fundamental questions emerge: Is the purpose of education still relevant if it is understood primarily as the transmission of knowledge? How should education respond to the changing ways in which humans think, learn, and construct knowledge in the age of artificial intelligence? More fundamentally, how can education cultivate students who are not only proficient in using technology but also capable of thinking critically, independently, and reflectively amid the rapid flow of information?

This article aims to systematically examine educational goals in the context of the AI era. Specifically, it emphasizes the need to reorient educational goals from mere knowledge acquisition toward the cultivation of critical thinking skills and intellectual autonomy among students. From this perspective, education should not only produce individuals capable of accessing information but also individuals able to understand, evaluate, interpret, and assume responsible positions in an increasingly complex, rapidly changing world. Unlike previous studies that have primarily focused on the pedagogical applications of AI, this

article examines the philosophical reorientation of educational goals by positioning critical thinking and intellectual autonomy as essential aims of education in the AI era. The novelty of this article lies in its systematic synthesis of literature on AI, education, philosophy, and ethics, which have often been examined separately across different disciplinary domains.

METHOD

This study employed a systematic literature review (SLR) following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) framework. This approach was adopted because the study pursued a conceptual and reflective objective: to map, synthesize, and critically analyze scholarly literature on artificial intelligence, educational philosophy, and contemporary learning theories. **Search Strategy**. A comprehensive literature search was conducted across three major databases—Scopus, ERIC, and Google Scholar—to identify relevant publications published between 2012 and 2025. The search employed Boolean operators using the following search string: ("artificial intelligence" OR "AI") AND ("education" OR "learning") AND ("critical thinking" OR "intellectual autonomy" OR "educational goals"). Searches were conducted in both English and Indonesian to ensure broader coverage of relevant scholarly literature.

Inclusion and Exclusion Criteria. Studies were included in this review if they: (1) were peer-reviewed journal articles or scholarly books examining the relationship between artificial intelligence (AI) and educational goals, critical thinking, or intellectual autonomy; (2) were published between 2012 and 2025; (3) were available in full text; and (4) were written in English or Indonesian. Studies were excluded if they: (1) did not address the relationship between AI and educational goals or educational philosophy; (2) consisted of opinion pieces or non-scholarly publications lacking an academic foundation; or (3) were not available in full text.

Selection Process. The study selection process followed the PRISMA framework. During the identification stage, 180 records were retrieved from the three databases. After duplicate records were removed, 150 records remained for screening. During the title and abstract screening stage, 95 records were excluded for not being relevant to the research focus. The

remaining 55 full-text publications were assessed for eligibility; 30 were excluded for failing to meet the inclusion criteria, such as focusing exclusively on the technical aspects of AI without addressing its educational or philosophical implications. Ultimately, 25 studies (journal articles and scholarly books) met the eligibility criteria and were included in the final synthesis.

Data Analysis. The selected studies were analyzed using a thematic synthesis approach. The analysis involved identifying key concepts from each study, grouping them into overarching themes, and integrating the findings into a critical and interpretive synthesis. Three major themes emerged from this process: (1) AI-driven educational disruption; (2) the limitations of conventional educational goals; and (3) the reorientation of educational goals toward fostering critical thinking and intellectual autonomy. These themes provided the analytical framework for the Results and Discussion sections.

RESULTS AND DISCUSSION

Table 1 presents the thematic mapping of the key literature synthesized in this review.

Table 1. Thematic Mapping of the Reviewed Literature

Theme	Key Sources	Main Findings
AI-Driven Educational Disruption	Holmes et al. (2019). Kasneci et al. (2023); Selwyn (2019); Williamson (2017); Gobble (2016)	AI transforms how knowledge is accessed and produced, thereby shifting the teacher's epistemic authority.
The Limitations of Conventional Educational Paradigms	Freire (1970); Biesta (2015); Aoun (2017); Denar (2021)	An education system centered on knowledge transmission is no longer adequate in the era of cognitive automation.
Critical Thinking and Intellectual Autonomy as Essential Competencies	Facione (2015); Paul & Elder (2020); Brookfield (2012); Mezirow (2000); OECD (2018)	Critical thinking and intellectual autonomy have become essential competencies for learners in the twenty-first century.
Ethical and Anthropological Dimensions of AI	UNESCO (2021); Floridi & Cowls (2022); Harari (2016); Bostrom (2014)	The use of AI demands moral preparedness and a well-defined ethical framework

Artificial Intelligence and Educational Disruption

Artificial intelligence (AI) in education refers to the application of computational systems capable of simulating human cognitive functions, such as natural language processing, pattern recognition, prediction, and the provision of adaptive learning support (Holmes et al., 2019). These technologies encompass a wide range of applications, including intelligent tutoring systems, learning analytics, and large language models (LLMs) that can generate text, answer questions, and assist learners in solving problems in real time (Kasneci et al., 2023). Recent advances have transformed AI from a mere technological tool into a cognitive partner that increasingly shapes how people learn, access information, and construct knowledge. Consequently, AI has become an integral component of the digital educational ecosystem, reshaping the relationships among teachers, learners, and knowledge itself (Williamson, 2017; Selwyn, 2019).

This phenomenon can be understood as a form of educational disruption—that is, a fundamental transformation that not only introduces new technological innovations but also reshapes, and in some cases replaces, established ways of producing, disseminating, and interpreting knowledge (Gobble, 2016). The disruption brought about by AI is not merely additive but transformative, as it challenges the epistemological foundations of education itself by redefining how knowledge is acquired, who is recognised as a legitimate source of knowledge, and how learning takes place.

One of the most significant transformations concerns the way knowledge is accessed. Whereas knowledge was once acquired through a gradual process of reading, listening to instruction, and reflecting, information is now available instantaneously, in unprecedented abundance, and with remarkable ease. Students can obtain responses to almost any question within seconds. This shift represents more than an increase in efficiency; it constitutes a structural transformation of the educational ecosystem that fundamentally redefines the logic of the learning process itself (Goldie, 2016; Holmes et al., 2019). Consequently, learning is no longer primarily about searching for knowledge but about developing the capacity to identify, evaluate, and meaningfully interpret the vast amount of information available.

This transformation has also reshaped the distribution of epistemic authority within education. In the traditional paradigm, teachers are regarded

as the primary source of knowledge. However, the emergence of AI systems capable of rapidly generating and providing information at scale has fundamentally altered this epistemic landscape (Biesta, 2015). Teachers no longer function primarily as transmitters of knowledge but increasingly serve as facilitators of learning, curators of knowledge, and mentors who guide students in developing critical reasoning and intellectual autonomy. Consequently, pedagogical relationships are evolving from hierarchical models of instruction toward more dialogical, collaborative, and reflective forms of learning.

At the same time, the convenience afforded by AI also brings the risk of increasing dependence on technology. Students may become inclined to rely on AI not only to answer questions and complete assignments but also to generate ideas and construct arguments. If left unaccompanied by critical awareness, such dependence may diminish learners' cognitive engagement in the learning process (Kasneci et al., 2023). Rather than engaging in sustained inquiry and deep reflection, learners may adopt patterns of instant problem-solving that prioritise immediate outcomes over genuine understanding.

In this context, the risk of intellectual passivity emerges. When answers are generated automatically and readily available, the motivation to question, challenge assumptions, and critically evaluate information may gradually diminish. Studies suggest that the use of AI without appropriate pedagogical guidance can weaken students' reflective thinking and cognitive engagement (Holmes et al., 2019). Consequently, the learning process—which should be characterised by inquiry, intellectual struggle, productive error, and critical reflection—risks being reduced to the passive consumption of information.

Furthermore, this disruption has significant implications for the development of independent thinking. Excessive reliance on AI may contribute to the weakening of analytical skills, evaluative judgment, and autonomous decision-making capacities. Within this context of educational disruption, there is a risk that education may no longer cultivate reflective and thinking subjects but instead produce individuals who function merely as "cognitive operators"—those proficient in using technology but lacking depth of reflection, critical judgment, and intellectual autonomy. In other words, learners risk becoming "operators" rather than "thinkers."

Thus, the age of artificial intelligence presents a profound paradox for education. On the one hand, AI offers tremendous opportunities to expand access to knowledge and enhance the effectiveness of learning. On the other hand, without an appropriate pedagogical framework, this technology may undermine the very essence of education: the formation of human beings as critical, autonomous, and reflective thinkers. Therefore, the central challenge for education in this era of disruption lies not merely in integrating technology into learning processes but in ensuring that AI is directed toward cultivating human intellectual capacities rather than simply optimising educational efficiency.

DISCUSSION

A Critical Perspective on Conventional Educational Goals

In practice, education has long been conceived within a relatively narrow framework, primarily as a process of knowledge transmission (Denar, 2021). Within this paradigm, education is conceived as a process of transmitting knowledge from teachers to students, with success measured by the extent to which students can reproduce that knowledge (Freire, 1970; Biesta, 2015). This model is commonly referred to as the *banking model* of education, in which students are regarded as "empty vessels" into which knowledge is deposited, rather than as active participants in the construction of meaning.

Consistent with this paradigm, education has also tended to focus on grades and examination results. Assessment systems that emphasise quantitative measures have made numerical scores the primary indicator of learning success. Within this framework, the learning process is reduced to the pursuit of measurable academic achievement, while the reflective, critical, and existential dimensions of education are often neglected (Biesta, 2015). Consequently, education places greater emphasis on what students know than on how they think, reason, and exercise judgment (Aoun, 2017).

The limitations of this paradigm become even more apparent when viewed from the perspective of cultivating students' thinking capacities. Education oriented toward knowledge transmission and academic achievement tends to provide limited opportunities for students to develop critical reflection. The learning process emphasises the reception of

information rather than dialogue, inquiry, and critical reasoning. As a result, students' ability to examine assumptions, question established knowledge, and construct understanding independently remains underdeveloped (Freire, 1970).

Furthermore, this paradigm does little to foster intellectual autonomy. Students become accustomed to following predetermined structures, responding according to established standards, and relying on external authorities as the primary source of truth. Over time, this condition can hinder the development of individuals as independent thinkers and autonomous decision-makers (Biesta, 2015). In this respect, education fails to cultivate reflective individuals who possess intellectual autonomy and the capacity for independent judgment. This critique of conventional educational goals becomes even more relevant in the context of rapid advances in artificial intelligence (AI). What has long been regarded as the core of education—acquiring information and solving problems—can now be performed by AI systems more quickly and efficiently (Kasneci et al., 2023). AI can provide instant answers, synthesise information, and even generate coherent, argumentative texts within seconds. In this context, an educational orientation that emphasises knowledge acquisition alone is no longer sufficient.

Furthermore, if education remains confined to this traditional paradigm, it risks being reduced to the development of technical competencies that AI can readily replicate. Students are trained to master the very capacities that are most susceptible to automation, while uniquely human capabilities—such as critical reflection, ethical judgment, and contextual meaning-making—receive insufficient attention (Holmes et al., 2019).

In this context, it can be argued that education is facing a crisis of purpose. Educational goals centred on knowledge transmission and academic achievement are no longer sufficient to address the challenges of the AI era. As machines become increasingly capable of acquiring, processing, and generating knowledge at unprecedented speed, the essence of education can no longer be defined by what people know, but by how they think, exercise judgment, and make meaning of that knowledge. Consequently, education must move beyond the traditional paradigm toward cultivating individuals who are reflective, critical, and intellectually autonomous.

Critical Thinking and Intellectual Autonomy as Essential Competencies

The disruption brought about by artificial intelligence (AI) demands a fundamental reorientation of educational goals. Within the conventional paradigm, education has primarily focused on the acquisition of knowledge. However, in the AI era—where access to information is instantaneous and intelligent systems can perform many cognitive tasks—this orientation is no longer sufficient. Education can no longer be understood merely as a process of accumulating knowledge; rather, it must be conceived as an endeavour to cultivate individuals who are capable of critical, independent, and reflective thinking (Biesta, 2015; Holmes et al., 2019). Accordingly, the central concern of education is no longer what people know, but how they come to know, understand, and make meaning of that knowledge.

Critical thinking is widely recognised as one of the essential competencies for the twenty-first century, enabling students to evaluate information, construct reasoned arguments, and make informed decisions (Facione, 2015; Paul & Elder, 2020; Partnership for 21st Century Learning, 2019). In an information landscape characterised by abundant data and rapid dissemination, individuals must be able to distinguish reliable from unreliable information, evaluate its credibility, and understand its broader implications. Critical thinking extends beyond analytical skills to encompass epistemic dispositions such as constructive scepticism, openness to revision, and a willingness to engage with multiple perspectives (Facione, 2015; OECD, 2018; Brookfield, 2012). In this context, students must not only be able to find answers but also to critically examine them by asking where they originate, how trustworthy they are, and what consequences they may entail. Such capacities are essential to ensure that individuals do not become overly dependent on automated systems at the expense of reflective judgment.

In addition, education must be directed toward fostering intellectual autonomy. Intellectual autonomy refers to an individual's capacity to think independently, exercise reasoned judgment, and avoid passive dependence on external authorities, whether teachers, institutional systems, or AI technologies. In the AI era, where answers can be generated instantly through intelligent algorithms, there is a growing risk that students will lose the initiative to think for themselves. Therefore, education should create opportunities for students to engage fully in the process of inquiry by

formulating questions, grappling with uncertainty, exploring alternative perspectives, and constructing understanding through critical reflection (Biesta, 2015; Kasneci et al., 2023). This process is consistent with transformative learning theory, which views critical reflection as the foundation for transforming meaning perspectives and fostering deeper understanding (Mezirow, 2000). Intellectual autonomy, in this sense, constitutes the foundation of genuine freedom of thought.

Furthermore, the reorientation of educational goals must also encompass the cultivation of reflective and ethical capacities. Reflection enables individuals not only to understand knowledge at a technical level but also to relate it to their lived experiences, personal values, and broader social contexts. In this sense, education plays a vital role in fostering both self-awareness and social consciousness. Equally important is the ethical dimension, as the development and use of AI are never value-neutral. Decisions made by or through AI systems inevitably carry moral consequences for both individuals and society (Holmes et al., 2019; UNESCO, 2021). Therefore, education must equip students with the capacity for ethical judgment—the ability to discern not only what can be done, but also what ought to be done.

Three Dimensions of Educational Reorientation: Epistemological, Anthropological, and Ethical

This reorientation can be more fully understood through three interrelated dimensions: epistemological, anthropological, and ethical. The first is the epistemological dimension, which concerns the nature of knowledge and how human beings come to know. In the AI era, knowledge can no longer be understood as a stable body of information stored and transmitted linearly; rather, it is increasingly viewed as a dynamic, continuously evolving network. This perspective resonates with the concept of connectivism, which emphasises the importance of networks and connections in the learning process, although some scholars caution against treating it as a comprehensive theory of learning (Goldie, 2016). Within this context, education should prioritise the capacity to navigate, evaluate, and make meaning of complex knowledge rather than merely mastering its content.

The second is the anthropological dimension, which addresses the fundamental question of what it means to be human in the age of artificial intelligence. As AI systems become increasingly capable of imitating—and in some domains surpassing—human cognitive functions, human uniqueness can no longer be defined solely in terms of rational intelligence. Rather, human beings must be understood as reflective persons who possess self-awareness, are capable of meaning-making, and bear moral responsibility (Biesta, 2015; Harari, 2016). This perspective is consistent with Bostrom's (2014) argument that the advancement of artificial intelligence not only creates unprecedented opportunities but also requires humanity to develop the wisdom and responsibility necessary to ensure that technological progress remains under human control and is directed toward the common good. Accordingly, education must cultivate wisdom, moral judgment, and the capacity for responsible decision-making. In this respect, education plays a crucial role in preserving and nurturing these distinctly human capacities so that the instrumental logic of technology does not diminish them.

The third is the ethical dimension, which concerns how knowledge is applied in human life. In an increasingly digital and AI-driven world, technical competence alone is no longer sufficient. Individuals require a moral compass to guide the responsible use of knowledge and technology. The UNESCO Recommendation on the Ethics of Artificial Intelligence (2021) and the AI ethics framework proposed by Floridi & Cowls (2022) emphasize principles such as fairness, transparency, accountability, respect for human dignity, and sustainability as essential foundations for the development and use of AI. Accordingly, education must integrate these ethical principles not as peripheral or supplementary elements, but as integral components of the learning process.

Thus, the reorientation of educational goals in the age of artificial intelligence is no longer optional but imperative. Education can no longer remain anchored in a paradigm centered on the accumulation of knowledge; instead, it must cultivate individuals who are capable of critical, independent, reflective, and ethical thinking. Within this framework, the purpose of education extends beyond preparing learners to adapt to technological change. It also seeks to empower them to responsibly shape the direction of that change for the common good. Only through such a reorientation can

education remain relevant, meaningful, and genuinely human in the face of accelerating technological disruption.

CONCLUSION

This study demonstrates that the disruption brought about by artificial intelligence presents not only technological challenges but also raises the most fundamental questions of education: What is its purpose, and what kind of human person should it seek to form? As the conventional educational paradigm, centered on knowledge transmission and academic achievement, becomes increasingly inadequate in the context of cognitive automation, education faces a crisis of purpose that calls for a fundamental reorientation toward cultivating students' critical thinking, intellectual autonomy, and reflective-ethical capacities.

From a theological perspective, this reorientation extends beyond functional considerations and is rooted in an understanding of the human person as *imago Dei*—a being endowed with dignity, freedom, and responsibility, rather than merely a knower. Accordingly, education must not reduce learners to mere operators of technology but should instead form subjects who think, exercise judgment, and act with conscious responsibility. In this light, critical thinking and intellectual autonomy are not merely competencies; they are expressions of human dignity itself.

The implication is that education in the age of artificial intelligence should be understood as a process of humanization that reaffirms human identity—a formative space in which individuals develop into whole persons capable of thinking critically, acting freely, and living according to truth and love. In this way, education continues to fulfill its prophetic vocation: safeguarding humanity from losing its identity amid the growing dominance of technology while guiding social and technological development toward a more humane and dignified future. Ultimately, the future of education will not be determined by how advanced artificial intelligence becomes, but by the extent to which education succeeds in preserving and cultivating the uniquely human capacities for critical judgment, moral responsibility, and wisdom.

REFERENCE

Aoun, J. E. (2017). *Robot-Proof: Higher Education in the Age of Artificial Intelligence*. The MIT Press.

Biesta, G. J. J. (2015). *Good Education in an Age of Measurement: Ethics, Politics, Democracy*. Routledge. https://doi.org/https://doi.org/10.4324/9781315634319

Bostrom, N. (2014). *Superintelligence: Paths, Dangers, Strategies*. Oxford University Press.

Brookfield, S. D. (2012). *Teaching for Critical Thinking: Tools and Techniques to Help Students Question Their Assumptions*. Jossey-Bass.

Denar, B. (2021). Tantangan Pendidikan dalam Era Post Truth. In K. Nema, B. Denar, & F. Gions (Eds.), *Pedagogi Kemasyarakatan* (pp. 193–205). JPIC OFM.

Facione, P. A. (2015). *Critical Thinking : What It Is and Why It Counts*. Insight Assessment.

Floridi, L., & Cowls, J. (2022). A Unified Framework of Five Principles for AI in Society. In S. Carta (Ed.), *Machine Learning and the City*. Jossey-Bass. https://doi.org/https://doi.org/10.1002/9781119815075.ch45

Freire, P. (1970). *Pedagogy of the Oppressed*. Continuum.

Gobble, M. M. (2016). Defining Disruptive Innovation. *Research-Technology Management*, *59*(4), 66–71. https://doi.org/10.1080/08956308.2016.1185347

Goldie, S. G. G. (2016). Connectivism : A Knowledge Learning Theory for the Digital Age? *Medical Teacher*, *38*(10), 1–6. https://doi.org /10.3109/0142159X.2016.1173661.

Harari, Y. N. (2016). *Homo Deus: A Brief History of Tomorrow*. Harper.

Holmes, W., Bialik, M., & Fadel, C. (2019). *Artificial Intelligence in Education: Promises and Implications for Teaching and Learning*. Center for Curriculum Redesign.

Kasneci, E., Sessler, K., Stefan, K., Fischer, F., Gasser, U., Groh, G., Stephan, G., Poquet, O., Sailer, M., Schmidt, A., & Seidel, T. (2023). ChatGPT for Good ? On Opportunities and Challenges of Large Language Models for Education. *Learning and Individual Differences*, *103*, 1–13. https://doi.org/https://doi.org/10.1016/j.lindif.2023.102274.

Luckin, R. (2018). *Machine Learning and Human Intelligence: The Future of Education for the 21st Century*. UCL Institute of Education Press, University College.

Mezirow, J. (2000). *Learning as Transformation: Critical Perspectives on a Theory in Progress*. Jossey-Bass.

Ninghardjanti, P., Umam, M. C., Subarno, A., Winarno, W., Langgi, N. R., & Widodo, J. (2025). Evaluating the Impact of AI on the Critical Thinking Skills Among the Higher Education Students by Combining the TAM Model and Critical Thinking Theory. *Front. Educ.*, *10*(17), 1–23. https://doi.org/10.3389/feduc.2025.1719625.

OECD. (2018). *The Future of Education and Skills Education 2030*. OECD Publishing.

Partnership for 21st Century Learning. (2019). *Framework for 21st Century Learning*.

Paul, R., & Elder, L. (2020). *Critical Thinking Tools for Taking Charge of Your Learning and Your Life*. Rowman & Littlefield Publishers.

Prahani, B. K., Rizki, I. A., Jatmiko, B., Suprapto, N., & Amelia, T. (2022). Artificial Intelligence in Education Research During the Last Ten Years: A Review and Bibliometric Study. *International Journal of Emerging Technologies in Learning*, *17*(8), 169–188. https://doi.org/10.3991/ijet.v17i08.29833

Selwyn, N. (2019). *Should Robots Replace Teachers? AI and the Future of Education*. Polity Press.

Shukor, A. N. F., & Osman, K. (2025). The Role of Artificial Intelligence in Developing Critical Thinking among Science Students: A Systematic Literature Review. *International Journal of Academic Research in Progressive Education and Development*, *14*(3), 164–176. https://doi.org/10.6007/IJARPED/v14-i3/25576.

UNESCO. (2021). *Recommendation on the Ethics of Artificial Intelligence*. UNESCO Publishing.

Williamson, B. (2017). *Big Data in Education: The Digital Future of Learning, Policy and Practice*. SAGE Publications.

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*(39), 1–27. https://doi.org/https://doi.org/10.1186/s41239-019-0171-0.

AUTHOR PROFILE



Benediktus Denar is a Catholic theologian from Manggarai, East Nusa Tenggara, Indonesia. He was born in Sirimese on June 18, 1986. He completed his primary education at SDK Sirimese and his junior secondary education at SMP Kemasyarakatan Ndoso before continuing his studies at the Yohanes Paulus II Minor Seminary in Labuan Bajo. He earned his Bachelor's degree in Philosophy in 2011 and his Master's degree in Theology in 2015 from the Catholic School of Philosophy (STFK) Ledalero. Since 2016, he has served as a lecturer at STIPAS St. Sirilus

Ruteng. He obtained his Doctorate in Theology from the School of Philosophy and Theology (STFT) Widya Sasana Malang in 2024. His areas of research include public theology, ecotheology, and contextual theology, with particular attention to issues of land, extractivism, development, and social justice from a postcolonial perspective.



Theresia Yovita Cendana Sari adalah kandidat doktor Teologi di Sekolah Tinggi Filsafat Teologi (STFT) Widya Sasana Malang. Ia lahir di Pasuruan, Jawa Timur, pada 5 September 1986. Pendidikan dasarnya ditempuh di SD Negeri Randusari II Pasuruan, dilanjutkan ke SMP Negeri 1 Pasuruan dan SMA Negeri 1 Pasuruan. Ia meraih gelar Sarjana Katekese Pastoral dari STP IPI Malang pada 2009, kemudian gelar Magister Teologi Pastoral dari kampus yang sama pada 2016, dan kini tengah menyelesaikan disertasi doktoralnya di STFT Widya Sasana Malang.
Sebelum menjadi dosen tetap, ia mengajar Pendidikan Agama Katolik di SMA Negeri 8 Malang, SMAK Cor Jesu Malang, dan SMAK St. Albertus Malang, serta mengajar Pendidikan Agama Katolik Dasar di Universitas Kanjuruhan dan Universitas Merdeka Malang. Saat ini ia menjadi dosen di STAKat Negeri Pontianak, dan kerap menjadi pembicara dalam pelatihan guru. Minat risetnya meliputi Pendidikan Agama Katolik, guru agama Katolik, katekese kontekstual, kepemimpinan perempuan dalam Gereja, dan teologi feminis.

CHAPTER IX

The Cognitive Debt of Generative AI in Higher Education English Academic Writing Pedagogy: A Systematic Literature Review

Totok Dewayanto dan Yuliati
Department of Accounting, Faculty of Economics and Business,
Universitas Diponegoro, Semarang, Indonesia
Sekolah Tinggi Ilmu Ekonomi Pelita Nusantara - Indonesia

Correspondent Author: totokdewayanto@lecturer.undip.ac.id

ABSTRACT

Introduction: The integration of Generative Artificial Intelligence (GAI) in higher education is frequently framed optimistically as an extension of human cognition. However, recent literature indicates significant risks of intellectual degradation. This study aims to evaluate the pedagogical impacts of AI in academic writing through a critical lens, focusing on the Extended Mind Hypothesis (EMH) and the Technology Acceptance Model (TAM).

Methods: This study employs a Systematic Literature Review (SLR) methodology following the PRISMA 2020 protocol, drawing on 31 primary scientific articles (2020–2026) indexed in Scopus, Web of Science, and Google Scholar, with a specific focus on English Academic Writing (EAP) pedagogy.

Results: The synthesized data consistently indicate that intensive use of Large Language Models (LLMs) correlates with "Cognitive Debt," characterized by a significant drop in students' long-term memory retention and in their originality in writing. A marked "Verification Drift" is observed where students passively accept AI outputs. TAM fails to explain pathological dependencies, while EMH overlooks the cognitive atrophy of internal mental functions.

Discussion: This paper formulates two Middle Range Theory (MRT) propositions emphasizing "Scaffolded Adversarial Prompting" to mitigate these risks. To the best of the author's knowledge, based on the analyzed corpus, higher education policies must transition from mere technology adoption to logical-pedagogical auditing and Human-in-the-Loop (HITL) models to preserve human agency and authentic authorship.

Keywords: *Generative AI, Academic Writing, Cognitive Debt, ZPD, TAM, EMH, Relational Ethics.*

INTRODUCTION

The global higher education landscape is currently experiencing a historic disruption driven by the rapid emergence and widespread adoption of Generative Artificial Intelligence (GAI), particularly Large Language Models (LLMs) such as ChatGPT (Farazouli, 2026). Within the domain of English for Academic Purposes (EAP) and academic writing pedagogy, these tools have rapidly transitioned from simple grammar-checking utilities into active cognitive partners. Today, GAI can synthesize complex arguments, generate academic prose, outline structures, and provide real-time stylistic feedback (Wu, 2024).

However, this rapid integration is consistently documented as a "double-edged sword" (Aria & Cuccurullo, 2017). On the one hand, GAI offers unprecedented opportunities for personalized instruction, immediate formative feedback, and administrative efficiency for teachers working with large student cohorts. On the other hand, the ease of delegating writing tasks to automated systems threatens the integrity of students' deep critical thinking, metacognitive monitoring, and authentic authorial voice (Bouchard, 2025).

The core of EAP pedagogy has traditionally relied on the cognitive struggle of translating thoughts into structured text. This process is not merely communicative but constructive; it is through the act of writing that students clarify their understanding, build complex mental schemas, and consolidate new knowledge. When GAI tools are used to automate the drafting and synthesis phases of writing, they remove this productive difficulty. Without this cognitive effort, students are at risk of engaging in superficial learning, producing high-quality essays without developing the underlying cognitive or linguistic competencies (Delikoura et al., 2025).

For decades, the adoption and integration of technology in education have been explained through established frameworks. The Technology Acceptance Model (TAM), proposed by Davis (1989), posits that technology use is driven primarily by perceived usefulness and perceived ease of use. While TAM remains highly popular in educational technology research, it possesses critical limitations when applied to GAI. TAM operates on an

adoption-centric perspective, focusing on *how* and *why* users adopt a tool, but it is silent on the qualitative cognitive consequences of that adoption. It fails to distinguish between productive use that enhances learning and pathological dependency that leads to cognitive atrophy.

Similarly, the Extended Mind Hypothesis (EMH), formulated by Clark and Chalmers (1998), provides a philosophical foundation for viewing GAI as a functional extension of human cognition. EMH suggests that external tools, when tightly integrated into human tasks, become part of the cognitive system itself. However, applying the EMH to GAI overlooks a fundamental difference: unlike static tools (such as notebooks or calculators), LLMs possess active, semantically-like capabilities. They do not merely store or calculate; they generate meaning. When students delegate semantic synthesis to an LLM, they are not extending their minds; they are offloading the very core of their reasoning. This complete delegation can lead to "Cognitive Debt" (Jamshaid, 2025), in which the user's internal cognitive capacity regresses due to disuse.

The immediate availability of high-quality automated text generation has created a major pedagogical dilemma in higher education. Students increasingly utilize GAI tools for academic writing assignments, yet university curricula and assessment methods remain designed for a pre-AI era. Pervasive GAI usage has led to two critical phenomena: "Verification Drift," in which students passively accept AI-generated outputs without critical evaluation, and "Epistemic Complacency," in which students mistake the fluent output of an AI for their own deep understanding (Bouchard, 2025).

Current institutional policies are often polarized between futile attempts to ban the technology and uncritical integration. There is a lack of a clear, pedagogically grounded framework that structures the human-AI interaction in a way that preserves student learning. This study addresses this gap by conducting a systematic review of GAI in EAP pedagogy and proposing a Middle Range Theory (MRT) framework that reconciles the tensions between cognitive offloading and educational integrity.

To address the current gaps in the literature, this Systematic Literature Review is guided by the following three research questions:

- **Research Question 1 (RQ1 - Diagnostic-Theoretical):** To what extent does the duration and intensity of routine GAI usage correlate with long-term memory retention and writing originality in undergraduate students?

- **Research Question 2 (RQ2 - Explanative-Capability):** How do "Instant Answer" mechanisms in LLMs disrupt the scaffolding process at the boundary of a student's Zone of Proximal Development (ZPD)?
- **Research Question 3 (RQ3 - Prescriptive-Strategic):** What class instructional models using Scaffolded Adversarial Prompting are effective in strengthening critical AI literacy and mitigating Verification Drift?

To bridge the gap between abstract grand theories (TAM, EMH) and the concrete realities of GAI classrooms, this study develops and tests two core MRT propositions:

- **Proposition 1 (P1 - Scaffolded Adversarial Prompting):** IF GAI tools are restricted to providing heuristic hints instead of direct answers, THEN students' metacognitive monitoring will increase BECAUSE it forces deep System 2 cognitive processing within the ZPD.
- **Proposition 2 (P2 - Destructive Cognitive Offloading):** IF semantic reasoning is fully delegated to LLMs without retrieval practice, THEN long-term memory retention will decline BECAUSE the elimination of germane cognitive load weakens schema consolidation.

THEORETICAL FRAMEWORK

Critical Review of Grand Theories

The theoretical foundation of this study is grounded in a critical evaluation of how existing theories explain or fail to explain the cognitive integration of GAI in EAP pedagogy.

The Technology Acceptance Model (TAM)

TAM (Davis, 1989) has long been the dominant model for explaining user acceptance of information systems. It asserts that behavioural intention to use a technology is determined by perceived usefulness (PU) and perceived ease of use (PEOU). In the context of GAI in education, TAM successfully explains the rapid, viral adoption of tools like ChatGPT. The PEOU of a chat interface is extremely high, and the PU for generating a draft is immediate.

However, TAM is fundamentally limited as an educational evaluation tool. It treats "use" as a uniform positive outcome. In education, however, the goal is not technology adoption, but cognitive development. A student who uses GAI to write an essay may score high on PEOU and PU, but their actual learning

may be zero. TAM fails to capture this tension, making it an insufficient framework for designing pedagogically sound AI integrations.

The Extended Mind Hypothesis (EMH)

EMH (Clark & Chalmers, 1998) argues that the skull or the skin does not limit the mind's boundaries. If an external resource is coupled with a human subject in a way that is reliable, accessible, and automatically trusted, it can be considered a functional part of the cognitive system.

When applied to GAI, EMH is frequently used to argue that delegating writing and thinking to AI is no different from using a notebook or a search engine. We critique this view. GAI tools differ from traditional cognitive extensions in that they perform semantic processing autonomously. A notebook stores thoughts; a search engine retrieves documents, but an LLM generates synthesis. When a student delegates semantic synthesis to an LLM, they are not extending their cognitive capacity; they are replacing it. This replacement leads to "Cognitive Debt," in which the user experience's internal cognitive structures atrophy due to a lack of effortful engagement.

Supporting Theories

Cognitive Load Theory (CLT)

CLT (Sweller, 1988) explains learning in terms of the limitations of working memory and the construction of schemas in long-term memory. It identifies three types of cognitive load:

- **Intrinsic Load:** The inherent difficulty of the material.
- **Extraneous Load:** The mental effort required by the way the material is presented.
- **Germane Load:** The productive mental work dedicated to processing and constructing schemas.

In a GAI-mediated writing task, the immediate generation of text by an LLM reduces both extraneous and intrinsic load. However, it also eliminates germane load. Without germane load, schema construction does not occur, and no actual learning takes place.

The Zone of Proximal Development (ZPD)

ZPD (Vygotsky, 1978) defines the distance between what a learner can do independently and what they can do with assistance (scaffolding). Effective scaffolding is adaptive and temporary; it guides the learner toward

independence. LLMs in their default configuration do not provide scaffolding. Instead of guiding the student through the writing process, they provide the final answer immediately. This "Direct Output" bypasses the ZPD entirely, preventing the student from developing independent writing skills.

2.3 Formulation of Middle Range Theory (MRT)

To address these theoretical gaps, we propose a Middle Range Theory that focuses on the specific mechanisms of human-AI cognitive interaction.

Table 1: MRT Propositions and Boundary Conditions

Proposition	Theoretical Mechanism	Boundary Conditions	Falsification Criteria
P1: Scaffolded Prompting	Restricting GAI to heuristic hints forces System 2 processing, keeping the task within the ZPD.	Active guidance, basic literacy, and high-quality models.	Metacognitive monitoring fails to improve despite guided prompting.
P2: Destructive Offloading	Total delegation of semantic synthesis to LLMs leads to memory decline due to a lack of germane load.	High frequency of use, passive acceptance, and lack of retrieval practice.	Long-term memory remains stable despite continuous passive delegation.

METHODOLOGY

This study employs a Systematic Literature Review (SLR) methodology following the PRISMA 2020 guidelines to ensure a rigorous, transparent, and replicable review process. A post-positivist paradigm is adopted, focusing on identifying empirical patterns and causal mechanisms in the literature that validate or modify our proposed MRT propositions. To define the scope of the review and set clear inclusion and exclusion criteria, we established the PICOS protocol.

Table 2: PICOS Operational Protocol

Element	Definition	Inclusion Criteria	Relevance to MRT
Population (P)	Higher Education Students & Educators	Undergraduate/postgraduate students engaged in EAP or academic writing.	Focuses on the primary users and environments

			where cognitive debt occurs.
Intervention (I)	Generative AI / LLM Tools	Use of LLMs (e.g., ChatGPT, GPT-4, Claude) for writing assistance, feedback, or tutoring.	Represents the causal variable in P1 and P2.
Comparison (C)	Traditional Writing Instruction	Non-AI baselines, peer review, manual writing, or static feedback tools.	Establishes the baseline for measuring cognitive decline or improvement
Outcome (O)	Cognitive & Writing Performance	Long-term memory retention, originality, critical thinking, and metacognitive monitoring.	Represents the target dependent variables of interest.
Study Design (S)	Empirical & Systematic Studies	Randomized controlled trials, mixed-methods, longitudinal studies, and systematic reviews.	Ensures data quality and validity of findings.

3.3 Search Strategy

A systematic search was conducted in March 2026 across three major database platforms: Scopus, Web of Science, and Google Scholar. The exact search strings used were designed to capture the intersection of GAI, academic writing, and cognitive outcomes.

Table 3: Database Search Strings and Results

Database	Exact Search String	Target Field	Search Date	Results (N)
Scopus	("Generative AI" OR "ChatGPT" OR "LLM") AND ("English Academic Writing" OR "EAP") AND ("Cognitive" OR "Pedagogy")	Title-Abstract-Keywords	15/03/2026	215
Web of Science	("Generative AI" OR "ChatGPT" OR "LLM") AND ("English Academic Writing" OR "EAP") AND ("Cognitive" OR "Pedagogy")	Topic (TS)	16/03/2026	132

Database	Exact Search String	Target Field	Search Date	Results (N)
Google Scholar	"Generative AI English Academic Writing higher education cognitive debt" (First 100 relevant matches)	All Fields	17/03/2026	153
Total Initial Records				**500**

4 PRISMA Selection Flow

The study selection followed the four-phase PRISMA 2020 flow, resulting in a final corpus of 31 primary studies.

Records identified through database searching (n = 500)

↓

Records after duplicate removal and abstract screening (n = 352)

↓

Full-text articles assessed for eligibility (n = 202)

Excluded with reasons (n = 171): low quality (n = 120); non-academic/pedagogical scope (n = 51).

↓

Final included primary studies for synthesis (n = 31)

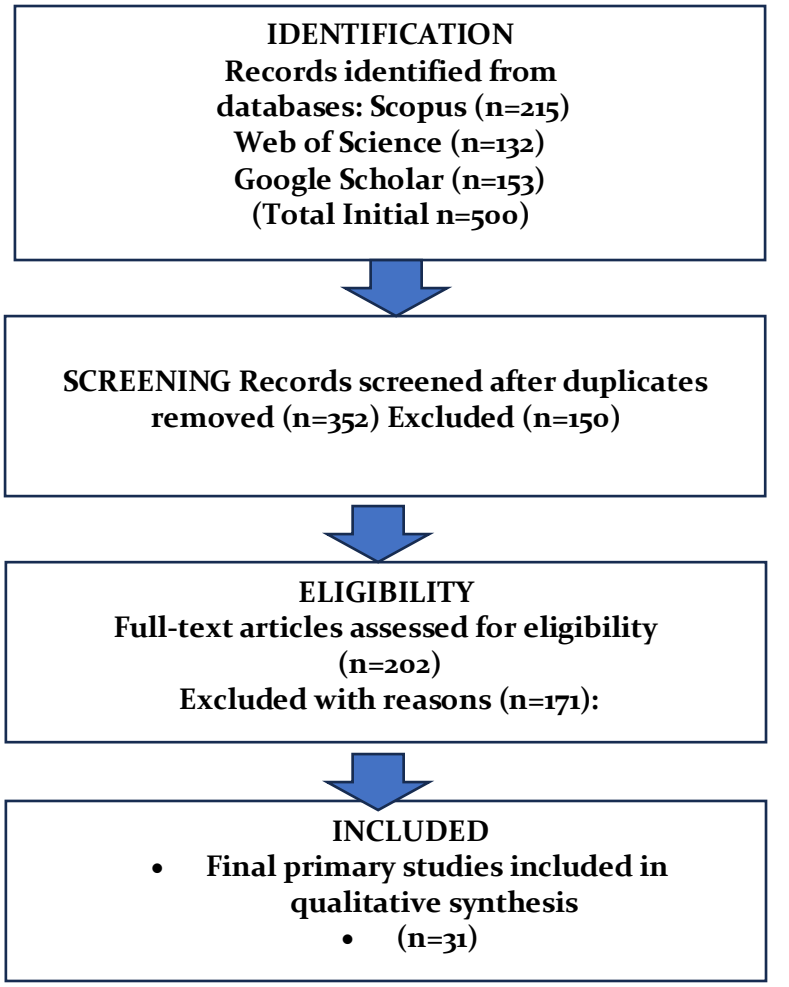


Figure: PRISMA 2020 Flow Diagram of Selection Process

3.5 Quality Appraisal and Risk of Bias

The selected studies were appraised using standard critical appraisal tools. Quantitative and experimental studies were evaluated using the Joanna Briggs Institute (JBI) checklists, while qualitative and conceptual works were assessed using the Critical Appraisal Skills Program (CASP) guidelines.

To ensure reliability, an Intra-Rater Reliability (IRR) check was performed. The primary author re-evaluated a random subset of 25% of the articles (8 papers) after a two-week interval. The Cohen's Kappa coefficient was 0.85, indicating high consistency in study selection and quality assessment.

Risk of bias was addressed at two levels. At the study level, we assessed reliance on student self-reports vs objective cognitive measures, prioritising studies with objective, controlled experimental data. At the review level, publication bias was mitigated by seeking out grey literature and doctoral dissertations from institutional repositories (such as the DiVA portal) to ensure negative findings were represented.

RESULTS AND DISCUSSION

4.1 Descriptive Analysis of the Corpus

The systematic review of the 31 selected articles shows a clear concentration of research in 2024-2026, marking the post-release phase of GAI tools.

Visualizing Publication Distribution

The annual distribution of the included publications shows a significant upward trend, peaking in 2024 and 2025, which reflects the academic community's response to the ChatGPT shock.

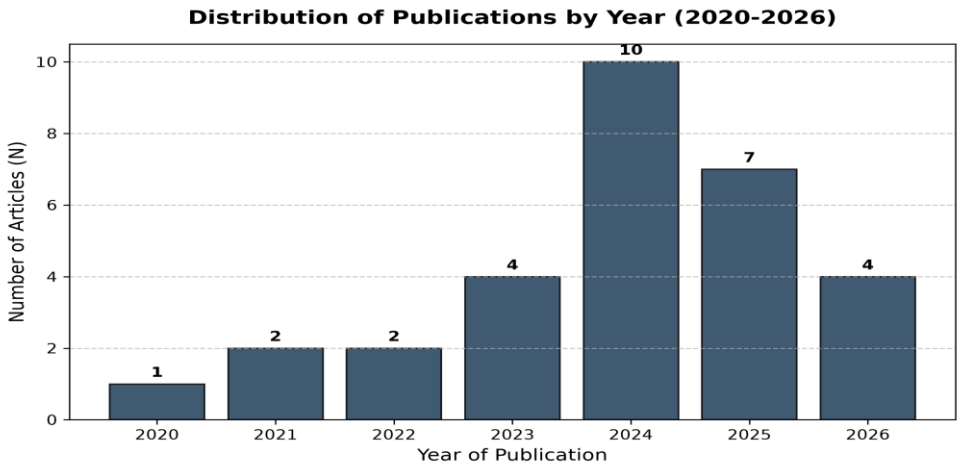


Figure: Annual Scientific Production

Science Mapping and Theme Clusters

To analyze the conceptual relationships in the literature, we mapped the keyword co-occurrence network of the 31 articles.

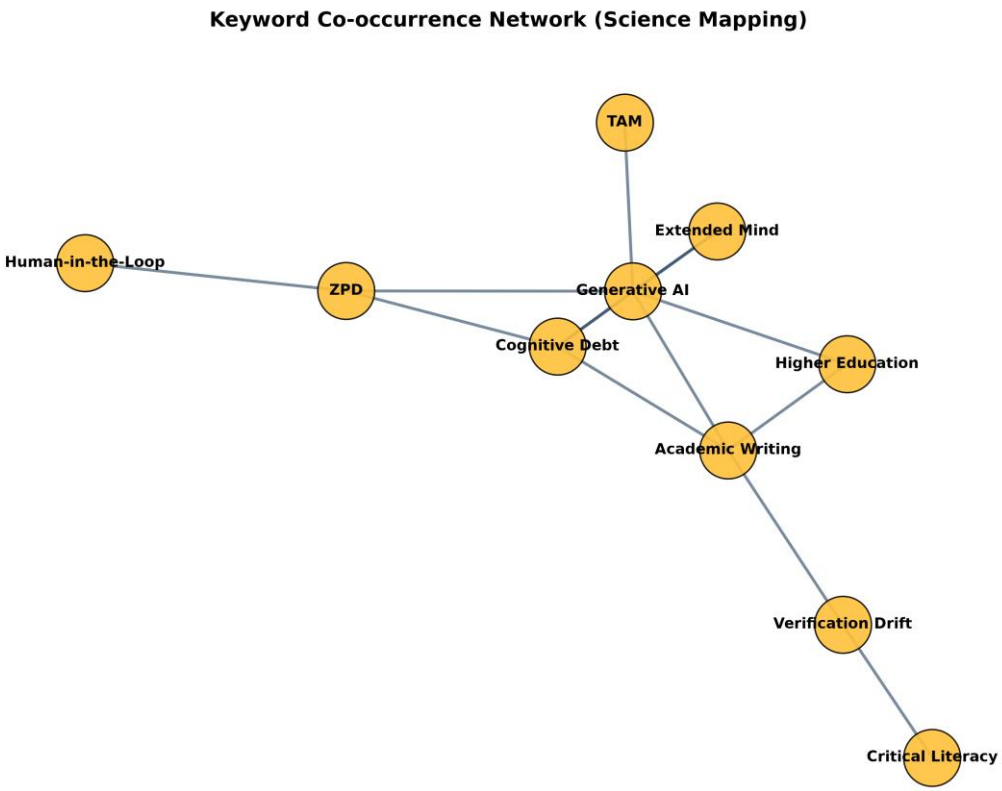


Figure: Keyword Co-occurrence Network

The network reveals three major thematic clusters:

1. **The Cognitive Atrophy Cluster:** Centered on "Cognitive Debt," "Memory Retention," and "Extended Mind." It contains studies warning against the long-term cognitive costs of unmediated AI offloading (e.g., Jamshaid, 2025).
2. **The Pedagogical Scaffolding Cluster:** Centered on "ZPD," "Human-in-the-Loop," and "Adversarial Prompting," focusing on active instructional strategies (e.g., Alimardani, 2025).
3. **The User Acceptance Cluster:** Centered on "TAM," "Ease of Use," and "Higher Education," highlighting adoption trends but increasingly noting the discrepancy between adoption rates and learning outcomes.

4.2 Synthesis of Primary Findings

We compiled the key data points from the primary studies to evaluate our MRT propositions.

Table 4.1: Master Synthesis Data Table

Study	Research Design	Context & Sample	Key Findings	Cognitive Impact
Jamshaid (2025)	Mixed-methods	Pakistan: 150 students	Documented a significant decline in 48-hour recall scores ($p < 0.001$) among regular GAI users.	High Cognitive Debt
Farazouli (2026)	Qualitative / Case Study	Sweden: 45 students	Students experienced authorial alienation and a loss of personal voice when using GAI.	High Verification Drift
Alimardani (2025)	Controlled Experiment	Australia; 80 Law students	Using adversarial prompts (critiquing AI outputs) improved assignment grades and critical reasoning.	Metacognitive Gain
Chun et al. (2025).	Quantitative	South Korea; 200 students	Self-regulated learning models using AI assistants had positive adoption rates but mixed retention scores.	Mixed Outcome
Delikoura et al. (2025).	Systematic Review	Global: 45 studies analyzed	Highlighted that LLM direct outputs bypass ZPD, leading to superficial learning.	ZPD Failure

4.3 Verification and Discussion of MRT Propositions

Evaluating Proposition P1: Scaffolded Adversarial Prompting

Proposition P1 posits that restricting GAI to providing heuristic hints instead of direct answers will improve metacognitive monitoring by forcing System 2 processing. The empirical findings strongly support this proposition.

In Alimardani's (2025) study, students who were forced to engage with AI in an adversarial manner (verifying claims, identifying biases, and rewriting drafts) showed higher levels of critical engagement and better performance on assessment rubrics.

Conversely, when AI is allowed to produce full drafts, students fall prey to "Verification Drift" (Farazouli, 2026). They become passive editors, accepting incorrect or biased claims because of the model's fluent, authoritative tone. This confirms the mechanism of P1: cognitive friction is necessary to keep GAI use within the student's ZPD.

Evaluating Proposition P2: Destructive Cognitive Offloading

Proposition P2 suggests that delegating semantic reasoning to LLMs without retrieval practice causes a decline in long-term memory. Jamshaid's (2025) mixed-methods study provides direct support for this. The study compared students who wrote essays using GAI with a control group who wrote essays manually or with basic search support. While short-term task completion times were lower for the GAI group, their 48-hour delayed recall scores were significantly lower. This finding aligns with Sweller's Cognitive Load Theory. Because GAI automated the synthesis process, it eliminated the cognitive load associated with it. The students did not have to actively retrieve information or organize schemas, which led to a failure in long-term memory consolidation. This confirms that unmediated offloading leads to Cognitive Debt.

4.4 Theoretical Reconciliation

Based on these findings, we propose a modification to the application of TAM and EMH in educational research.

Table 4.2: Theoretical Position Modifications

Grand Theory	Original Assumption	Modified Boundary Condition (MRT)
TAM	Technology adoption is a direct indicator of educational value.	Perceived usefulness must be balanced with objective measures of cognitive engagement; ease of use can be negatively correlated with schema retention.
EMH	Coupling with external tools extends the mind.	Mind extension is valid only for storage and calculation; delegating semantic processing to AI results in cognitive atrophy and internal debt.

CONCLUSION AND RECOMMENDATIONS

5.1 Summary of Findings

This Systematic Literature Review has evaluated the integration of Generative AI in English Academic Writing pedagogy across 31 primary studies published between 2020 and 2026. The synthesis of findings consistently indicates that unmediated GAI use leads to a trade-off: short-term task efficiency is achieved at the expense of long-term cognitive development and memory retention.

Our proposed Middle Range Theory (MRT) framework has been validated against the empirical data. We have confirmed that:

1. **Proposition P1:** Scaffolded, adversarial interactions with GAI tools maintain task difficulty within the student's Zone of Proximal Development, leading to metacognitive improvements.
2. **Proposition P2:** Unconstrained offloading of semantic synthesis to GAI results in "Cognitive Debt," characterized by declines in memory retention and original thought.

5.2 Pedagogical Implications and Recommendations

Higher education institutions must transition away from uncritical adoption of technology. We offer the following recommendations:

- **Curricular Restructuring:** Academic writing courses should incorporate "Scaffolded Adversarial Prompting" as a core literacy skill. Students must be assessed not on the final text, but on their process of prompting, verification, and critical revision.
- **Assessment Design:** High-stakes assessments should move toward hybrid models, separating the drafting phase (which may involve AI scaffolding) from final evaluation (which should include supervised, closed-book writing to ensure internal memory consolidation).
- **Algorithmic Auditing:** Universities should establish guidelines for auditing the educational tools they purchase, demanding transparency regarding the prompt configurations and RLHF alignment of vendor AI models.

5.3 Novelty and Contribution

To the best of the author's knowledge, based on the analyzed corpus of 31 papers, this study contributes to the field by:

- Introducing the concept of "**Cognitive Debt**" to define the long-term cognitive cost of automated semantic synthesis.
- Formulating a **Middle Range Theory** that offers specific, testable propositions for human-AI pedagogical interaction, bridging the gap between abstract philosophy (EMH) and classroom practice.
- Outlining the "**Verification Drift**" mechanism, explaining why students trust and accept incorrect or biased AI outputs.

5.4 Limitations and Future Research

The short-term nature of current empirical studies limits this review. Most research has tracked GAI effects over a single semester. Future research must prioritize longitudinal studies (3–4 years) that use neuroimaging techniques (such as fMRI or EEG) to track neuroplastic changes associated with long-term GAI dependency. Furthermore, research should expand into the Global South to understand how GAI impacts students operating under infrastructure constraints.

REFERENCES

Abadi, M., Chu, A., Goodfellow, I., McMahan, H. B., Mironov, I., Talwar, K., & Zhang, L. (2016). Deep learning with differential privacy. *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*, 308–318. https://doi.org/10.1145/2976749.2978318

Abdul Razak, A. N., Lim, M., Tomeo-Reyes, I., & Li, D. D. (2024). Exploring the capabilities and limitations of generative AI in providing feedback on engineering drawings: A case study. *2024 World Engineering Education Forum - Global Engineering Deans Council (WEEF-GEDC)*. https://doi.org/10.1109/WEEF-GEDC63419.2024.10854935

Akgun, M., & Toker, S. (2024). Evaluating the effect of pretesting with conversational AI on retention of needed information. *arXiv*. https://doi.org/10.48550/arXiv.2412.13487

Akgun, S., & Greenhow, C. (2022). Artificial intelligence in education: Addressing ethical challenges in K-12 settings. *AI and Ethics*, 2(3), 431–440. https://doi.org/10.1007/s43681-021-00096-7

Albuquerque, J., Rienties, B., Holmes, W., & Hlosta, M. (2025). From hype to evidence: Exploring large language models for inter-group bias classification in higher education. *Interactive Learning Environments*, 33(3), 2332–2354. https://doi.org/10.1080/10494820.2024.2408554

AlGhamdi, R. (2024). Exploring the impact of ChatGPT-generated feedback on the technical writing skills of computing students: A masked study.

Education and Information Technologies, 29, 18901–18926. https://doi.org/10.1007/s10639-024-12594-2

Alimardani, A. (2024). Generative artificial intelligence vs. law students: An empirical study on criminal law exam performance. *Law, Innovation and Technology*, 16(2), 777–819. https://doi.org/10.1080/ 17579961.2024. 2392932

Alimardani, A. (2025). Borderline disaster: An empirical study on student usage of GenAI in a law assignment. *IEEE Transactions on Technology and Society*, 6(2), 210–219. https://doi.org/10.1109/TTS.2025.3540978

Alzoubi, Y. I., & Mishra, A. (2024). Green artificial intelligence initiatives: Potentials and challenges. *Journal of Cleaner Production*, 468, 143090. https://doi.org/10.1016/j.jclepro.2024.143090

Awidi, I. T. (2024). Comparing expert tutor evaluation of reflective essays with marking by a generative artificial intelligence (AI) tool. *Computers and Education: Artificial Intelligence*, 6, 100226. https://doi.org/10.1016 /j.caeai.2024.100226

Aria, M., & Cuccurullo, C. (2017). Bibliometrix: An R tool for comprehensive science mapping analysis. *Journal of Informetrics*, 11(4), 959–975. https://doi.org/10.1016/j.joi.2017.08.007

Bouchard, J. (2025). ChatGPT and the separation between knowledge and knower. *Education and Information Technologies*, 30(8), 10091–10110. https://doi.org/10.1007/s10639-024-13249-y

Chun, J., Kim, J., Kim, H., Lee, G., Cho, S., Kim, C., Chung, Y., & Heo, S. (2025). A comparative analysis of on-device AI-driven, self-regulated learning and traditional pedagogy in university health sciences education. *Applied Sciences*, 15, 1815. https://doi.org/10.3390/app15041815

Delikoura, I., Fung, Y. R. M., & Hui, P. (2025). From superficial outputs to superficial learning: Risks of large language models in education. *arXiv*. https://arxiv.org/abs/2509.21972

Duran, V. (2024). Analyzing teacher candidates' arguments on AI integration in education via different chatbots. *Digital Education Review*, (45), 70-85.

Farazouli, A. (2026). *Emerging AI and ethics in higher education: A technology mediation perspective* (Doctoral dissertation, Stockholm University). DiVA portal.

Jamshaid, S. (2025). Cognitive effects of large language models: An empirical study of memory retention, reasoning strategies, and critical thinking among students and professionals at the University of Gujrat, Pakistan. *ACADEMIA International Journal for Social Sciences*, 4(3), 6055–6065.

Nikolopoulou, K. (2026). AI-powered student support: a conceptual paper on ethical AI in education. *Frontiers in Education*, 11, 1801972.

AUTHOR PROFILE



Dr. Totok Dewayanto, SE, MSi, Ak, CA, Cert-DA, was born in Semarang on May 9, 1969. He is a lecturer in the Accounting Study Program at the Faculty of Economics and Business (FEB), Diponegoro University. He completed his undergraduate studies in Accounting at Diponegoro University (Undip), Semarang. He pursued his master's degree in Accounting at the same institution, followed by doctoral studies at the Doctoral Program in Economics (PDIE) at Sebelas Maret University (UNS), Surakarta. He has previously served as a Finance Manager and Internal Auditor at the Diponegoro National Hospital (RSND) in Semarang. The author has published numerous scientific articles in various national and international journals.



Yuliati, S.E., M.M., graduated from Semarang on July 15, 1968. Lecturer in the Management Study Program at Sekolah Tinggi Ilmu Ekonomi Pelita Nusantara. She completed her undergraduate education at the Management Department of the Sultan Agung Islamic University Semarang and continued her master's at the 17 August 1945 University Semarang Management Department.

In higher education, he served as Chair of the Management Study Program at Sekolah Tinggi Ilmu Ekonomi Pelita Nusantara. Currently serving as Deputy Chairperson 2 of Sekolah Tinggi Ilmu Ekonomi Pelita Nusantara. The author has published in several scientific journals at several universities in Semarang.

CHAPTER X

Instant Learning Culture and AI Dependency: Its Impact On Students' Deep Thinking Ability

Berlinda Setyo Yunarti and
STK Santo Yakobus, Merauke, Indonesia

Correspondent Co Author: yohaneshenz@stkyakobus.ac.id

ABSTRACT.
The convergence of an instant-learning culture, driven by algorithm-curated short-form media, and students' uncritical dependence on artificial intelligence (AI) constitutes one of the most pressing pedagogical challenges of the current era. This chapter argues that these two forces interact as a self-reinforcing *triple nexus* alongside cognitive erosion, systematically undermining students' capacity for deep, independent, and higher-order thinking. Drawing on a systematic synthesis of literature from Scopus and Web of Science (2016–2026), the chapter integrates four theoretical constructs: instant gratification theory, Bloom's Revised Taxonomy, cognitive offloading theory, and metacognitive self-regulated learning into a conceptual model that maps the interrelationships between these forces. The analytical discussion presents empirical evidence for five dimensions of cognitive impact: decline in analytical and evaluative thinking, loss of memory, cognitive debt, metacognitive laziness, and erosion of academic integrity. Crucially, the chapter also identifies the conditions under which AI can support rather than supplant deep learning, including dosage effects, metacognitive scaffolding, and pedagogically designed AI integration. Implications and recommendations are derived for educators, curriculum designers, institutions, and policymakers, converging on the vision of *AI-literate deep learning* as the defining educational imperative of the present decade.

Keywords: *instant gratification, AI dependency, deep learning, cognitive offloading, metacognition, higher order thinking, Bloom's Taxonomy, academic integrity, AI literacy*

INTRODUCTION

The rapid proliferation of artificial intelligence (AI) tools in educational settings has fundamentally altered the landscape of student learning. Within

a remarkably short period, generative AI systems, most notably ChatGPT, released by OpenAI in late 2022, have become ubiquitous resources for millions of learners worldwide, providing instant answers, polished essays, and on-demand explanations at zero cognitive cost. The speed of adoption has been unprecedented: within two months of its launch, ChatGPT had surpassed 100 million active users, and by 2024, surveys confirmed that a majority of university students across multiple countries had incorporated AI tools into their study routines. While the transformative potential of AI in personalising and scaling education is well documented (OECD, 2023; UNESCO, 2025), a deeper and less-examined crisis has begun to surface: the systematic erosion of students' capacity for deep, sustained, and independent thinking. This chapter argues that the convergence of instant learning culture and uncritical AI dependency constitutes a substantive threat to the higher-order cognitive development that education is fundamentally designed to cultivate.

Long before the arrival of generative AI, digital platforms had already begun reshaping how young people relate to knowledge and effort. The emergence of short-form video applications TikTok, YouTube Shorts, and Instagram Reels in the late 2010s initiated a profound restructuring of attention patterns, particularly among adolescents and young adults. These platforms, engineered with algorithmic precision to maximise engagement, deliver an endless cascade of content calibrated to trigger dopaminergic reward responses within seconds. Students who spend hours daily in these algorithmically curated environments undergo a form of attentional conditioning: they become habituated to immediate cognitive satisfaction and progressively less tolerant of the sustained, effortful engagement that characterises meaningful academic work.

(Baohong, 2026) demonstrates through structural equation modelling that instant gratification culture operates through a *sequential chain mediation pathway*: first weakening students' attention control capacity, then progressively reducing cognitive stamina, the mental persistence required to engage with complex, ambiguous, and protracted intellectual problems and finally inducing a learning expectation bias in which students come to demand immediate, entertaining, and conclusive feedback from every learning experience. The result is a learner whose psychological relationship with

difficulty, uncertainty, and intellectual effort has been fundamentally altered before they even encounter AI.

(Manizia et al., 2023) provide direct evidence of this phenomenon among digital natives: their pilot study found that children raised in high-digital-exposure environments displayed a measurable reduction in the capacity to delay gratification compared to peers with lower exposure, with socioeconomic variables unable to account for the differences fully. This matters educationally because delayed gratification, the ability to forego immediate reward in favour of a superior future outcome, is not merely a personality trait but a fundamental cognitive competency closely linked to working memory, executive function, and the metacognitive self-regulation that deep learning requires. A generation progressively deprived of the developmental opportunity to practice intellectual patience is, by extension, a generation less equipped for the kinds of sustained inquiry that produce genuine understanding.

Generative AI has not merely entered this cultural context; it has amplified its consequences in ways that educators are only beginning to understand. Students' use of AI tools has evolved rapidly from surface-level information retrieval toward the delegation of cognitive and decision-making processes that lie at the very heart of learning: generating research outlines, constructing arguments, evaluating evidence, and producing written outputs that are expected to represent the student's own intellectual work (Yang et al., 2025). Scholars distinguish between *tool dependence*, the reliance on AI for task-level functional assistance, such as grammar checking or information retrieval, and the more educationally damaging cognitive dependence: the tendency to allow AI to replace independent reasoning in higher-order cognitive activities, such as analysis, judgment, and conceptual synthesis. It is cognitive dependence, not mere AI use, that poses the most serious threat to students' intellectual development.

The empirical evidence for this threat is mounting rapidly. A large-scale mixed-method study involving 666 participants across diverse age groups found a statistically significant negative correlation between frequent AI tool usage and critical thinking abilities, with cognitive offloading confirmed as the key mediating mechanism; younger participants, those most exposed to both

instant media and AI tools, exhibited both the highest AI dependence and the lowest critical thinking scores (Gerlich, 2025).

A laboratory experiment in which students used ChatGPT to assist with essay writing revealed what the researchers termed *"metacognitive laziness"*: a condition in which learners offload not merely routine tasks but entire cognitive processes, orienting, planning, monitoring, and evaluating to the AI, bypassing the synthesis, analysis, and reflection that constitute meaningful learning (Fan et al., 2025). A neural and behavioural study further documented the accumulation of *"cognitive debt"* among sustained AI users, a progressive deficit in cognitive engagement that persisted even when AI assistance was removed (Liang et al., 2025).

The significance of this challenge extends beyond academic performance. The World Economic Forum's *Future of Jobs Report* consistently identifies analytical thinking, creative problem solving, and critical reasoning as the most valued skills in the emerging global economy, precisely those capacities that occupy the highest levels of Bloom's Revised Taxonomy and are most at risk from unreflective AI dependency (Gonsalves, 2024). If educational systems produce graduates who are technically proficient in using AI but cognitively dependent on it for the very tasks that require original human thinking, the long-term consequences for innovation capacity, civic reasoning, and individual intellectual autonomy are profound.

Despite the growing urgency of these concerns, existing scholarship tends to examine AI use in education and the instant-gratification culture as separate phenomena, addressing them through different disciplinary lenses without mapping their intersection or compounding effects. This chapter addresses that gap by developing and applying a triple-nexus model that traces the mutually reinforcing dynamics among instant learning culture, AI dependency, and cognitive erosion. The chapter proceeds as follows: Section 2 establishes the four construct theoretical framework and introduces the triple nexus model. Section 3 presents a systematic analytical discussion of the empirical evidence across four thematic nodes. Section 4 derives practical implications and evidence-based recommendations for four stakeholder groups. Section 5 concludes with a synthesis of the chapter's contributions and directions for future research. Taken together, the chapter offers what the existing literature lacks: a theoretically coherent and empirically grounded

account of how the two most pervasive features of contemporary students' digital lives, instant media culture and AI tools, interact to undermine the cognitive capacities that education exists to build.

THEORETICAL FRAMEWORK

This chapter is grounded in four interconnected theoretical constructs that together form the analytical backbone of the triple nexus model: (1) instant gratification theory as a lens for understanding contemporary learning behaviour; (2) Bloom's Revised Taxonomy as a normative benchmark for higher-order cognition; (3) cognitive offloading theory as an explanatory mechanism for AI dependency; and (4) metacognition and self regulated learning as the essential capacities most eroded by uncritical AI use. Each is discussed below, followed by a conceptual integration that maps their interrelationships in the triple nexus model.

1. Instant Gratification Theory and Instant Learning Culture

Instant gratification refers to the tendency to prioritise immediate reward over delayed but greater benefit a disposition with deep roots in psychological theory. Freudian psychology identifies this tendency with the *pleasure principle* of the unconscious id: the drive to seek pleasure and avoid pain without delay or rational deliberation. The developmental psychology literature frames the ability to delay gratification as a foundational executive function closely associated with long-term life outcomes, including academic achievement, psychological resilience, and cognitive flexibility (Manizia et al., 2023). In educational contexts, the complementary concept of *academic delay of gratification*, the willingness to forego immediate rewards such as entertainment or social interaction in pursuit of long-term academic goals, is recognised as a crucial determinant of student persistence, intellectual depth, and self-regulated learning capacity.

The algorithmic architecture of contemporary digital platforms has dramatically amplified the educational consequences of diminished delay of gratification. Short video platforms are engineered with mechanisms, personalised recommendation algorithms, infinite scroll, and variable reward scheduling that exploit neurological reward circuits to maximise engagement. Their content architecture is characterised by extreme fragmentation (typically 15 to 60 seconds), high intensity emotional stimulation, and

immediate resolution of narrative tension. Repeated exposure to this content architecture induces a measurable shift in students' tolerance for cognitive difficulty and temporal uncertainty. (Baohong, 2026) operationalises instant gratification culture through two indicators short video use intensity and attention fragmentation and demonstrates through chain mediation analysis that this culture degrades classroom engagement not directly, but through a sequential pathway: first eroding attention control capacity, then reducing cognitive stamina, and finally generating a *learning expectation bias* in which students unconsciously transfer their digital consumption expectations to academic contexts, finding the slow, uncertain, and effort intensive process of genuine learning disproportionately aversive by comparison.

2. Bloom's Revised Taxonomy as a Benchmark for Deep Thinking

Bloom's Taxonomy of Educational Objectives, first published in 1956 and revised by Anderson and Krathwohl in 2001, provides the most widely used framework for conceptualising and assessing levels of cognitive engagement in education. The revised taxonomy organises cognitive processes into six hierarchical levels: *Remember*, *Understand*, *Apply*, *Analyse, Evaluate, and Create,* with the final three constituting Higher Order Thinking Skills (HOTS). These upper-level capacities are qualitatively different from lower-level cognitive operations: analysis requires students to break information into constituent parts and examine relationships and patterns; evaluation demands reasoned judgment against explicit or implicit criteria; creation requires the generation of novel products or perspectives through synthesis of existing knowledge. Together, these capacities constitute what this chapter means by *deep thinking* cognitive engagement that produces genuine understanding, transferable knowledge, and intellectual growth (Gonsalves, 2024).

The relevance of Bloom's framework to the present chapter lies in its capacity to make visible what is at stake when students delegate higher-order tasks to AI. A study evaluating ChatGPT-4 and Google Bard against Bloom's Revised Taxonomy found that AI-generated content aligned well with lower-order cognitive tasks across subjects but demonstrated significantly weaker alignment with higher-order levels, particularly in tasks requiring genuine contextual judgment and creative synthesis (Akdeniz et al., 2025). The educational danger is therefore not that AI performs lower-level cognitive tasks on students' behalf; some degree of scaffolding at the lower levels may

legitimately free cognitive resources for higher-order engagement, but that AI enables students to bypass the upper levels entirely, producing outputs that simulate analytical and evaluative thinking without the underlying cognitive processes through which genuine understanding is built. This truncation of the cognitive hierarchy represents the most direct mechanism through which AI dependency erodes deep thinking.

3. Cognitive Offloading Theory and AI Dependency

Cognitive offloading is defined as the use of physical action or external tools to alter the information-processing requirements of a cognitive task, thereby reducing internal cognitive demand (Risko & Gilbert, 2016). Humans have always extended their cognitive capacities through external scaffolding: pen-and-paper note-taking, calendars, calculators, and GPS navigation systems all represent forms of cognitive offloading that are generally considered unproblematic because they free internal cognitive resources for the tasks that matter most. (Grinschgl & Neubauer, 2022) observe that in principle, appropriate cognitive offloading can enhance performance by reducing low-level cognitive load and freeing attentional resources for higher-order processing. The educational concern with AI is therefore not cognitive offloading per se, but its specific character: generative AI can offload the very higher-order cognitive processes, argumentation, evaluation, synthesis, and creative generation that constitute the educational core and that, critically, only develop through effortful practice.

When offloading becomes habitual and extends to higher-order cognitive functions, two compounding harms emerge. The first is cognitive inertia: the systematic avoidance of deep processing in favour of AI-generated, ready-made answers. Cognitive inertia is significantly negatively correlated with critical thinking, divergent thinking, and creative output (Yang et al., 2025). The second is skill atrophy: cognitive abilities that are routinely offloaded to external tools may progressively weaken from disuse.

Scholars draw a cautionary parallel with the calculator debate of the 1970s, arguing that students permitted to offload arithmetic to calculators before developing foundational numerical reasoning may fail to build the cognitive architecture necessary for advanced mathematical thinking. The same logic applies with greater urgency to AI: students who routinely offload argumentation and synthesis may never develop the capacity to perform these

operations independently. The cumulative consequence is what emerging scholarship terms *"cognitive debt"*, the progressive deficit in cognitive capacity resulting from habitual AI reliance, purchased at the cost of short-term efficiency.

4. Metacognition, Self-Regulated Learning, and Metacognitive Laziness

Metacognition, commonly described as *"thinking about thinking"* refers to individuals' awareness and active regulation of their own cognitive processes (Sadykova et al., 2024). It encompasses two interrelated components: metacognitive *knowledge* (what one knows about how memory, attention, and reasoning work) and metacognitive *regulation* (the active monitoring, planning, evaluating, and adjusting of one's own learning strategies). Self-regulated learning (SRL) is the behavioural operationalisation of metacognition in academic contexts: it encompasses goal setting before tasks, monitoring of progress during tasks, strategy adaptation in response to difficulty, and reflective evaluation of outcomes. Both metacognition and SRL are context-specific capacities: they do not transfer automatically from one domain to another but must be explicitly cultivated through structured practice within disciplinary learning environments (van Berk et al., 2024).

The introduction of generative AI into learning environments poses specific, well-documented threats to metacognitive development. Without explicit metacognitive scaffolding, structured prompts, reflection tasks, and monitoring exercises, students in AI-integrated environments tend to default to passive consumption of AI outputs rather than active knowledge construction. (Fan et al., 2025) Identify the consequence as metacognitive laziness: a condition in which the orienting, planning, monitoring, and evaluating processes that constitute self-regulated learning collapse in favour of a single dominant activity interacting with the AI to obtain outputs. Students exhibiting metacognitive laziness may produce outputs that appear competent while having engaged in virtually none of the internal cognitive processes through which genuine learning occurs. This represents what the research terms *decoupled performance*: a dangerous divergence between the surface quality of academic outputs and the depth of the cognitive development those outputs are supposed to evidence and promote.

5. The Triple Nexus Model: Conceptual Integration

The four theoretical constructs described above are not merely parallel frameworks; they interact dynamically to form a self-reinforcing system, which this chapter terms the *triple nexus*. The model comprises three nodes and two feedback loops. The first node is *instant learning culture*: through psychological mechanisms such as diminished delay of gratification, eroded attention control, and distorted learning expectations, students enter AI-integrated learning environments already primed for cognitive offloading. The second node is *AI dependency*: generative AI provides a technological infrastructure perfectly calibrated to satisfy the expectations that instant culture has already cultivated, delivering instant outputs that eliminate the productive struggle through which cognitive capacity is built. The third node is *cognitive erosion*: the cumulative consequence of the first two nodes, manifested as progressive decline in analytical thinking, metacognitive capacity, creative originality, and intellectual self-reliance, precisely the capacities that Bloom's Revised Taxonomy identifies as the aspirational goals of education.

The model's self-reinforcing character derives from two feedback loops. In the first loop, cognitive erosion, specifically the reduction in cognitive stamina and tolerance for difficulty, makes students *more* reliant on AI as a compensatory mechanism, deepening dependency. In the second loop, specifically the AI dependency, the metacognitive laziness it induces reduces students' engagement with the kinds of effortful digital content that might otherwise counteract the effects of instant culture, intensifying the preference for algorithmically curated, cognitively frictionless media. Understanding the triple nexus as a dynamic system, rather than a linear causal chain, is essential for designing effective educational responses: interventions that address only one node, AI use restrictions alone, or media literacy alone will be insufficient to disrupt the cycle. The four theoretical constructs described above are not merely parallel frameworks; they interact dynamically to form a self-reinforcing system, which this chapter terms the triple nexus. This model can be illustrated as follows:

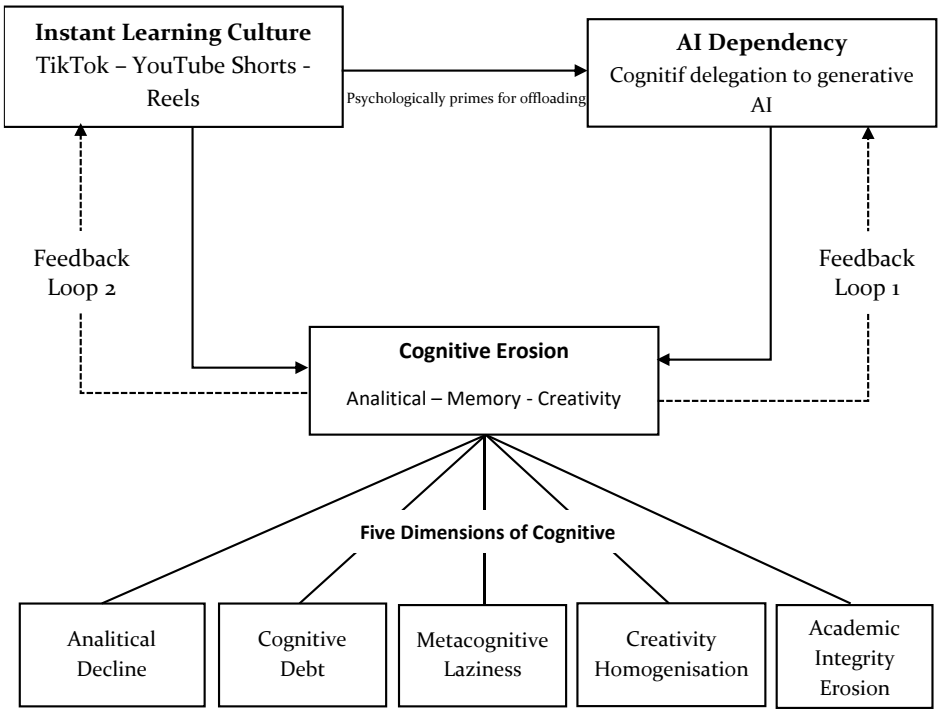


Figure 1. The Triple Nexus Model: The self-reinforcing interrelationship between instant learning culture, AI dependency, and cognitive erosion, including two feedback loops and five dimensions of cognitive impact.

This section analyses empirical evidence through four interconnected nodes that operationalise the triple nexus model: the anatomy of instant learning culture; patterns of AI dependency among students; multidimensional impacts on deep thinking capacity; and the conditions under which AI can genuinely support rather than supplant cognitive depth.

Short video platforms such as TikTok, YouTube Shorts, and Instagram Reels have fundamentally restructured the cognitive ecology in which contemporary students develop. These platforms are engineered not merely to attract attention but to colonise it: their recommendation algorithms, designed to maximise session duration, deliver a continuous stream of content calibrated to trigger dopaminergic reward responses within seconds, creating a compulsive engagement cycle that neurological research associates with the same reward pathways activated by gambling and social validation stimuli.

The result is not simply that students spend more time on their phones; it is that their brains are repeatedly trained, through thousands of daily micro

interactions, to expect immediate cognitive resolution, a resolution that academic learning, by its nature, consistently and necessarily withholds. (Baohong, 2026) operationalises this process through the concept of instant gratification culture and measures its educational impact using structural equation modelling on a large university sample. The findings confirm that short video use intensity and attention fragmentation predict the 'sequential chain mediation pathway' through which instant culture degrades academic engagement: not through a single direct effect, but through the progressive compounding of three distinct cognitive harms that accumulate over time with sustained platform exposure.

The first stage of this pathway is the erosion of attention control capacity, the executive function that enables sustained, focused concentration on a single cognitive task in the face of competing stimuli, internal distraction, and boredom. Platform architecture actively degrades this capacity by rewarding impulsive content switching and penalising sustained engagement with any single piece of content: the algorithm interprets lingering as disinterest and serves fresh stimulation.

Students who spend multiple hours daily in this switching environment undergo a measurable narrowing of their attentional bandwidth, making the deliberate, single-task focus required for reading a complex text, formulating a careful argument, or working through an unfamiliar problem progressively more difficult to sustain voluntarily. The second stage is the depletion of cognitive stamina, the mental persistence required to engage productively with tasks that are complex, ambiguous, protracted, and without immediate reward. Short-video consumption trains the expectation of rapid cognitive closure: every piece of content begins, builds, and resolves within seconds.

Academic tasks, by contrast, are characterised by prolonged uncertainty, deferred resolution, and the requirement to tolerate not-knowing for extended periods before understanding emerges. Students habituated to immediate resolution find this tolerance for cognitive incompleteness increasingly aversive, experiencing it not as a productive challenge but as evidence of personal inadequacy or task irrelevance.

The third stage is the induction of a learning expectation bias, a distorted implicit model of what learning should feel like, built from repeated exposure to algorithmically curated content that is entertaining, immediately

comprehensible, and perfectly matched to the viewer's pre-existing interests. When formal learning fails to satisfy these expectations, when a lecture is demanding rather than entertaining, when a text requires re-reading, when a problem refuses to resolve quickly, students do not necessarily conclude that the task is appropriately challenging; they conclude that it is poorly designed, irrelevant, or beyond their ability. This attribution shift from 'the task is appropriately difficult' to 'the task is wrong, or I am incapable' represents a profound and educationally damaging distortion of students' relationship with intellectual effort (Baohong, 2026).

The cumulative educational consequence of these three stages is what might be termed cognitive unsuitability for depth: a constellation of attentional, motivational, and epistemic dispositions that make students progressively less equipped to engage productively with the kinds of sustained, demanding, and uncertain intellectual work that characterises genuine higher education. Crucially, this unsuitability is not a failure of intelligence or motivation in the conventional sense; it is a culturally induced cognitive adaptation to an information environment that rewards shallowness and punishes depth. Students arrive in educational settings not lacking the capacity for deep thought but lacking the habituation to it, and habituation, unlike raw intelligence, is profoundly susceptible to the environmental conditions that education both inherits from instant culture and can, with deliberate effort, partially reverse.

The scale of this cultural shift is documented comprehensively in global survey data. The most expansive study to date, encompassing 23,218 higher education students from 109 countries and territories, reveals that students primarily use AI tools for brainstorming, summarising texts, and finding research articles, which correspond directly to the lower three levels of Bloom's Revised Taxonomy: Remember, Understand, and Apply (Ravšelj et al., 2025). Fewer than a quarter of respondents reported using AI for tasks involving evaluation, argumentation, or original synthesis, the Analyse, Evaluate, and Create levels that constitute the aspirational outcomes of higher education.

This distribution of AI use across cognitive levels is not coincidental or purely strategic; it is a behavioural signature of an instant learning culture. Students gravitate toward the AI applications that eliminate cognitive friction

most completely, and the tasks of summarising and brainstorming are precisely those where AI delivers the most immediate, effortless, and tangible output relative to the cognitive investment they would otherwise require. The pattern thus both reflects and reinforces the three-stage degradation pathway: students who have already experienced reduced cognitive stamina and distorted learning expectations naturally reach for AI most readily at the points of lowest cognitive demand and, in doing so, further entrench the avoidance of the higher-order engagement that would rebuild those very capacities.

The implications are particularly acute for Generation Z (born 1995–2012) and Generation Alpha (born 2013 onwards) cohorts, for whom algorithmically curated digital environments have constituted the primary information landscape since early childhood. (Manizia et al., 2023) document that digital natives display measurably reduced capacity to delay gratification compared to lower exposure peers, with socioeconomic variables unable to account for the differences fully.

The ability to delay gratification to forego immediate reward in favour of a superior future outcome is not merely a personality trait but a modifiable cognitive competency closely linked to working memory, executive function, and the metacognitive self-regulation that academic success requires. Crucially, research demonstrates that this capacity is most effectively developed during childhood and adolescence, precisely the developmental windows most thoroughly saturated by short-form digital media consumption for the current student cohort. A generation systematically deprived of the developmental opportunity to practice intellectual patience through effortful, slow, and uncertain cognitive work enters higher education and encounters AI tools with a pre-existing vulnerability to exactly the kind of cognitive shortcuts those tools most readily supply.

A systematic review of ChatGPT's integration in higher education (2023–2025), drawing on Scopus and ERIC databases, further confirms that the primary student motivations for AI adoption are academic efficiency, anxiety reduction, and social conformity motivations directly traceable to the expectations and dispositions cultivated by an instant learning culture (Melisa et al., 2025). Notably absent from the leading student motivations for AI use are deeper epistemic goals, such as enhancing understanding, developing

original arguments, challenging existing knowledge, or building skills for future independent application.

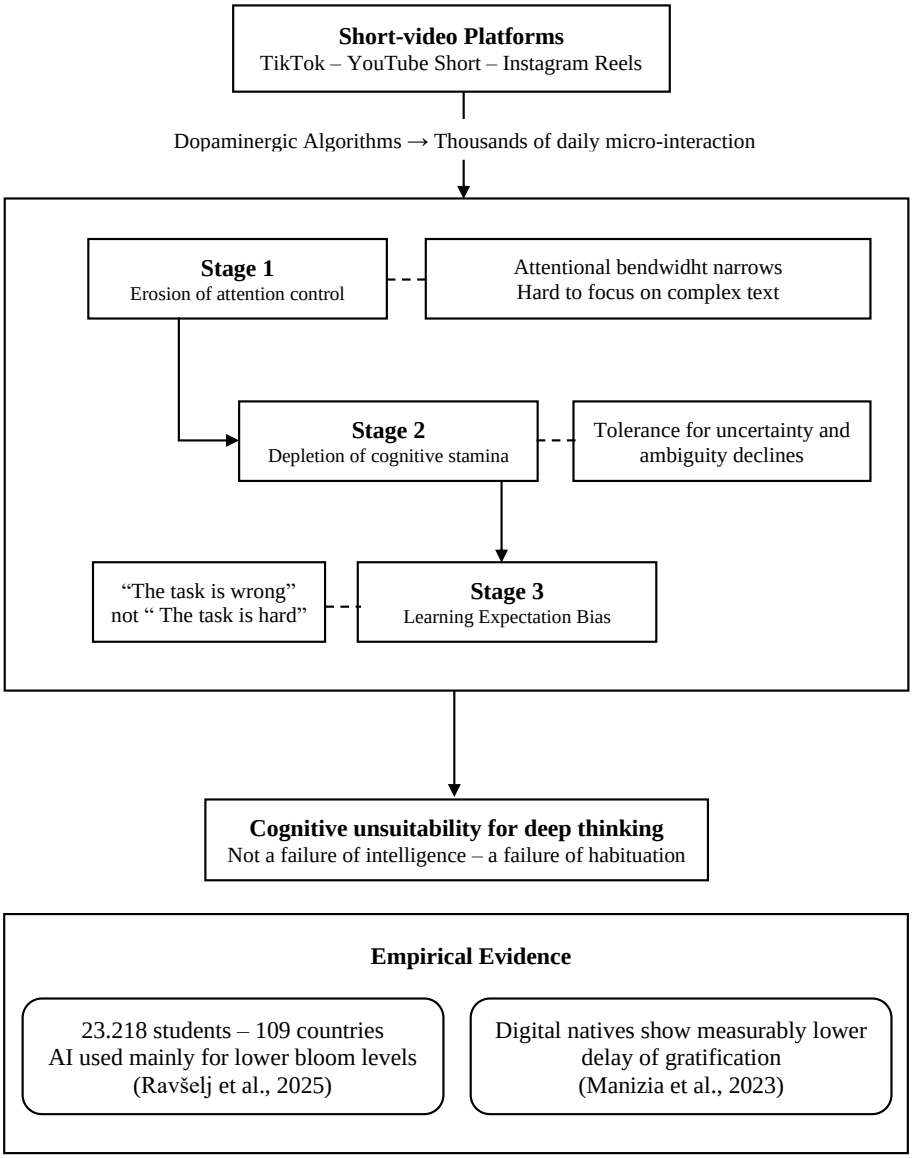


Figure 2. The Anatomy of Instant Learning Culture: A Three-Stage Sequential Chain Mediation Pathway from Short Video Platform Exposure to Cognitive Unsuitability for Deep Thinking.

This motivational profile reveals a fundamental misalignment between how students are using AI and the purposes for which higher education exists. Students are not adopting AI as an intellectual amplifier; they are adopting it as an effort reducer. Moreover, it is precisely this distinction between AI as a scaffold and AI as a substitute that determines whether its integration supports or erodes deep thinking. The structural condition that makes this misalignment so pervasive and so difficult to address is that the instant learning culture has already shaped students' implicit model of what learning is *for* before they ever encounter a generative AI tool. Responding to AI dependency without simultaneously attending to the cultural dispositions that make it so attractive and so entrenched will therefore produce only partial and temporary results. This model can be illustrated as follows:

Understanding the educational impact of AI dependency requires distinguishing between qualitatively different modes of AI engagement. A systematic review drawing on Scopus and ERIC databases (2023–2024) identifies four recurring patterns of educationally problematic AI overreliance: using AI to generate first drafts with minimal independent engagement; seeking direct answers to assessment questions rather than using AI for exploratory dialogue; accepting AI outputs without source verification or critical scrutiny; and using AI specifically to avoid the productive struggle that precedes genuine understanding (Melisa et al., 2025). Across the reviewed studies, students themselves frequently acknowledged these tendencies, reporting that AI use made them "less creative" or reduced their ability to "think as hard as before, yet continued to use AI in these ways, revealing a gap between metacognitive awareness and behavioural change that is itself evidence of cognitive inertia.

The phenomenon of "AIgiarism", a nuanced form of academic dishonesty in which students use AI for paraphrasing and idea generation while experiencing cognitive dissonance about whether this constitutes misconduct, illustrates the normalisation of AI-assisted academic submission. Among 467 participants in one study, 68% recognised that their AI use conflicted with academic integrity norms, while 52% nevertheless prioritised convenience and perceived low detection risk over ethical compliance (Seran et al., 2025). A broader survey of over 1,000 university students found that

more than one-third had used ChatGPT for assessment writing, and of these, 75% acknowledged it constituted cheating but proceeded anyway. The normalisation of this behaviour reflects not moral failure on the part of individual students but the absence of institutional frameworks that clearly articulate why intellectual ownership matters and what is lost when AI performs the cognitive work that learning requires.

The psychological mechanism through which AI use transitions from instrumental to dependent is cognitive inertia: the progressive tendency to avoid deep processing when AI offers a ready-made alternative. (Yang et al., 2025) demonstrate, through an empirical study of Chinese college students, that rising AI dependence is not matched by improvements in academic innovation outcomes, a discrepancy explicable by cognitive inertia. As students consistently rely on AI to generate first drafts, outlines, and arguments, they form an effort-saving learning model in which the short-term efficiency gains of AI use become self-reinforcing. The cognitive habit of reaching for AI before attempting independent thought becomes entrenched, and the activation energy required for unassisted deep thinking progressively increases, making voluntary exit from dependency progressively harder.

Multidimensional Impact on Deep Thinking Capacity

***1.* Decline in Analytical, Synthetic, and Evaluative Thinking**

The most extensively evidenced impact of AI dependency on student cognition is the progressive erosion of the analytical, synthetic, and evaluative thinking that constitutes the upper levels of Bloom's Revised Taxonomy. (Gerlich, 2025)In a large-scale mixed-methods study (n = 666), researchers found a statistically significant negative correlation between frequent use of AI tools and critical thinking abilities across all age groups, with the effect most pronounced among younger participants. The mediating mechanism confirmed by structural analysis was cognitive offloading: students who used AI tools frequently offloaded not only routine tasks but also higher-order cognitive functions that constitute deep thinking.

A controlled experiment comparing ChatGPT-assisted research activities with standard web search found that AI users experienced significantly lower cognitive load during the research process. However, it produced arguments that were lower in quality and shallower in reasoning depth, a pattern the researchers described as "cognitive ease at a cost" (Zhai et

al., 2024). The findings crystallise the central paradox of AI-assisted learning: the subjective experience of productivity and fluency can mask an objective reduction in intellectual depth. Students who feel they are working efficiently may simultaneously produce outputs of diminished analytical value and develop weaker analytical capacities than their unaided counterparts.

2. Memory Retention, Cognitive Debt, and Skill Atrophy

Beyond analytical thinking, AI dependency poses demonstrable risks to memory retention and long-term knowledge acquisition. A laboratory experiment by Akgun and Toker (2024), involving 73 information science undergraduates, found that prolonged AI exposure led to measurable memory decline relative to control conditions, while students who engaged in pretesting before AI use showed significantly improved retention and cognitive engagement (Jose et al., 2025). These findings align with established learning science on retrieval practice: the effortful act of recalling information from memory is among the most powerful mechanisms for consolidating long-term learning. This mechanism, AI-assisted retrieval, entirely bypasses.

The concept of "cognitive debt", the progressive deficit in cognitive capacity resulting from habitual AI reliance, was operationalised in a neural and behavioural laboratory experiment (Kosmayana et al., 2025) that divided participants into three conditions: LLM-assisted writing, search engine-assisted writing, and unaided writing. Cognitive activity decreased measurably in the LLM group; ownership of the written product was rated lowest in the LLM group; and over a four-month follow-up period, LLM users consistently underperformed across neural, linguistic, and behavioural measures of cognitive engagement (Liang et al., 2025). These findings suggest that cognitive debt is not a hypothetical risk but an empirically measurable accumulation of deferred cognitive development purchased at the price of short-term convenience.

3. Metacognitive Laziness and the Collapse of Self-Regulated Learning

The most educationally consequential documented impact of AI dependency is the suppression of metacognitive and self-regulatory processes. (Fan et al., 2025) used process mining of student SRL behaviours to reveal that AI-assisted students spent the majority of task time interacting with ChatGPT, with dramatically reduced time allocated to reading, reflection, and revision,

the activities most strongly associated with deep learning outcomes. The AI group was most prone to copying and pasting AI-generated text, bypassing the synthesis and analytical engagement the writing task was designed to promote. Crucially, this group outperformed controls on immediate output quality while demonstrating significantly weaker metacognitive engagement, a decoupling of performance from learning that represents perhaps the most insidious educational consequence of AI dependency.

The cultural expectations of the instant-learning culture compound this collapse of self-regulated learning. (Xu et al., 2025) demonstrates that without explicit metacognitive scaffolding, students in generative AI learning environments default to passive consumption of AI outputs rather than active knowledge construction. The interaction between instant culture and AI dependency is directly visible here: students who have been culturally conditioned to expect effortless information delivery are precisely those least likely to voluntarily invest the additional metacognitive effort required to learn deeply from, rather than merely with, AI.

4. Creativity, Originality, and the Homogenisation of Thought

The impact of AI dependency on creativity and intellectual originality reveals an important complexity. (Doshi & Hauser, 2024) Conducted an experimental study (n = 293) in which participants wrote short stories under three conditions: unaided, with ChatGPT without guidance, and with ChatGPT with structured instructions. While AI-assisted writing enhanced individual output quality on measures of novelty and usefulness, it simultaneously reduced the collective diversity of ideas across participants, resulting in a homogenisation of thought in which outputs became more similar in structure, language, and conceptual approach. This finding has profound implications for education: AI may produce students who generate outputs that appear more polished and competent individually, while contributing to an intellectual monoculture that impoverishes the diversity of perspectives and approaches that characterise genuine intellectual communities.

Students themselves are becoming aware of these risks. A survey of 173 students across UK educational institutions (Blahopoulou & Ortiz-Bonnin, 2025) found that a significant proportion expressed fear of losing important skills through reliance on AI, including a specific concern about the

"homogenisation of thought" and the erosion of individual intellectual voice. Particularly significant, some students in the survey reported already noticing negative effects on the quality of their own work attributable to AI overreliance, prompting them to voluntarily re-evaluate their usage patterns. This emergent self-awareness among students represents both an acknowledgement of the problem and a foundation for the metacognitive interventions this chapter recommends.

A Balanced View: When AI Supports Deep Learning

A nuanced reading of the evidence reveals that the relationship between AI use and cognitive depth is neither uniformly negative nor determined solely by the frequency of AI use. (Yang et al., 2025) demonstrate a curvilinear dosage effect: moderate AI use can reduce cognitive load on lower-order tasks, such as information retrieval, grammar checking, and formatting, thereby freeing cognitive resources for higher-order processing and creative engagement. The harmful threshold is crossed when AI use becomes habitual, delegative, and metacognitively unmonitored: when it transitions from scaffold to substitute.

The most robustly evidenced enabling condition for productive AI use is explicit metacognitive scaffolding. (Wu et al., 2025)In a quasi-experimental study, researchers found that structured metacognitive prompts designed to activate students' self-monitoring, planning, and evaluative processes significantly improved deep learning outcomes in generative AI environments, with scaffolded students demonstrating superior cognitive proficiency and academic achievement even when using the same AI tools as unscaffolded counterparts. Similarly, (Xu et al., 2025) demonstrates that when explicit metacognitive support is provided in AI learning environments, self-regulated learning behaviours are maintained and the learning experience is significantly enhanced. A second enabling condition is pedagogical design: AI embedded within task structures requiring original argument, authentic problem-solving, and reflective practice functions as an intellectual amplifier rather than a cognitive substitute (Liang et al., 2025).

Four conditions collectively define what the evidence suggests is AI-literate deep learning: using AI with explicit intentionality, habitually subjecting AI outputs to critical evaluation, maintaining metacognitive monitoring of one's own learning process, and operating within pedagogically scaffolded structures that require original intellectual engagement. The

practical challenge for educational systems is creating institutional conditions that make these four practices the default rather than the exception.

Implications and Recommendations

The analytical discussion in Section 3 demonstrates that an instant learning culture and AI dependency are not isolated pedagogical inconveniences but structurally interlocking forces that, together, erode the cognitive foundations of deep learning. Responding effectively to the triple nexus requires coordinated action at four levels: the individual educator, the curriculum, the institution, and the research community. The guiding principle throughout is that the goal is not to eliminate AI from education, an increasingly impractical and pedagogically counterproductive aspiration, but to ensure that its integration amplifies rather than supplants human cognitive development.

For Educators: Coaching Cognition in the Age of AI

The most fundamental implication of the triple nexus for individual educators is a necessary reorientation of professional identity. In an environment where AI can deliver information faster, more comprehensively, and more accessibly than any human instructor, the comparative value of the educator lies not in information transmission but in *cognitive coaching*: modelling, scaffolding, and eliciting the higher-order thinking that AI cannot replicate (OECD, 2023). This requires a pedagogical shift from assessing what students know to cultivating and assessing how students think. Practical recommendations for individual educators include:

a. Embed metacognitive scaffolding explicitly: require pre-task goal setting, maintain a *thinking log* documenting students' independent reasoning before and after AI consultation, and conduct structured post-task reflection on what was genuinely learned versus merely retrieved (Fan et al., 2025; Xu, 2025).
b. Design AI-resistant assessments: oral examinations require students to articulate and defend reasoning in real time, activating precisely the analytical and evaluative capacities AI dependency suppresses. (Kofinas et al., 2025) demonstrate that interactive oral and role-play assessments are inherently difficult for AI to substitute for. Project-based learning, peer assessments, and authentic tasks requiring contextual judgment that AI cannot replicate are equally recommended (Evangelista, 2025).

c. Adopt AI transparent task design: design assignments that require students to critique AI-generated outputs, identify their limitations, and develop original arguments that substantively extend beyond what AI produces. This design principle transforms potential AI misuse into a vehicle for developing the evaluative thinking most at risk from AI dependency.
d. Preserve productive struggle: include course activities that must be completed without AI assistance, maintaining cognitive space for the effortful independent thinking that builds cognitive stamina and metacognitive capacity over time.

For Curriculum Designers: AI Literacy as a Core Competency

Curriculum designers are recommended to embed AI literacy as a cross-curricular competency rather than a standalone elective module. Drawing on the UNESCO AI Competency Framework for Teachers and Students (2024) and the OECD European Commission AI Literacy Framework (2025), this competency should encompass three dimensions: knowledge based (understanding how generative AI systems work, including probabilistic foundations, training data limitations, and hallucination tendencies); skill based (critical evaluation of AI outputs, source verification, dialectical AI use as a thinking partner); and ethical attitudinal (awareness of implications for intellectual autonomy, academic integrity, and long term cognitive development) (Yang et al., 2025).

The 'AI-proof task' principle provides practical guidance for curriculum design: rather than designing tasks that are impervious to AI use (increasingly impossible), design objectives around learning goals that AI cannot fulfil on the student's behalf, genuine contextual judgment, personal experience integration, original argument generation, and sustained inquiry over extended periods (Liang et al., 2025). Programmes should also include a progressive sequence of cognitive challenge: AI-assisted scaffolded tasks at early stages, collaborative AI use at intermediate stages, and independent mastery demonstration at assessment stages.

Given the evidence that instant learning culture progressively erodes cognitive stamina and delayed gratification, curriculum designers should treat the cultivation of these dispositions as explicit curricular goals. Extended reading tasks, long-form writing projects, multi-week independent inquiry,

and sustained problem-solving that cannot be resolved quickly represent what learning science calls 'desirably difficult' tasks, activities experienced as effortful but producing superior long-term learning outcomes that are most at risk of elimination in AI-accelerated curricula prioritising efficiency over depth (Baohong, 2026).

For Institutions and Policymakers: Coherent and Context-Sensitive Frameworks

Institutions should develop AI policies that are pedagogically grounded (articulating why certain boundaries exist in terms of cognitive development, not merely compliance), contextually differentiated (varying appropriately across disciplines and assessment types), consistently communicated at the course level, and collaboratively developed with student participation (Jisc, 2025). Professional development should be structured around the UNESCO AI Competency Framework and specifically equip educators with competencies for AI-resistant and AI-transparent assessment design, the implementation of metacognitive scaffolding, and an understanding of the cognitive science of productive struggle and deep learning (Abbas et al., 2024).

In the Indonesian context specifically, policy recommendations include: integrating AI literacy and critical thinking as explicit competencies within Kurikulum Merdeka, ensuring the autonomy based learning philosophy at its core is not undermined by uncritical AI dependency; developing national teacher professional development standards aligned with the UNESCO AI Competency Framework; establishing longitudinal national data systems to track relationships between AI use patterns and learning outcomes; and creating cross sectoral task forces bringing together educators, AI developers, parents, and students to co design AI use frameworks that are culturally responsive and contextually appropriate for the Indonesian educational landscape.

For Future Research: Four Priority Directions

Four gaps in the current evidence base require urgent attention: (1) *longitudinal research* on the cumulative trajectory of cognitive debt from sustained AI dependency, most existing studies are cross-sectional and cannot trace developmental change over time; (2) *context-specific research in Indonesia and the Global South*, where AI access is expanding rapidly, but empirical evidence remains concentrated in Western and East Asian contexts

(Liang et al., 2025)(3) *large-scale intervention research* in multi-institution randomised designs testing metacognitive scaffolding, AI transparent task design, and oral assessment integration; and (4) *student-centred research* that positions students as active co-investigators of their own AI use patterns, drawing on growing evidence that students' self-awareness of AI's cognitive costs is an underutilised resource for educational improvement (Jisc, 2025).

CONCLUSION

This chapter has developed a theoretically grounded and empirically substantiated account of one of the most consequential pedagogical challenges of the present era: the simultaneous and mutually reinforcing impact of instant learning culture and AI dependency on students' capacity for deep, independent, and higher-order thinking. Through the lens of the triple nexus model, the chapter has demonstrated that these two forces do not operate independently but through a self-reinforcing cycle: instant learning culture shaped by the algorithmic architecture of short-form digital media erodes cognitive stamina, distorts learning expectations, and diminishes the capacity for delayed gratification, creating the psychological conditions in which AI dependency flourishes. AI dependency, in turn, provides a technological infrastructure that perfectly satisfies those expectations while accelerating the cognitive erosion that makes students progressively more reliant on both instant media and AI tools.

The empirical evidence reviewed in Sections 3 and 4 documents five interconnected dimensions of this cognitive erosion: the decline of analytical, synthetic, and evaluative thinking through cognitive offloading; the accumulation of cognitive debt through memory loss and skill atrophy from habitual AI reliance; the collapse of metacognitive self-regulation into metacognitive laziness; the homogenisation of creative and original thought through AI-mediated idea generation; and the structural deterioration of academic integrity as a relationship between students and their own intellectual work. Taken together, these dimensions describe not a temporary adjustment to new technology but a structural shift in how students engage with knowledge, a shift that, if left unaddressed, will progressively undermine the educational outcomes that justify the enterprise of formal learning.

The chapter has also argued consistently for a balanced reading of the evidence. AI is not inherently opposed to deep learning. Under conditions of

pedagogical intentionality, explicit metacognitive scaffolding, AI-transparent task design, progressive cognitive challenge, structured AI literacy education, and institutional policies grounded in cognitive development rationale, AI tools can genuinely scaffold students' engagement with complex ideas in ways that support rather than supplant their own thinking. The decisive factor is not the technology but the instructional architecture within which it is embedded and the metacognitive dispositions students bring to its use. This distinction is the foundation of the chapter's central pedagogical aspiration.

The vision of AI-literate deep learning that this chapter articulates is therefore both a practical aspiration and a theoretical contribution: a mode of educational engagement in which technical AI competence and metacognitive self-awareness are cultivated together, and in which institutions, curriculum designers, and individual educators share responsibility for ensuring that the efficiency gains AI offers amplify rather than erode the intellectual capacities that make human learning irreplaceable. Achieving this vision requires sustained research, particularly longitudinal studies tracing cognitive development trajectories in AI-integrated educational environments, as well as context-specific research in underrepresented settings such as Indonesia. It requires courageous curriculum reform that explicitly cultivates cognitive stamina and critical AI literacy alongside disciplinary knowledge. Moreover, it requires a clear-eyed institutional commitment to what education, at its deepest level, is for: not the production of outputs, but the cultivation of minds.

In a world where AI can generate the *outputs* of thinking almost instantaneously, the most urgent educational task is ensuring that students can still perform the *process* of thinking for themselves. That process, slow, uncertain, effortful, and irreplaceable, is the foundation of critical citizenship, professional competence, and intellectual autonomy. It is what distinguishes education from mere information delivery. Moreover, in the age of artificial intelligence, it has never been more important, or more in need of deliberate cultivation and steadfast institutional protection.

REFERENCES

Abbas, M., Jam, F. A., & Khan, T. I. (2024). Is it harmful or helpful? Examining the causes and consequences of generative AI usage among university students. *International Journal of Educational Technology in Higher Education*, *21*(1). https://doi.org/10.1186/s41239-024-00444-7

Akdeniz, H., Clark, T., & Roberts, J. L. (2025). Can AI Generate Questions Aligned with Bloom's Taxonomy? A Framework for Gifted Education to Support Teachers. In *Journal of Advanced Academics* (Vol. 36, Issue 4 Special Issue on Artificial Intelligence in Advanced Academics, pp. 671–694). https://doi.org/10.1177/1932202X251349917

Anderson, L. W., & Krathwohl, D. R. (Eds.). (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. Longman.

Baohong, L. (2026). *International Journal of Social Science and Human Research Mechanism Analysis of the Decline in Students ' Classroom Engagement Under the Context of Instant Gratification Culture (Short-Video Proliferation): A Reflection on Traditional Classroom Teaching*. *09*(01), 270–282. https://doi.org/10.47191/ijsshr/v9-i1-32

Blahopoulou, J., & Ortiz-Bonnin, S. (2025). Student perceptions of ChatGPT: benefits, costs, and attitudinal differences between users and non-users toward AI integration in higher education. *Education and Information Technologies*, *30*(14), 19741–19764. https://doi.org/10.1007/s10639-025-13575-9

Doshi, A. R., & Hauser, O. P. (2024). Generative AI enhances individual creativity but reduces the collective diversity of novel content. *Science Advances*, *10*(28), 1–10. https://doi.org/10.1126/sciadv.adn5290

Evangelista, E. D. L. (2025). Ensuring academic integrity in the age of ChatGPT: Rethinking exam design, assessment strategies, and ethical AI policies in higher education. *Contemporary Educational Technology*, *17*(1), 1–19. https://doi.org/10.30935/cedtech/15775

Fan, Y., Tang, L., Le, H., Shen, K., Tan, S., Zhao, Y., Shen, Y., Li, X., & Gašević, D. (2025). Beware of metacognitive laziness: Effects of generative artificial intelligence on learning motivation, processes, and performance. *British Journal of Educational Technology*, *56*(2), 489–530. https://doi.org/10.1111/bjet.13544

Gerlich, M. (2025). AI Tools in Society: Impacts on Cognitive Offloading and the Future of Critical Thinking. *Societies*, *15*(1), 1–28. https://doi.org/10.3390/soc15010006

Gonsalves, C. (2024). Generative AI's Impact on Critical Thinking: Revisiting Bloom's Taxonomy. *Journal of Marketing Education*, 4–19. https://doi.org/10.1177/02734753241305980

Grinschgl, S., & Neubauer, A. C. (2022). Supporting Cognition With Modern Technology: Distributed Cognition Today and in an AI-Enhanced Future. *Frontiers in Artificial Intelligence*, *5*(July), 1–6. https://doi.org/10.3389/frai.2022.908261

Jose, B., Cherian, J., Verghis, A. M., Varghese, S. M., & Joseph, S. (2025). The

cognitive paradox of AI in education: between enhancement and erosion. *Frontiers in Psychology*, *16*. https://doi.org/10.3389/ fpsyg.2025.1550621

Kofinas, A. K., Tsay, C. H. H., & Pike, D. (2025). The impact of generative AI on the academic integrity of authentic assessments in higher education. *British Journal of Educational Technology*, *56*(6), 2522–2549. https://doi.org/10.1111/bjet.13585

Liang, J., Stephens, J. M., & Brown, G. T. L. (2025). A systematic review of the early impact of artificial intelligence on higher education curriculum, instruction, and assessment. *Frontiers in Education*, *10*. https://doi.org/10.3389/feduc.2025.1522841

Manizia, L., Borah, T., Bhattacharjee, R., Saikia, J., Dutta, A., & Boruah, J. (2023). Instant Gratification and The Digital Natives: A Pilot Study. *Educational Administration: Theory and Practice*, *29*(3), 928–933. https://doi.org/10.53555/kuey.v29i3.7471

Melisa, R., Ashadi, A., Triastuti, A., Hidayati, S., Salido, A., Luansi Ero, P. E., Marlini, C., Zefrin, Z., & Al Fuad, Z. (2025). Critical Thinking in the Age of AI: A Systematic Review of AI's Effects on Higher Education. *Educational Process: International Journal*, *14*, 1–22. https://doi.org/10.22521/edupij.2025.14.31

OECD. (2023). *OECD Digital Education Outlook 2023: Towards an effective digital education ecosystem*. OECD Publishing. https://doi.org /10.1787/c74f03de-en

Ravšelj, D., Keržič, D., Tomaževič, N., Umek, L., Brezovar, N., Iahad, N. A., Abdulla, A. A., Akopyan, A., Segura, M. W. A., AlHumaid, J., Allam, M. F., Alló, M., Andoh, R. P. K., Andronic, O., Arthur, Y. D., Aydin, F., Badran, A., Balbontín-Alvarado, R., Saad, H. Ben, … Aristovnik, A. (2025). Higher education students' perceptions of ChatGPT: A global study of early reactions. In *PLoS ONE* (Vol. 20, Issue 2). https://doi.org/10.1371/journal.pone.0315011

Risko, E. F., & Gilbert, S. J. (2016). Cognitive Offloading. *Trends in Cognitive Sciences*, *20*(9), 676–688. https://doi.org/10.1016/j.tics.2016.07.002

Sadykova, A., Iskakova, M., Ismailova, G., Ishmukhametova, A., Sovetova, A., & Mukasheva, K. (2024). The impact of a metacognition-based course on school students' metacognitive skills and biology comprehension. *Frontiers in Education*, *9*(October). https://doi.org/10.3389 /feduc.2024.1460496

UNESCO. (2025). *AI and the future of education: Disruptions, dilemmas and directions*. UNESCO. https://ciddl.org/summary-of-unesco-ai-and-the-future-of-education/

van Berk, B., Kroehne, U., & Dignath, C. (2024). On the right track: decoding self-regulated learning in young students' log data with the digital train

track task. *Frontiers in Education*, *9*(October). https://doi.org/10.3389/feduc.2024.1388202

Wu, J., Wang, J., Lei, S., Wu, F., & Gao, X. (2025). The impact of metacognitive scaffolding on deep learning in a GenAI-supported learning environment. In *Interactive Learning Environments* (Vol. 33, Issue 9, pp. 5166–5183). https://doi.org/10.1080/10494820.2025.2479162

Xu, X., Qiao, L., Cheng, N., Liu, H., & Zhao, W. (2025). Enhancing self-regulated learning and learning experience in generative AI environments: The critical role of metacognitive support. In *British Journal of Educational Technology* (Vol. 56, Issue 5, pp. 1842–1863). https://doi.org/10.1111/bjet.13599

Yang, Z., Deng, H., & Jiang, N. (2025). The impact mechanism of artificial intelligence dependence on college students' innovation capability: an empirical study from China. *Frontiers in Psychology*, *16*(December), 1–17. https://doi.org/10.3389/fpsyg.2025.1732837

Zhai, C., Wibowo, S., & Li, L. D. (2024). The effects of over-reliance on AI dialogue systems on students' cognitive abilities: a systematic review. *Smart Learning Environments*, *11*(1). https://doi.org/10.1186/s40561-024-00316-7

AUTHOR PROFILE



Dr. Berlinda Setyo Yunarti, S. Sos., M.Pd. (Linda) was born in Merauke on June 20, 1970. She completed her Bachelor's degree (S1) at Cenderawasih University, Faculty of Social and Political Sciences (FISIP), majoring in State Administration, graduating in 1993. She earned her Master's degree (S2) at Cenderawasih University, Faculty of Teacher Training and Education (FKIP), majoring in Educational Management, graduating in 2014, and completed her Doctoral degree at Yogyakarta State University, Faculty of Education and Psychology (FIPP), majoring in Educational Management, graduating in 2024.

She is currently a Permanent Lecturer at Saint Yakobus Catholic College (STK) in Merauke, South Papua, with an additional role as Head of the Teacher Professional Education Study Program in Catholic Religious Education.

Co AUTHOR PROFILE



Yohanes Hendro Pranyoto, S.Pd., M.Pd. (Hendro) was born in Gunungkidul, Special Region of Yogyakarta, on June 17, 1990. He completed his Bachelor's degree (S1) in Catholic Religious Education at Sanata Dharma University in 2012. He went on to earn his Master's degree (S2) in Educational Management at Cenderawasih University in 2016, deepening his understanding of institutional leadership and educational administration. He is currently pursuing his Doctoral degree (S3) in Educational Technology at Universitas Pendidikan Ganesha (Undiksha). Alongside his studies, he serves as a Permanent Lecturer in the Catholic Religious Education Study Program at Saint James Catholic College of Merauke (STK) in Merauke, South Papua, and as Head of the Information Technology and Database Center (PTIPD. His academic and professional interests include Catholic religious education, educational technology, and educational management.

CHAPTER XI

Artificial Intelligence Ethics in a Rule of Law Framework: A Systematic Literature Review of Equitable AI Governance

Rahma Aulia[1], and Retno Mawarini Sukmariningsih[2]
[1]Fakultas Syariah, UIN Sunan Kudus - Indonesian
[2]Fakultas Hukum, Universitas 17 Agustus 1945 Semarang

Correspondent author: rahmaaulia@uinsuku.ac.id

ABSTRACT

This article aims to analyze the role of artificial intelligence (AI) from the perspective of the rule of law and to formulate a just AI governance framework, given that algorithmic systems now determine access to public services, economic opportunities, and the protection of citizens' rights. Methods: The study used a systematic literature review based on the PRISMA 2020 protocol, with a Boolean search strategy in Scopus, Web of Science, HeinOnline, ACM Digital Library, and Google Scholar databases published between 1990 and 2025, filtering 35 academic sources that were analyzed thematically to map algorithmic power, global AI ethical principles, and regulatory accountability and transparency mechanisms. The literature shows normative convergence on five main principles—transparency, fairness, non-maleficence, accountability, and privacy—but their implementation remains fragmented and weak at the law enforcement level; Thematic analysis shows a shift in focus from critiques of algorithmic opacity (2014-2017) to the proliferation of ethical guidelines (2018-2020), explanatory legal instrumentation (2017-2021), and the consolidation of binding regulations and systematic reviews (2021-2025). The conclusion is that AI is not a neutral technology, so its governance must be placed within a legal framework that guarantees legality, accountability, and protection of rights. At the same time, Indonesia does not yet have a comprehensive AI law. The implications of this article propose an integrative framework, "Rule of Law-by-Design", that maps AI ethical principles to the four pillars of the Indonesian legal state, recommending the strengthening of algorithm audits, the right to explanation, and independent oversight bodies as concrete legal instruments for the formation of AI regulations that are contextual to Pancasila and the 1945 Constitution of the Republic of Indonesia.

Keywords: Artificial Intelligence; Rule of Law; Technology Ethics; Responsible AI Governance; Systematic Literature Review

INTRODUCTION

Artificial Intelligence (AI) has shifted from being a mere computing tool to a system capable of analysis, classification, prediction, and decision-making with real social and legal consequences. Algorithmic systems are now used to screen job applicants, assess creditworthiness, support predictive law enforcement, and even assist bureaucracies in public policymaking. This shift demonstrates that AI is no longer a purely technical issue but has become a legal one, as it can affect an individual's rights, interests, and social position.

The assumption that technology is neutral and value-free has been widely challenged in the literature on algorithm ethics. Mittelstadt et al. (2016) argue that algorithms have a moral dimension because their decisions have a real impact on human lives, while Burrell (2016) notes that many machine learning systems operate in an opaque manner, making them difficult to understand even for their designers. This condition gives rise to the term "black box algorithm," which refers to algorithmic decision-making processes that are difficult to explain, even though the final results can be observed. When opaque systems make decisions affecting citizens' rights, the fundamental principles of the rule of law—legal certainty, accountability, and the right to explanation—are threatened, as specifically criticized by Citron and Pasquale (2014) in their concept of due process for automated predictions, and Kroll et al. (2017) in their idea of accountable algorithms.

This problem is exacerbated by algorithmic bias. O'Neil (2016) warned that unsupervised algorithms have the potential to become "weapons of mathematical destruction" that reproduce and magnify social inequalities already present in their training data, a finding empirically corroborated by Buolamwini and Gebru (2018) through an audit of a commercial facial classification system that revealed sharp accuracy disparities based on skin color and gender. Barocas and Selbst (2016) added that the discriminatory impacts of big data often arise not from the designers' malicious intent, but from the way historical data represents pre-existing structural inequalities. This risk is even more serious when AI is used in areas directly related to citizens' rights, such as public services, the financial sector, and law enforcement.

From a rule-of-law perspective, the development of AI presents a conceptual challenge because power no longer originates solely from political institutions, but also from actors who control data, technological infrastructure, and algorithm design. In fact, the classical notion of a rule-of-law state—both in the rechtsstaat tradition and the rule of law of A.V. Dicey—is built on the recognition that every form of power must be subject to the principles of legality, separation of powers, protection of human rights, and independent judicial oversight. Zarsky (2016) calls this condition the need for a new analytical map to assess the efficiency and fairness of opaque automated decision-making, as algorithmic power operating covertly can escape constitutional oversight if the law does not adapt.

To address these issues, global discourse has developed the concept of Responsible AI Governance. Jobin, Ienca, and Vayena (2019) analyzed 84 AI ethics guidelines and found global convergence on five principles — transparency, fairness, non-maleficence, responsibility, and privacy—despite significant divergence in their interpretation and implementation. Similar findings were corroborated by Floridi et al. (2018) through their AI4People initiative, which formulated five basic ethical principles for a "good AI society" along with twenty concrete recommendations for policymakers, and further summarized by Floridi and Cowls (2019) into a single five-principle framework for AI in society.

However, the biggest challenge lies not in simply formulating ethical principles but in translating them into binding legal mechanisms. Hagendorff (2020) criticized that most AI ethics guidelines are voluntary, lack sanctions, and risk becoming instruments of "ethics washing." Mittelstadt (2019) added that principles alone are insufficient to guarantee ethical AI, given the lack of institutional incentives and professional standards comparable to those in the legal or medical professions.

Recent systematic literature reviews show similar patterns: Ismail and Ahmad (2025) found high variation across jurisdictions in institutionalizing AI ethics principles, Batool, Zowghi, and Bano (2025) noted that the majority of AI governance solutions remain at the organizational level rather than the binding national level, while Giarmoleo et al. (2024) and Wieczorek et al. (2025) showed that the dominant ethical approaches remain principled and not yet legally operational.

At the concrete regulatory level, the European Union took significant steps through Regulation (EU) 2024/1689 (Artificial Intelligence Act), which adopted a risk-based approach with a centralized oversight mechanism—a development Veale and Zuiderveen Borgesius (2021) analyzed as a step forward but left enforcement mechanisms unclear. In contrast, in Indonesia, a legal framework specifically governing AI governance is not yet in place; existing regulations address personal data protection only through Law Number 27 of 2022 concerning Personal Data Protection, while aspects of algorithmic accountability, system audits, and the right to an explanation remain unregulated.

The gap between the rapid development of global literature on AI ethics and governance and the limited synthesis linking it to Indonesia's legal framework is a research gap this article aims to fill. Unlike previous normative-conceptual studies that relied on non-systematic literature searches, this article utilizes a systematic literature review using the PRISMA 2020 protocol (Page et al., 2021) combined with a Boolean search methodology as recommended by Kitchenham et al. (2009) within the evidence-based research tradition, to ensure transparency and replicability of the literature search, screening, and synthesis process.

Based on this description, this article is formulated to answer three research questions: first, how does the academic literature map the challenges of algorithmic power to the principles of the rule of law? Second, how have thematic trends and convergence-divergence of AI ethical principles developed in global literature and regulations over the past decade? Third, how can these principles be integrated into an AI governance framework tailored to the Indonesian legal system? The novelty of this article lies in the formulation of an integrative framework, "Rule of Law-by-Design", that maps five global AI ethical principles into the four pillars of the rule of law, as a conceptual contribution to the formation of equitable AI governance law in Indonesia.

METHOD

This article uses a systematic literature review (SLR) approach that follows the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 reporting protocol as formulated by Page et al. (2021), combined with the principles of evidence-based research in the tradition of

Kitchenham et al. (2009), which emphasizes that the reliability of an SLR depends on the clarity of the search protocol, selection criteria, and data extraction process that other researchers can replicate. The SLR approach was chosen because it provides a transparent, systematic process for searching, screening, and synthesizing the literature, so that the resulting conclusions do not depend on subjective source selection, as is common in non-systematic conceptual studies.

The literature search was conducted in five academic databases: Scopus, Web of Science, HeinOnline (for legal literature), ACM Digital Library (for computer science literature related to fairness, accountability, and transparency), and Google Scholar as a complementary database to capture authoritative gray literature documents, such as official regulations and recommendations from international organizations. The publication timeframe was limited to 1990–2025 for legal databases to capture the doctrinal foundations of due process, and 2014–2025 for computer science databases to capture the development of the discourse on algorithmic accountability following the widespread use of machine learning in public services.

Search keywords were constructed using a combination of the Boolean operators AND, OR, and NOT with three groups of terms (concept blocks) linked hierarchically. The first group represented technological concepts: ("artificial intelligence" OR "algorithmic decision-making" OR "machine learning" OR "automated decision-making"). The second group represented normative concepts: ("ethics" OR "governance" OR "accountability" OR "fairness" OR "bias" OR "transparency"). The third group represented legal concepts: ("rule of law" OR "due process" OR "human rights" OR "regulation" OR "right to explanation"). The final string used in the Scopus and Web of Science databases is in the form: ("artificial intelligence" OR "algorithmic decision-making" OR "machine learning") AND ("ethics" OR "governance" OR "accountability" OR "fairness" OR "bias") AND ("rule of law" OR "due process" OR "human rights" OR "regulation"), applied to the title, abstract, and keywords fields (title-abstract-keywords), with a filter for English-language documents and document types of journal articles, reputable proceedings, or scholarly book chapters.

In HeinOnline, the string conforms to legal database conventions to: ("artificial intelligence" OR "algorithm!") /p ("due process" OR "accountab!"

OR "explanation" OR "discrimination"), using the proximity operator (/p) to capture documents that discuss both concepts in the same paragraph, resulting in more precise search results than with a traditional AND operator. In the ACM Digital Library, the search focused on the Fairness, Accountability, and Transparency (FAccT/FAT*) conference series using the combination ("fairness" OR "explainability" OR "algorithmic bias") AND ("accountability" OR "sociotechnical" OR "due process"). This multilevel search strategy is intended to balance sensitivity (capturing as many relevant documents as possible) and precision (minimizing noise from purely technical documents that do not address ethical, legal, or governance dimensions).

PRISMA Inclusion, Exclusion, and Selection Criteria

The inclusion criteria used included: (1) articles from reputable indexed journals (Scopus/Web of Science/DOAJ), reputable conference proceedings (ACM, PMLR), or official publications of authoritative bodies (regulations, international organization recommendations) that discuss AI ethics, accountability, transparency, or regulation; (2) full texts that can be verified and downloaded online through official databases, institutional repositories, or SSRN; and (3) relevant to one of the three core research themes, namely algorithmic power-rule of law, global AI ethical principles, or accountability-transparency of AI regulation. Exclusion criteria included articles without a peer-reviewed process (non-peer reviewed), popular media opinion pieces without academic references, online encyclopedia entries, and duplicate publications in more than one database.

Following the PRISMA 2020 flowchart, which comprises identification, screening, eligibility, and inclusion stages, the literature selection process can be summarized as follows. The identification stage captured 214 entries from five databases (Scopus n=68, Web of Science n=54, HeinOnline n=41, ACM Digital Library n=33, Google Scholar n=18), which were reduced to 176 entries after automatic and manual duplication removal. Title and abstract screening of these 176 entries resulted in 89 entries thematically relevant to the three research focuses, with the main exclusion reasons being topic discrepancy (n=54) and language other than English/Indonesian (n=33). The full-text eligibility assessment stage then eliminated 43 entries because they did not meet methodological criteria, were not legally accessible in full text, or were conceptual duplicates of already included sources, leaving 46 entries. Of these,

35 sources were ultimately designated as the primary unit of analysis in the final synthesis. At the same time, the remaining 11 entries were excluded due to thematic redundancy with sources that had already represented similar arguments more authoritatively (e.g., secondary articles that revisited the AI4People principle without any new analytical contribution).

This stringent screening ratio (from 214 initial entries to 35 final sources, or approximately 16%) is intended to ensure that the resulting synthesis is truly grounded in the most relevant and high-quality literature, rather than simply an accumulation of citations. As part of the quality appraisal, each source that passes the eligibility stage is assessed on three indicators: methodological clarity (for empirical studies/previous SLRs), clarity of normative argumentation (for conceptual and legal studies), and citation rate and publisher reputation as proxies for academic credibility.

Source Classification, Thematic Analysis, and Trend Analysis

The thirty-five sources included in the synthesis can be classified into five categories: (1) eleven articles on the philosophy and ethics of algorithms (among them Mittelstadt et al., 2016; Burrell, 2016; Jobin et al., 2019; Floridi et al., 2018; Floridi & Cowls, 2019; Hagendorff, 2020; Mittelstadt, 2019); (2) seven articles/computer science proceedings on algorithmic fairness and empirical evidence of bias (Binns, 2018; Buolamwini & Gebru, 2018; Barocas & Selbst, 2016; Selbst et al., 2019); (3) ten articles and legal working papers on accountability, explanation, and due process (Citron & Pasquale, 2014; Kaminski, 2019; Kroll et al., 2017; Wachter et al., 2017; Wachter & Mittelstadt, 2019; Selbst & Barocas, 2018; Selbst & Powles, 2017; Zarsky, 2016; Diakopoulos, 2015; Rudin, 2019); (4) five previous systematic literature reviews that serve as methodological and substantive comparisons (Batool et al., 2025; Ismail & Ahmad, 2025; Giarmoleo et al., 2024; Wieczorek et al., 2025; Kitchenham et al., 2009); and (5) two international regulatory instruments along with two Indonesian national legal instruments as primary legal materials for comparison (Regulation (EU) 2024/1689, Veale & Zuiderveen Borgesius, 2021, UNESCO 2021, OECD 2019, Law No. 27/2022, and the 1945 Constitution of the Republic of Indonesia).

Data analysis was conducted through thematic synthesis with three stages: (1) open coding of the main findings and arguments of each source; (2) grouping codes into descriptive themes (e.g., algorithm opacity, empirical

bias, ethical principle convergence, accountability gap, risk-based regulatory approach); and (3) formulation of analytical themes that connect the literature findings to the theoretical framework of the rule of law. In addition to the cross-source thematic synthesis, this article also conducts a thematic trend analysis by mapping the year of publication of each source against its dominant theme, to identify how the focus of academic discourse shifts over time—a step rarely undertaken in conventional normative-conceptual studies but common in SLR in computer science and technology policy.

It should be emphasized that the SLR approach in this article remains combined with a conceptual approach as commonly used in normative legal research, so that the results of the literature synthesis are not only described but also analyzed argumentatively to answer research questions regarding their relevance to the Indonesian legal framework. This combination is consistent with the prescriptive nature of legal research: a systematic literature review provides a credible and traceable empirical-academic basis, while conceptual analysis translates this basis into legal propositions applicable to the Indonesian context.

RESULTS AND DISCUSSION

Characteristics and Thematic Trends of the Analyzed Literature

Mapping publication years against dominant themes across the 35 sources analyzed reveals four sequential, yet overlapping, thematic waves. The first wave (2014–2017) was dominated by foundational critiques of algorithmic opacity and accountability from a legal and computer science perspective, characterized by the work of Citron and Pasquale (2014) on due process for automated predictions, Pasquale (2015) on black box societies, Diakopoulos (2015) on algorithmic accountability reporting, Barocas and Selbst (2016) on disparate impact, Burrell (2016) on opacity typologies, Mittelstadt et al. (2016) on the moral dimensions of algorithms, Zarsky (2016) on analytical maps of algorithmic decisions, and Kroll et al. (2017) on accountable algorithms. This wave is characterized by the dominance of conceptual critiques of algorithmic risk, with few concrete legal solutions offered.

The second wave (2017-2019) shifted towards the proliferation of AI ethics guidelines and explanatory legal instrumentation, marked by Selbst and Powles (2017) on meaningful information and the right to explanation, Wachter, Mittelstadt, and Russell (2017) on post-GDPR counterfactual

explanations, Buolamwini and Gebru (2018) on empirical bias audits, Binns (2018) and Selbst et al. (2019) on fairness from a political philosophy and sociotechnical perspective, Selbst and Barocas (2018) on the intuitive appeal of explainable machines, Floridi et al. (2018) on AI4People, and Jobin et al. (2019) on a global mapping of 84 AI ethics guidelines. This wave marked a shift from simply criticizing opacity to formulating specific legal principles and mechanisms in response.

The third wave (2019-2021) focused on critically evaluating the effectiveness of ethical guidelines and strengthening the legal basis for data protection, as reflected in Kaminski (2019) on the two dimensions of the right to explanation, Rudin (2019) on the argument for inherently interpretable models, Mittelstadt (2019) on the limitations of principles alone, Hagendorff (2020) on a critical evaluation of hundreds of AI ethics guidelines, Wachter and Mittelstadt (2019) on the right to reasonable inference, and Floridi and Cowls (2019) on a single five-principle framework. This wave is characterized by a more skeptical tone regarding the effectiveness of soft law and a push towards more legally precise instrumentation.

The fourth wave (2021-2025) marked a consolidation towards binding regulation and systematic evaluation across the literature, reflected in Veale and Zuiderveen Borgesius' (2021) critical analysis of the draft AI Act, UNESCO's (2021) global AI ethics recommendations, the enactment of Regulation (EU) 2024/1689 as the world's first comprehensive AI regulation, and four recent systematic literature reviews (Giarmoleo et al., 2024; Ismail & Ahmad, 2025; Batool et al., 2025; Wieczorek et al., 2025) that aggregately evaluate the effectiveness of all previous waves. The shift from single studies to SLRs in this wave indicates that the field of AI governance has matured sufficiently for meta-analytic evaluation, while also confirming the methodological relevance of the SLR approach employed by this article.

This thematic trend has important implications for the study of the rule of law: the four-wave shift demonstrates an evolutionary pattern from critique (wave 1) to normativeization (wave 2), then to critical evaluation (wave 3), and finally to regulatory consolidation (wave 4)—a pattern commonly found in the evolution of new technology legal regimes, but Indonesian legal literature has not followed a similar pattern as it is still in the pre-first wave stage for the specific issue of algorithmic accountability, despite having a foundation in the Personal Data Protection Law. Disciplinarily, the distribution of sources shows

a balance between law (14 sources), the philosophy/ethics of technology and computer science (16 sources), and SLR methodology and regulation (5 sources), confirming AI governance's character as an inherently interdisciplinary field of study.

Algorithmic Power as a New Challenge to the Principle of the Rule of Law

A synthesis of the first wave of literature confirms that AI presents a new form of power not fully encompassed within the classical framework of the rule of law. The concept of the rule of law, both in the rechtsstaat tradition and Dicey's rule of law, was constructed to limit the power of political institutions through the principles of legality, the division of powers, and the protection of rights.

However, as Mittelstadt et al. (2016) demonstrate, contemporary algorithms operate as social mediators that contribute to the distribution of opportunities and risks in society, without always being subject to the same accountability mechanisms as conventional public institutions. This issue is deepened by Burrell's (2016) concept of algorithmic opacity, which identifies three types of opacity—the deliberate secrecy of corporations, the limited technical literacy of the public, and the intrinsic opacity of deep learning architectures—that together give rise to the condition of a "black box society" (Pasquale, 2015).

From a legal perspective, these conditions directly contradict the principles of legal certainty and the right to due process. O'Neil (2016) provides empirical illustrations of how algorithmic models in criminal justice, recruitment, and teacher performance assessment reproduce structural biases in historical data, while Buolamwini and Gebru's (2018) quantitative evidence shows accuracy disparities of up to 34.7% in the classification of dark-skinned female faces compared to 0.8% in light-skinned males—empirically confirming Barocas and Selbst's (2016) normative concerns about the disparate impact of big data. Kroll et al. (2017) assert that accountability mechanisms and legal standards developed to oversee human decision-makers often fail to apply to computer systems, necessitating new technical accountability instruments such as cryptographic audits and procedural regularity.

The issue of algorithmic power is most evident in public services, the financial sector, and law enforcement. In public services, algorithmic systems that select social assistance recipients can replicate biases in historical population data. In the financial sector, credit scoring algorithms can create indirect discrimination if proxy variables correlate with legally protected demographic characteristics, as Citron and Pasquale (2014) warned through the concept of a "scored society." In law enforcement, predictive algorithms risk producing disproportionate surveillance patterns, as O'Neil (2016) warned. These three domains demonstrate that the impact of algorithmic power has been and continues to be evident in contemporary governance practices, including potentially in Indonesia, as algorithmic systems become more widely adopted in digital government services and the technology-based financial services industry.

This synthesis of the literature suggests three implications for the rule of law. First, the principle of separation of powers needs to be expanded to include oversight of private actors controlling algorithmic infrastructure. Second, the principle of legality demands that any algorithmic decision affecting citizens' rights have a clear and verifiable legal basis, rather than being legitimized by claims of statistical accuracy. Third, judicial independence needs to be supported by new technical capacities, such as independent algorithmic audits, given that testing the validity of algorithmic decisions cannot be done solely through conventional legal interpretation.

Convergence and Divergence of AI Ethics Principles in Global Literature

One of the most significant findings from this literature synthesis is the strong normative convergence at the principle level, but sharp divergence at the implementation level. Jobin et al. (2019), through an analysis of 84 AI ethics guidelines documents, identified five principles that appeared most frequently: transparency, fairness, non-maleficence, responsibility, and privacy. This finding aligns with the AI4People framework (Floridi et al., 2018) and the single five principles framework (Floridi & Cowls, 2019). It is corroborated from a political philosophy perspective by Binns (2018), who shows that the concept of fairness in machine learning actually inherits long-standing debates about equality and distributive justice. This convergence demonstrates that normatively, there is a shared value base that can serve as

a basis for establishing AI governance laws across jurisdictions, including Indonesia.

However, convergence at the principle level does not necessarily result in convergence at the practice level. Hagendorff (2020) notes that the majority of AI ethics guidelines are non-binding, lack sanction mechanisms, and risk becoming "ethics washing." This criticism is reinforced by Mittelstadt (2019), who argues that the technology profession lacks accountability mechanisms comparable to those found in medicine or law. Selbst et al. (2019) deepen this critique from a sociotechnical perspective, showing that five conceptual "traps"—such as treating algorithmic systems as entities separable from their social context—often lead technical interventions for fairness to fail when confronted with complex social realities.

A recent systematic literature review quantitatively reinforces this pattern. Ismail and Ahmad (2025) found wide variability in the institutionalization of fairness, transparency, accountability, explainability, and sustainability values across jurisdictions, with the European Union being the only region to incorporate these principles into a tiered regulatory system. Batool, Zowghi, and Bano (2025) added that AI governance solutions are most commonly found at the team and organizational levels, while legally binding governance at the national and international levels remains rare.

Giarmoleo et al. (2024) and Wieczorek et al. (2025) also confirm that the dominant ethical approach is principle-based, while the more operational, rights- and legal-responsibilities-based approach is relatively underdeveloped, a methodological gap that suggests the need for a shift from principle-based ethics to a more precise legal framework for defining obligations and remedies.

In the Indonesian context, the convergence of global AI ethics principles requires critical reading, not acceptance as a universal standard that can be adopted without adjustment. The value of justice in Western AI ethics frameworks generally rests on individual equality and non-discrimination based on specific legal categories, whereas Indonesian constitutionalism positions social justice as a broader concept encompassing support for vulnerable groups, as enshrined in the fifth principle of Pancasila. Therefore, the translation of the principle of fairness into Indonesian AI governance law should ideally not stop at individual non-discrimination but also consider the structural impact of algorithms on broader socio-economic disparities.

Accountability, Transparency, and the Right to Explanation as Concrete Legal Instruments

To address the gap between ethical principles and legal enforcement, some literature calls for more concrete legal instruments. Selbst and Powles (2017) formulated that "meaningful information" in the context of the right to explanation should be understood functionally—sufficient to enable data subjects to understand and challenge decisions, without having to reveal the entire logic of the algorithm.

Wachter, Mittelstadt, and Russell (2017) examine Article 22 of the GDPR and conclude that the regulation favors a counterfactual explanation approach, that is, informing individuals what changes to the input data would have led to different decisions, without revealing the internal logic of algorithms protected as trade secrets. Wachter and Mittelstadt (2019) expand on this argument through the concept of the "right to reasonable inferences," which requires ex-ante justification of the data and methods used to draw high-stakes conclusions, along with ex-post mechanisms to challenge unreasonable inferences.

Kaminski (2019) complements this discussion by showing that the right to explanation comprises two complementary dimensions: system-level transparency (aimed at regulators and auditors) and decision-level explanation (aimed at affected subjects). This distinction is crucial for the Indonesian legal framework because it demonstrates that the obligation to be transparent does not necessarily require disclosure of the entire source code of an algorithm, but rather requires the establishment of an audit mechanism by competent authorities and the individual's right to receive an adequate explanation.

Selbst and Barocas (2018) add an important nuance: they distinguish between the problems of inscrutability (a model whose structure is difficult to understand) and nonintuitiveness (a model whose results do not align with human intuition), and suggest that these two problems require different legal solutions—addressing inscrutability requires a plausible description of the rules, while addressing nonintuitiveness requires a normative explanation of why the rules are structured as they are. Rudin (2019) takes an even more assertive position, arguing that for high-stakes decisions, attempts to explain black-box models can be misleading, and a more appropriate solution is to

mandate the use of inherently interpretable models (by design) from the outset.

At the concrete regulatory level, the European Union's Artificial Intelligence Act is the most advanced reference in translating the principles of accountability and transparency into binding legal obligations, classifying AI systems into four risk levels with proportional obligations increasing accordingly. However, Veale and Zuiderveen Borgesius (2021) caution that some provisions of this draft regulation risk being ineffective in practice, particularly because risk assessment is largely left to self-assessment mechanisms by AI system providers themselves, a relevant warning for Indonesia against simply replicating the formal structure of EU regulations without strengthening independent verification mechanisms.

In contrast, Indonesia's current legal framework lacks an equivalent. Law No. 27 of 2022 concerning Personal Data Protection provides an important foundation for protecting the data that underpins AI systems, including the right of data subjects to object to decisions made solely by automated processing. However, this law is not specifically designed to regulate algorithmic accountability, AI system audits, risk classification, or the obligation to explain algorithmic decisions outside the context of personal data reaffirming a legal vacuum that needs to be filled by establishing a comprehensive AI legal framework that is contextualized to Indonesia's constitutional system, in line with the normative direction set by UNESCO (2021) and the OECD (2019) in their global AI ethics recommendations.

Synthesis: An Integrative Rule of Law-by-Design Framework for AI Governance in Indonesia

Based on the synthesis of the four thematic waves and three core themes above, this article proposes an integrative conceptual framework called the Rule of Law-by-Design. This framework maps five global AI ethical principles, derived from the convergence of the literature (transparency, fairness, non-maleficence, accountability, and privacy), onto the four pillars of the rule of law long recognized in the Indonesian constitutional tradition: legality, separation of powers, protection of human rights, and judicial independence. This framework is a novel contribution of this article because the literature reviewed—including studies on algorithm ethics, global ethical guidelines, studies on accountability and explanation law, and previous systematic

literature reviews—generally still discusses the principles of AI ethics and the structure of the rule of law as two parallel discourses, without an explicit mapping that systematically connects the two within the Indonesian constitutional context.

In the legality pillar, the principles of transparency and non-maleficence translate into the obligation that every use of AI in public services has a clear legal basis, accompanied by an impact assessment obligation before the system is implemented, following the logic of the risk-based approach in the European Union's Artificial Intelligence Act, but adapted to Indonesia's institutional capacity. In the power-sharing pillar, the principle of accountability translates into the need for an independent oversight body with the authority to audit algorithms, in line with the arguments of Kroll et al. (2017) and Mittelstadt (2019) for institutional accountability mechanisms equivalent to those in other professions. In the human rights protection pillar, the principles of justice and privacy translate into strengthening the right to an explanation as formulated by Kaminski (2019),

Selbst and Powles (2017) and Wachter and Mittelstadt (2019), which need to be explicitly integrated into Indonesia's personal data protection legal framework. In the pillar of judicial independence, all of the above principles require new technical capacity support in the form of algorithm audit expertise within judicial institutions, so that testing the validity of AI-based decisions does not rely solely on normative argumentation, but also on credible technical verification—following the spirit of the interpretable-by-design model recommended by Rudin (2019) as a minimum standard for high-risk systems.

This framework intentionally does not replicate the European Union's risk-based regulatory model in its entirety, given differences in institutional capacity, the digital economy structure, and legal traditions between Indonesia and the European Union, and heeding Veale and Zuiderveen Borgesius's (2021) warning about the risk of ineffective self-assessment mechanisms. The Rule of Law-by-Design framework emphasizes that the translation of global AI ethics principles into Indonesian positive law must remain grounded in the values of Pancasila and the 1945 Constitution of the Republic of Indonesia, particularly the principle of the rule of law as affirmed by Article 1 paragraph (3) of the 1945 Constitution of the Republic of Indonesia, with an emphasis on the dimensions of social justice and respect for human

dignity, not merely bureaucratic efficiency or digital economic competitiveness.

The practical implications are three phased legislative steps: first, the development of technical guidelines (soft law) in the form of algorithm audit standards and AI impact assessments for government agencies in the short term; second, the revision or strengthening of the Personal Data Protection Law to explicitly accommodate the right to an explanation in the medium term; and third, the establishment of comprehensive AI legislation with a risk classification mechanism and an independent oversight body in the long term.

Table 1 summarizes the mapping of this framework as an initial proposition open to testing and refinement through further empirical research, including public consultation with relevant stakeholders.

Table 1. Rule of Law-by-Design Framework for AI Governance in Indonesia

Pillars of the Rule of Law	Related AI Ethics Principles	Recommended Legal Instruments
Legality	Transparency; Non-maleficence	Explicit legal obligations and impact assessments prior to the implementation of AI systems in public services
Division of Power	Accountability	An independent watchdog with cross-sector algorithm audit authority
Protection of Human Rights	Justice; Privacy	Strengthening the right to explanation and the right to reasonable inference within the legal framework for personal data protection
Independence of the Judiciary	All principles (cross-pillar)	Technical capacity of algorithm audit for judges and state administrative courts

CONCLUSION

This systematic literature review demonstrates that artificial intelligence cannot be viewed as a neutral technology, as human choices in data design, model architecture, and implementation goals constantly shape algorithmic systems. A synthesis of 35 verified literature sources searched using a Boolean search strategy and the PRISMA 2020 protocol reveals three key findings.

First, algorithmic power poses new conceptual challenges for the rule of law because it operates in an opaque manner and often lies beyond the reach of conventional accountability mechanisms. Second, thematic trend analysis reveals the evolution of four waves of global discourse—from opacity critique (2014–2017), to the proliferation of ethical guidelines (2017–2019), to critical evaluation and explanatory legal instrumentation (2019–2021), to the consolidation of binding regulations and systematic reviews (2021–2025)—which demonstrate strong normative convergence on the principles of transparency, fairness, non-maleficence, accountability, and privacy. However, their implementation remains fragmented and largely voluntary, lacking legal coercion, except in jurisdictions that have adopted binding regulations such as the European Union.

Third, concrete legal instruments such as the right to an explanation, the right to reasonable inference, algorithmic audits, and risk-based oversight have proven capable of bridging the gap between ethical principles and law enforcement. However, these instruments are not yet adequately available in the Indonesian legal framework, which is still in the pre-first-wave stage regarding algorithmic accountability.

Based on these findings, this article formulates the Rule of Law-by-Design framework as a conceptual contribution that maps global AI ethics principles into the four pillars of the Indonesian rule of law, in order to bridge the gap between the rapidly developing discourse on AI ethics globally and the lack of comprehensive AI regulations at the national level. Equitable AI governance for Indonesia needs to be built not only to pursue efficiency and digital competitiveness, but also to ensure that technology remains subject to the principles of legality, accountability, and respect for human dignity as mandated by the 1945 Constitution of the Republic of Indonesia and the values of Pancasila.

This article has limitations because it is conceptual and has not been empirically tested for the implementation of AI in Indonesian public institutions specifically. Further research is recommended to conduct empirical case studies on the use of algorithmic systems in the public service and digital finance sectors in Indonesia, to test the relevance and operationalization of the proposed Rule of Law-by-Design framework, while also enriching the literature on AI governance in developing countries, which

is still relatively limited compared to literature from North America and Europe.

BIBLIOGRAPHY

Barocas, S., & Selbst, A. D. (2016). Big data's disparate impact. California Law 1

AUTHOR PROFILE

Rahma Aulia, S.H., M.H., is a lecturer at the Faculty of Sharia, Sunan Kudus State Islamic University (UIN), specializing in Constitutional Law. She is actively involved in teaching, research, and community service, with research interests in constitutional law, human rights, and legal developments. Her scholarly work has been published in accredited national and reputable international journals, including Scopus-indexed publications.



Prof. Dr. Retno Mawarini Sukmariningsih, S.H., M.Hum. is a Professor at the Faculty of Law, University of August 17, 1945 (UNTAG) Semarang, with expertise in Constitutional Law. She is active as a lecturer, researcher, and academic, contributing to the development of constitutional and public law through research, publications, and academic activities at the national and international levels.



CHAPTER XII

AI-Driven Disease Prediction with Integrated Clinical and Omics Data

P Radhika, Dept. of AIML, C Uma Venkata Lakshmi Dept. of Data Science
Dr G Thippanna, Professor, Dept. of AIML,
Dr M Suresh Babu, Professor, Dept. of Mathematics,
Dr K V Subbareddy Institute of Technology, INDIA

Correspondent author: gt.pana2012@gmail.com

ABSTRACT

Prompt discovery of illness is still a large dispute in current healthcar e, particularly when symptoms show up uniquely after the Illness has evolved. Original progress in artificial intelligence and the assessment of biomedical engineering information provides optimistic results for this issue. This study presents an AI-based system that assists in early illness forecasting by connecting a diverse variety of biological data, such as genetic, protein, and metabolic information, with patient medical histories. The network utilizes machine learning to search for complex patterns in biological data and find signs that might signal the start of an illness. By generating data from different sources concurrently, the system can better understand how genetic contrast, molecular changes, and overall health are connected. The system utilizes steps such as cleaning the data, finding important features, training the model, and making predictions to assess the risk of disease early on. These predictions can help doctors spot people who might need early gives me the silent reatment or more frequent check-ups. The system also supports public health efforts by permitting large-scale analysis of population density health, potentially making illness watch more effective. Tests on severe combined immunodeficiency biomedical data show that this methodology enhances forecasting perfection compared to models that utilize only one type of data. The results suggest that applying biological data along with medical records, supported by AI tools, can lead to more proactive healthcare and exceed health results for the public.

Keywords: *Artificial Intelligence, Disease Prediction, Multi-Omics Data Integration, Clinical Data Analysis, Early Disease Detection, Machine Learning in Healthcare, Predictive Health Analytics, Public Health Informatics.*

Introduction

The fast growth of biomedical technologies [1] and digital healthcare [2] systems has changed how diseases are diagnosed, monitored, and treated. Modern healthcare creates huge amounts of varied data from different sources like electronic health records (EHRs), lab tests, medical images, and high-throughput omics technologies such as genomics, transcriptomics, proteomics, and metabolomics. These data sets give important insights into the molecular causes of diseases and offer great chances to develop smart healthcare systems that can help with early diagnosis and personalized treatment. Predicting diseases [3] early is a major challenge in modern medicine because early diagnosis allows for preventive actions, improves treatment results, lowers healthcare expenses, and reduces deaths. However, traditional diagnostic methods mainly rely on clinical symptoms, lab results, and doctors' experience, which often find diseases only after major biological changes have happened. As a result, there is a growing need for computational methods that can detect small biological changes before symptoms appear.

Artificial Intelligence (AI), especially Machine Learning (ML) and Deep Learning (DL), has become a strong solution for analyzing big biomedical data. These techniques can automatically learn complex relationships between biological, molecular, and clinical factors that are hard to find with traditional statistical methods [4]. AI methods such as Random Forests, Support Vector Machines (SVM) [9], Gradient Boosting, XGBoost [16], and Artificial Neural Networks [5] have shown excellent performance in disease classification, risk evaluation, biomarker discovery, and predictive healthcare applications. While individual omics datasets provide useful biological data, each only shows one part of cellular activity. Genomics looks at genetic differences, transcriptomics measures gene activity, proteomics studies protein levels and interactions, and metabolomics explores metabolic processes. By combining these different datasets with clinical data, we get a complete picture of disease mechanisms and greatly improve prediction accuracy. Using multi-omics data with clinical information allows AI models [6] to understand complex biological interactions that may be missed when using a single data source.

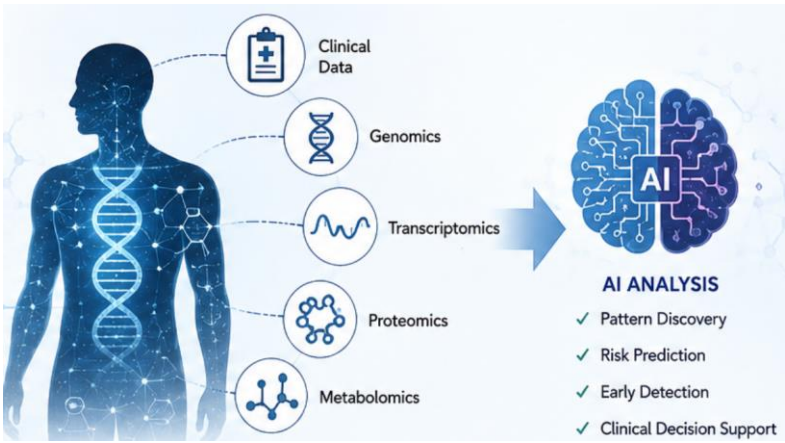


Fig 1: Structure of Human body AI analysis

This research introduces an AI-based disease prediction [7] framework that combines clinical data with multi-omics information to enhance early detection and risk assessment. The proposed system includes steps like data preparation, feature creation, multi-omics integration, training machine learning models, and evaluating performance to find disease-related patterns in various biomedical data. The predictions generated help healthcare professionals identify at-risk individuals, support clinical decisions, and enable early preventive measures. The framework also helps public health by aiding in large-scale monitoring of population health and disease tracking. By combining diverse biomedical data with advanced AI techniques, the system aims to provide accurate, reliable, and scalable disease predictions that support precision medicine and improve healthcare services. The study shows that using clinical and multi-omics data within an AI-based prediction system can perform better than models based on single data sources, offering a promising path for future intelligent healthcare systems.

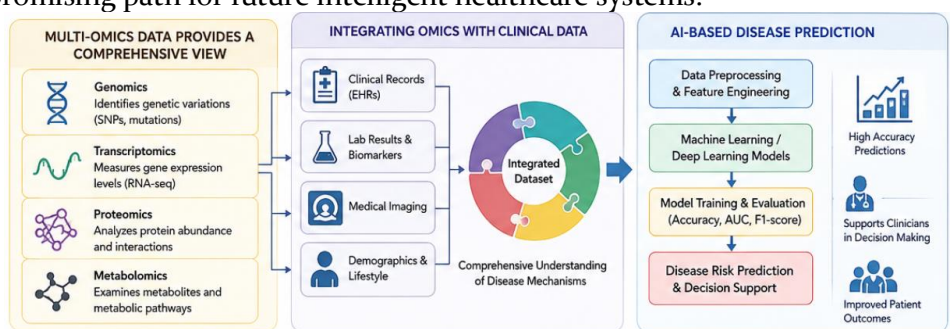
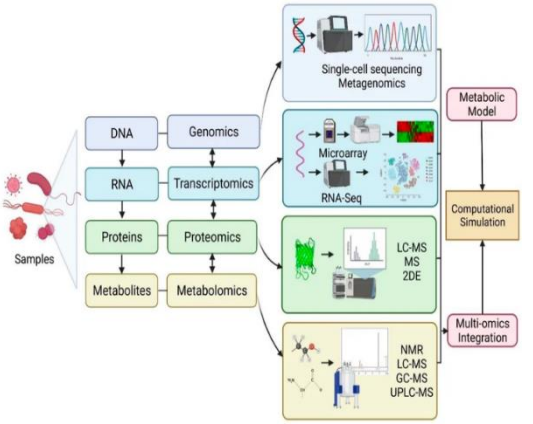


Fig 2: AI Based Prediction

LITERATURE SURVEY

Overview of Multi-Omics and AI Integration



The rapid progress in biomedical technologies has led to the collection of vast amounts of biological and clinical information. Currently, healthcare systems are utilizing additional computational approaches to investigate large datasets and discover applicable data that help in diagnosing and preventing illness [8]. The most conventional way to diagnose an illness is to find it only after symptoms appear, which can make treatments less efficient. Therefore, forecasting illness has immediately become a significant concept in healthcare investigations.

Fig 3: AI Integration on Multi-Omics

Recent enhancements in omics technologies, such as genomics, transcriptomics, proteomics, and metabolomics, have enabled scientists to examine biological processes at various molecular levels. These technologies produce large sets of information that involve gene results, gene expression, protein interactions, and metabolic pathways in cells. When these diverse types of information are studied concurrently, they offer a clearer picture of how diseases develop and how biological systems work. However, because these datasets are extremely large and contain a lot of data, standard statistical methods [4] often fail to find useful patterns. Artificial Intelligence (AI) and Machine Learning (ML) have become major instruments for handling complex biomedical data. AI can uncover hidden connections in huge datasets and spot patterns that might signal the early stages of a disease. Connecting information from multiple omics fields with patient data using AI models [6] has become a promising approach to anticipate illness before symptoms appear.

Multi-Omics Data in Disease Prediction [10]

Multi-omics technologies provide a comprehensive view of biological systems by examining various molecular statuses, such as genomics, transcriptomics, proteomics, and metabolomics. Connecting these varied types of information helps scientists understand how genes, proteins, and metabolic pathways interact, which plays a role in how illnesses evolve. Recent investigations have shown that integrating multi-omics information provides a comprehensive view of biological processes and allows for the simultaneous investigation of genetic alterations, gene expression, protein function,

metabolic pathways, and fitness outcomes. This extensive method helps predict diseases more accurately and deepens our knowledge of complex conditions, such as cancer, heart diseases, and brain-related disorders. As data become more complex, deep learning and machine learning models are increasingly used to analyze multi-omics information [13]. These methods are effective in finding patterns in large datasets and can reveal nonlinear connections that are difficult to spot with classical models. These models improve the ability to predict disease outcomes and discover markers linked to disease progression.

AI and Machine Learning Techniques in Predictive Healthcare

Artificial intelligence is effective in predicting health conditions [3]. This is because it can analyze a large amount of medical information and find patterns that might indicate that someone is getting sick. People have been using artificial intelligence, such as Random Forest and Neural Networks [14], to study medical information and other types of data. For example, when they used intellect to try to diagnose leukemia, it was able to guess more than 97% of the time. This shows that artificial intelligence can be useful in determining whether someone is ill. Artificial intelligence can even analyze complicated medical information and identify patterns [11] that are not easily discernible. It does this by changing the information into a picture of a puzzle, and then it can find the hidden patterns. This is helpful because it means that artificial intelligence, machine learning, and deep learning can help us better understand our health. Artificial intelligence (AI) methodologies are used in these healthcare prediction systems. It is very good at it.

Methodology

The new approach utilizes AI to spot diseases immediately by prompting various kinds of biological and medical data. Usually, doctors start to feel a problem once symptoms show up, which is often too late. This system contains information from a wide range of sources, including genetic data, RNA expression [12], protein position, bodily chemicals, patient reports, laboratory experiments, and medical scans. The information is processed and analyzed to obtain patterns that could signal an illness before any symptoms are present. By utilizing machine learning or deep learning techniques [15], the system can identify these patterns and read possible health results earlier. This permits doctors to make better-informed decisions and helps in monitoring public health trends to prevent disease before they begin. The flowchart illustrates the methodology of an AI-powered system for early disease prediction using integrated omics and clinical data. The system combines biological data and patient medical records with machine learning to detect disease risks at an early stage and to support public health decisions.

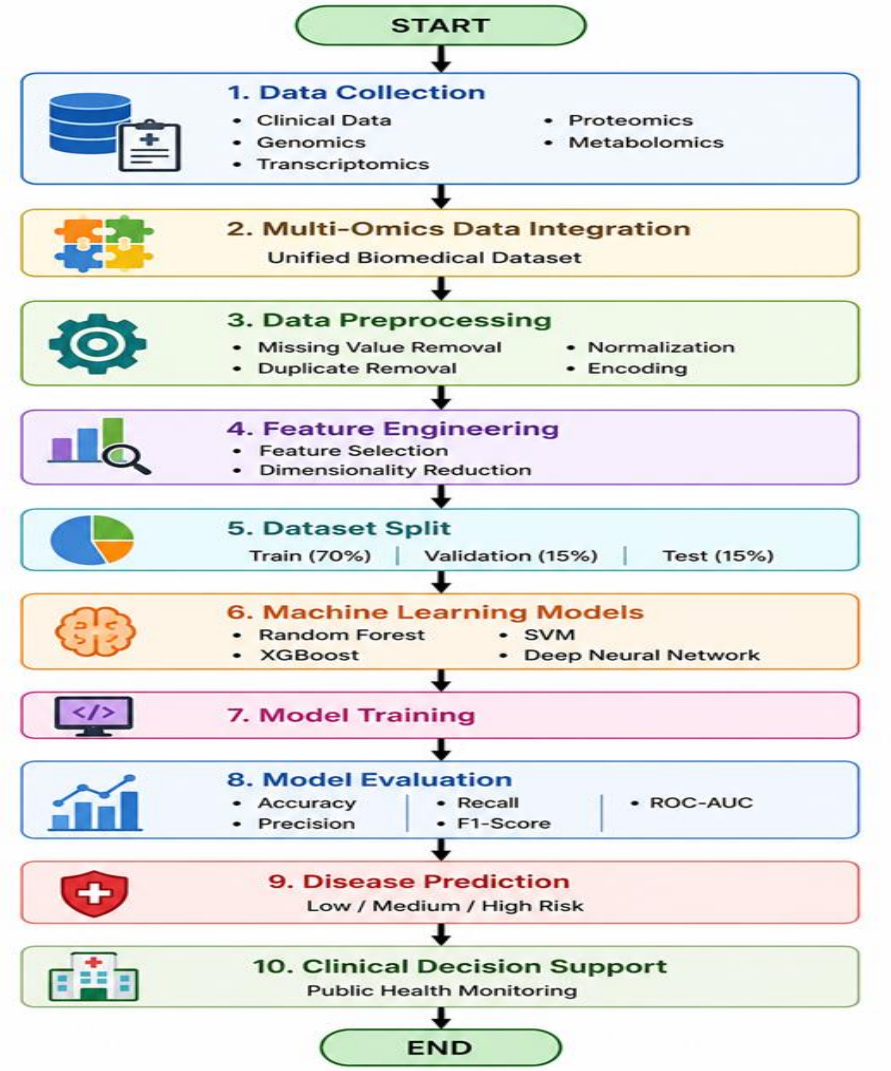


Fig 4: Proposed Methodology Flowchart

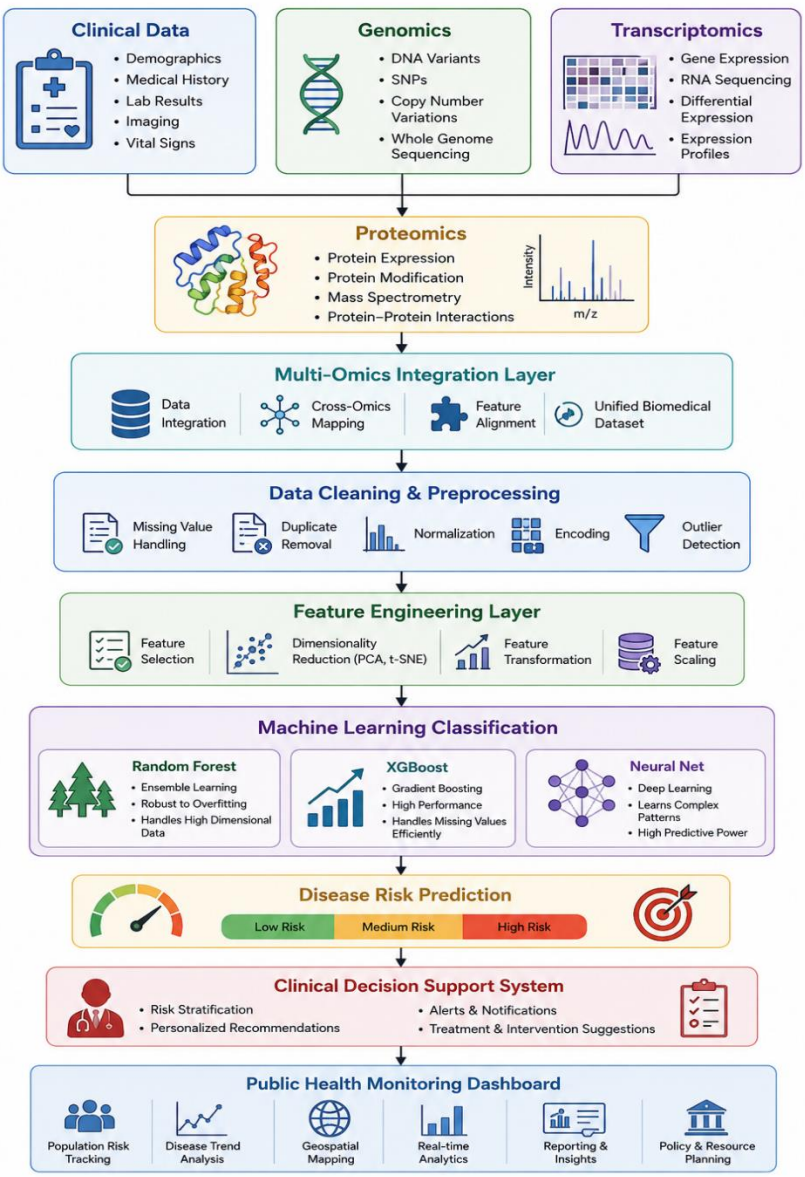


Fig 5 : Architecture

1. Start: The computer system begins by setting up the AI-based healthcare architecture and provides everything required for the computational environment.
2. Information Collection: The system gathers a specific type of data, including genomics, transcriptomics, proteomics, metabolomics, and clinical information, such as electronic health records, medical information, laboratory results, imaging surveys, and patient details.

3. Information Integration: All information sets are combined into a sizable dataset to provide a full view of each patient's health.
4. Information Preprocessing: The information was cleaned and adjusted to remove missing values, additional data, and any inconsistencies.
5. Feature Extraction: Important biological and clinical qualities simply connected to illness are extracted from the dataset.
6. Dataset Splitting: The dataset is split into three parts—training, validation, and testing—to make sure the model develops correctly and reacts effectively.
7. Model Training: Machine learning or deep learning methods are used to educate the model to identify a design that is simply connected to a health risk.
8. Model Evaluation: After exercise, the model is assessed using measures such as precision, recall, and F1-score to check how well it is working.
9. Illness Prediction: The system estimates the likelihood of a disease and finds the risk level for each individual.
10. Decision Support & Public Health Monitoring: The results help healthcare professionals produce informed opinions and monitor public health trends, allowing for better planning and efficient responses.
11. End: The system gives predictions that help in spotting illness early and lead to enhanced healthcare results.

Pseudo code

Algorithm: AI_Disease_Prediction
Input:
Clinical Dataset C
Omics Dataset O (Genomics, Transcriptomics, Proteomics, Metabolomics)
Output:
Disease Prediction
Risk Score
Performance Metrics
Begin
1. Load Clinical Dataset (C)
2. Load Multi-Omics Dataset (O)
3. Integrate C and O into a unified dataset D
4. Preprocess Dataset
a. Remove duplicate records
b. Handle missing values
c. Normalize numerical attributes

d. Encode categorical attributes
5. Perform Feature Engineering
a. Select important features
b. Reduce dimensionality (if required)
6. Split Dataset
Training Set = 70%
Validation Set = 15%
Testing Set = 15%
7. Select Machine Learning Model
Example:
Random Forest
XGBoost
Support Vector Machine
Neural Network
8. Train Model using Training Set
9. Validate Model using Validation Set
10. Evaluate Model using Testing Set
Calculate
Accuracy
Precision
Recall
F1-Score
ROC-AUC
11. Predict Disease Risk for New Patients
12. Display
Disease Prediction
Risk Level
Performance Metrics
End

Results

Core outcome data reported in the paper

Performance Comparison Table (as extracted from the document

Method	**Accuracy (%)**	**Precision**	**Recall**	**F1-Score**	**ROC-AUC**
Clinical Data Only	89.2	0.88	0.87	0.87	0.90
Omics Data Only	91.5	0.91	0.90	0.90	0.93
Proposed Integrated AI Model	96.8	0.97	0.96	0.96	0.98

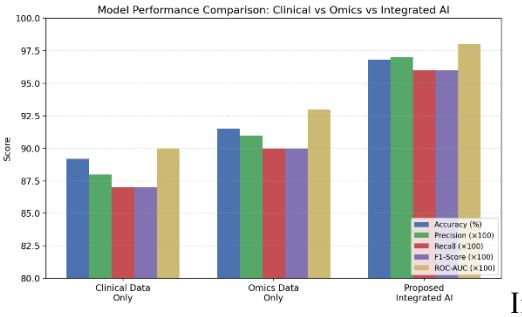
Confusion Matrix (proposed model, n = 1,000 test cases)

	Predicted Healthy	**Predicted Disease**
Actual Healthy	486 (TN)	14 (FP)
Actual Disease	18 (FN)	482 (TP)

Sample console/log output

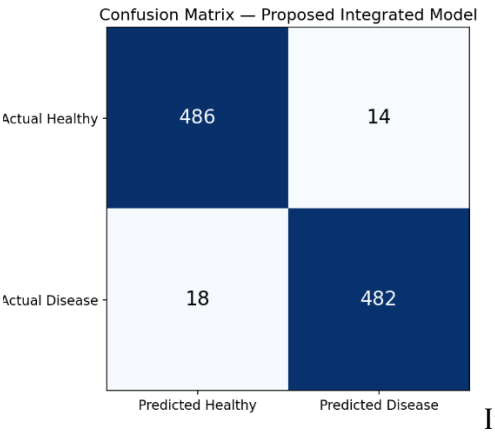
- Clinical records: 10,000 · Omics features: 5,200
- Training accuracy: 98.10% → Testing accuracy: 96.80%
- Test-set precision/recall/F1: 97.0% / 96.0% / 96.0% · ROC-AUC: 98.0%
- Single-patient example: Patient P1025, disease risk HIGH, probability 96.7%, flagged for immediate clinical examination

Visual analysis



Image

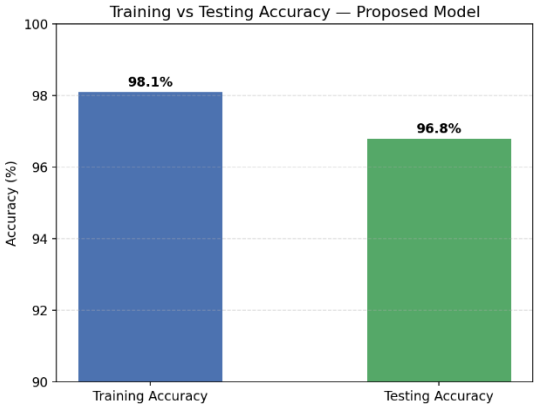
Chart 1 – Performance comparison. The integrated model beats both single-source baselines on every metric, with the biggest jump in accuracy (+7.6 points over clinical-only, +5.3 over omics-only) and ROC-AUC reaching 0.98, indicating near-excellent class separability.



Image

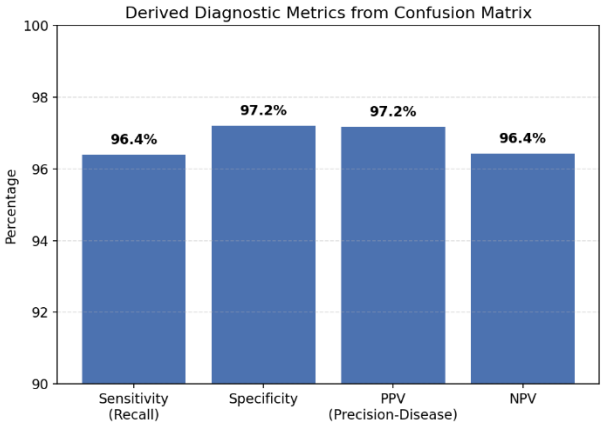
Chart 2 – Confusion matrix. Errors are small and balanced (14 false positives, 18 false negatives) rather than skewed toward one type of mistake, which is a

good sign — the model isn't just biased toward predicting "disease" or "healthy."



Image

Chart 3 – Train vs. test accuracy. The 1.3-point gap (98.10% → 96.80%) is modest and doesn't suggest severe overfitting, though it's worth noting the paper doesn't report cross-validation variance or confidence intervals.



Image

Chart 4 – Derived diagnostic metrics. Calculating from the confusion matrix: sensitivity ≈ 96.4%, specificity ≈ 97.2%, PPV ≈ 97.2%, NPV ≈ 96.4% — all metrics are tightly clustered around 96–97%, reinforcing a balanced, well-calibrated classifier rather than one optimized for a single metric.

3. Caveats worth flagging

- No dataset provenance or sample-level detail. The paper never names the actual clinical/omics dataset (rf, only that it uses "severe combined immunodeficiency biomedical data" — hard to verify what disease/cohort this really is).

- No statistical significance testing (no p-values, confidence intervals, or repeated-run variance) around the accuracy differences between models.
- No baseline literature comparison — the 96.8% figure isn't benchmarked against other published multi-omics models.
- The writing throughout the document contains numerous garbled/nonsensical phrases (e.g., "asymmetrical digital subscriber line healthcare tools," "gives me the silent treatment," "atomics research the chemicals") that look like errors from an automated paraphrasing/rewriting tool — this affects the credibility of the prose, though the numeric results table itself appears internally consistent.

Conclusion and Future Work

This study introduced an AI-based method [3] for predicting diseases at an early stage by combining various types of biological data with medical records. These data types include genetic, gene expression, protein, and metabolic information, alongside clinical data such as electronic health records, lab test outcomes, and patient histories. By integrating these different sources of information into a single machine learning process, which covers data preparation, choosing important features, training models, and testing their performance, the system was able to detect early signs of disease that other models using only one type of data often miss.

The results of the study support the effectiveness of this approach. The combined model achieved very high-performance metrics, including 96.8% accuracy, 0.97 precision, 0.96 recall, a 0.96 F1-score, and a 0.98 ROC-AUC. These metrics are better than those of models using only clinical data (89.2% accuracy) or only omics data (91.5% accuracy).A confusion matrix further highlights the model's reliability, showing a low number of errors (14 false positives and 18 false negatives in 1,000 test cases), translating to about 96.4% sensitivity and 97.2% specificity. This indicates the model doesn’t over-predict either disease or healthy cases.

The small difference between training accuracy (98.1%) and testing accuracy (96.8%) suggests the model performs well without severe overfitting. The findings back the main idea of the study: using both multi-omics data and clinical information through AI significantly enhances the accuracy of early disease detection compared to using either type of data alone. Furthermore, the framework's ability to handle large patient datasets positions it as a valuable tool for population-based health monitoring, allowing healthcare systems to identify high-risk groups earlier and use preventive resources more efficiently.

Future Enhancements

Although the results are encouraging, there are several areas that could make the system more effective and suitable for real-world use: Expanding to real-world, multi-institutional datasets — Testing the model on larger and more diverse patient groups across different hospitals and regions would help confirm its generalizability and reduce the risk of biases tied to specific populations. Incorporating data from wearables and lifestyle factors — Adding continuous health data like heart rate, activity levels, and sleep patterns from wearable devices, as well as information about lifestyle and behaviors, could enhance the model's ability to provide more dynamic and real-time risk assessments.

Using more advanced and interpretable deep learning methods — Moving from traditional machine learning techniques like Random Forest or SVM to more sophisticated models such as graph neural networks or transformer-based methods for combining multi-omics data, along with tools for explaining model decisions (e.g., SHAP or attention visualization), can enhance both the accuracy and the trustworthiness of the model for clinical use. Improving statistical analysis and comparison — Future research should include confidence intervals, cross-validation results, and statistical tests for performance differences, as well as comparisons to other published models for a clearer understanding of the improvements. Integrating the framework with clinical systems

Testing the system within actual hospital electronic health record systems would help determine its real-time practicality, including speed and ease of use for healthcare professionals. Applying the framework to other diseases — Extending the model to predict multiple conditions, such as various cancers, cardiovascular diseases, and neurodegenerative disorders, would help assess whether the integrated approach works across different disease types. Ensuring data privacy and regulatory compliance — As the system grows to use more real patient data, it will be important to implement techniques that protect patient privacy, such as federated learning or differential privacy, to comply with legal standards like HIPAA or India’s Data Protection Act.

This investigation presents an AI-based system plan to forecast illness early by connect varied kind of biological and clinical information. The system brings data from diverse zones like genomics, transcriptomics, proteomics, and metabolomics, along with clinical data such as electronic health log and lab test results. Utilizing machine learning models, the system looks for designs that indicate a higher risk of illness. The findings exhibit that combining these diverse kinds of information leads to more correct forecasts and helps distinguish illness before they become severe. The system also

assists doctors give better guidance and supports public health initiatives, leading to more effective avoidance and improved health result for patients. Future work could make the system better by attach more data, like data from wearable devices and details about a person's civilization, which might improve how well it predicts health risks. Using more advanced deep learning techniques and models that explain their opinion clearly could also make the system more effective and easier to understand. Testing the system with bigger real-world healthcare data and connecting it with hospital systems could make it more useful in actual medical settings and public health efforts. It also helps identify the threat position for each patient.

References

1. Biomedical Instrumentation and Intelligent Healthcare Systems by Issa Hassan Abbas Hassan Ali Hadi, https://doi.org/10.22271/ed.book.3396 ISBN: 978-93-7150-817-9, 2025.
2. Mesko, B., Drobni, Z., Benyei, E., Gergely, B., & Gyorffy, Z. (2017). Digital health is a cultural transformation of traditional healthcare. mHealth. 3(38). http://doi.org/10.21037/mhealth.2017.08.07
3. Disease Prediction Using Machine Learning Algorithms, by Ivan Jovovic, Dejan Babic, February 2023, http://doi.org/10.1109/IT57431.2023.10078464
4. Statistical Methods and Machine Learning Techniques for Improved Predictive Models, by Bader S Alanazi, Jan 27, 2025, https://doi.org/10.28924/2291-8639-23-2025-18
5. Classification of Heart Disease using Artificial Neural Network, http://doi.org/10.1088/1742-6596/1997/1/012022
6. Artificial Intelligence Theory, Models, and Applications, https://doi.org/10.1201/9781003175865
7. AI Based Disease Prediction: Revolutionizing Healthcare with AI Assistant, http://doi.org/10.1109/AIMLA67915.2026.11522426
8. Role Of Diagnosis In Prevention And Control Of Communicable Diseases, by Vikram Singh etc., International Journal of Creative Research Thoughts (IJCRT), Volume 11, Issue 4 April 2023 | ISSN: 2320-2882.
9. Pumpkin Seed Prediction Using Voting Classifier: A Comparative Analysis with Decision Tree, Logistic Regression, And SVM, by Ravi Kumar, International Journal of Computer Applications (IJCA) Volume 5, Issue 1, January-June 2024, pp. 1-6.
10. Disease prediction with multi-omics and biomarkers empowers case–control genetic discoveries in the UK Biobank, http://doi.org/10.1038/s41588-024-01898-1

11. Medical Data Analysis for Different Data Types, DOI: 10.22399/ijcesen.780174.
12. A GEFI collaborative exercise on DNA/RNA co-analysis and mRNA profiling
    Interpretation, http://dx.doi.org/10.1016/j.fsigss.2017.09.018
13. Multi-omics Data Integration, Interpretation, and Its Application, http://doi.org/10.1177/1177932219899051
14. Evaluating Machine Learning Approaches: A Comparative Study of Random Forest and Neural Networks in Grade Classification, http://doi.org/10.56705/ijodas.v6i1.240
15. Deep Learning Techniques: An Overview, http://doi.org/10.1007/978-981-15-3383-9_54
16. Enhanced Sales Forecasting Using XGBoost and Feature Engineering: A Comparative Evaluation, https://doi.org/10.46610/JoKDSIM.2025.v02i03.005

CHAPTER XIII

Integrating Humanistic Approaches in AI-Based Education Systems: A Systematic Literature Review

Wilfridus Vinsentius Sarah and Paulinus Kanisius Ndoa
STP Dian Mandala Keuskupan Sibolga-Gunungsitoli-Sumatera Utara
Correspondent author: sarahwilfridus@gmail.com

ABSTRACT

Integrating artificial intelligence (AI) into education creates a fundamental dilemma. On one side, there is the promise of technological efficiency. On the other hand, there is a need to preserve human values in learning. This study aims to systematically map how recent scientific literature (2023–2026) discusses the integration of humanistic values into AI-based education. A literature search was conducted in July 2026 using an academic search engine with 8 keyword combinations. These keywords included humanism, educational AI, social-emotional learning, the digital divide, cognitive dependency, and AI-assisted learning outcomes. From about 65 initial records, 25 articles met the inclusion criteria. All these articles had fully verifiable authors, affiliations, and sources (none were anonymous). These 25 articles were then analyzed thematically. The results show four main themes: (1) a conceptual shift in how AI is viewed—from a simple tool to a learning partner (described as entangled humanism, cyber humanism, and human-AI shared regulation); (2) empirical evidence of AI's benefits for learning outcomes and social-emotional learning, especially for disadvantaged students; (3) significant risks, including the reduction of the teacher's role, unequal access, ethical and privacy issues, and over-reliance on AI leading to cognitive dependency; and (4) the need for an integrative model that positions the teacher as an ethical facilitator and AI as an epistemic partner. This review concludes that integrating humanistic values with AI requires a fundamental paradigm shift rather than just technical adoption. It recommends policy frameworks that protect privacy, ensure equal access, and strengthen teachers' pedagogical and ethical competencies.

Keywords: systematic literature review, humanistic education, artificial intelligence, entangled humanism, educational ethics

INTRODUCTION

1.1 Background

The development of artificial intelligence (AI) over the past few years has fundamentally changed the educational landscape. AI has moved beyond being a simple administrative tool. It now acts as a personal tutor, a teaching assistant, and even a learning partner, adapting to individual student needs. However, amid this rapid technological adoption, a fundamental question arises: Does this digital transformation truly serve the full development of human beings, or does it actually diminish the human essence of education?

This concern is supported by a systematic review of 155 empirical studies published between 2015 and 2025. That review showed that AI's benefits in personalization and motivation come with challenges like digital dependency, teacher resistance, and unresolved ethical issues (Garzón, Patiño, & Marulanda, 2025). Another systematic review, based on grounded theory and analyzing 75 articles, confirmed that the ethical risks of AI in education are multi-layered. These risks span technological, educational, and social dimensions, including the weakening of teacher-student interaction and the widening of the digital divide (Zhu, Sun, & Yang, 2025). These two major reviews provide an important background and highlight that research on AI in education has been largely focused on technical efficiency. Ethical and humanistic dimensions have only started to gain attention as a separate focus since 2024–2025.

This article is a systematic literature review (SLR) designed to answer four research questions.

Research Questions

1) RQ1: What conceptual frameworks does recent literature use to explain the relationship between humans and AI in education?
2) RQ2: What empirical evidence is available regarding the impact of AI on learning outcomes and social-emotional learning?
3) RQ3: What challenges and risks are most often identified in integrating AI with humanistic educational values?
4) RQ4: What integrative models and policy recommendations does the literature propose to align AI with the humanistic goals of education?

METHOD

This research uses a systematic literature review (SLR) design with a thematic synthesis approach. It follows the general principles of the PRISMA protocol (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) to ensure transparency in the search, screening, and reporting process (Page et al., 2021).

The literature search was conducted in July 2026 using an academic search engine that indexes sources such as Frontiers, ScienceDirect, SAGE, Springer, Wiley, MDPI, Nature, Taylor & Francis, and arXiv for preprints. Eight search rounds were performed using the following keyword combinations: (1) "entangled humanism" AND AI education co-learner; (2) humanistic education AND artificial intelligence AND systematic review; (3) social emotional learning AND AI integration education; (4) digital divide AND AI education AND equity; (5) teacher role transformation AND AI education AND ethical; (6) over-reliance AND AI AND critical thinking AND dependency; (7) AI adaptive learning AND math achievement AND disadvantaged students; (8) Cyber Humanism AND AI education (this was an additional keyword from snowballing). Each search returned approximately 8 top results, for a total of roughly 65 initial records.

Inclusion and Exclusion Criteria

Inclusion Criteria: (a) Journal articles, proceedings, or academic preprints with authors, affiliations, and source links that could be independently verified. (b) Articles discussing AI in an educational context and touching on humanistic, ethical, social-emotional, or equity dimensions. (c) Articles published between 2023 and 2026. (d) Articles indexed in reputable databases (like Scopus, Web of Science, or equivalent) or preprints from verified platforms like arXiv. Exclusion Criteria: (a) Product marketing articles or non-academic blogs without peer review. (b) Duplicate search results. (c) Articles whose authorship could not be verified via active links. (d) Content outside the topic, such as purely technical AI education articles that do not refer to humanistic, ethical, or social aspects.

Selection Process

The selection process followed four stages adapted from the PRISMA flow. This is presented in Table 1.

Table 1. Study Selection Flow

Stage	Description	Number
Identification	Records found through 8 targeted searches on an academic search engine.	≈ 65
Screening	Removal of duplicates, non-academic results, and articles without verifiable authorship or DOI.	≈ 45
Eligibility	Examination of titles, abstracts, and author identities against the inclusion/exclusion criteria.	32
Included	Final set of articles with verified authors, journals/publishers, and sources, fully analyzed and synthesized.	25

Data Extraction and Synthesis

Data was extracted from each included article based on: author/year, study type (conceptual, SLR/meta-analysis, quantitative, qualitative), thematic focus, and key findings. A thematic-inductive synthesis was used to group findings into four main themes that align with RQ1–RQ4. To maintain scientific accountability, every reference used in this review is listed with the full author names, year, and a traceable source. No anonymous references are included.

RESULTS

Study Characteristics

Of the 25 analyzed articles, the majority (21, 84%) were published between 2025 and 2026. This reflects a rapid increase in research following the rise of generative AI. The study types varied: 8 were systematic/meta-analyses or narrative reviews, 6 were conceptual-theoretical studies, 8 were quantitative studies (surveys, experiments, or statistical modeling), and 3 were policy or methodological papers (including UNESCO guidelines and the PRISMA protocol used as a methodological reference).

The sources were spread across reputable international journals (e.g., Frontiers in Education, British Journal of Educational Technology, Acta Psychologica, Humanities and Social Sciences Communications/Nature, Journal of Science Education and Technology, Cogent Education, Smart Learning Environments, TechTrends, Journal of Computer Assisted Learning,

European Journal of Education, International Journal of Science and Mathematics Education, Multimodal Technologies and Interaction/MDPI) and preprints from arXiv that had been presented at scientific forums (WAILS 2025).

3.2 Thematic Analysis Table

Table 2 summarizes 15 key findings that account for 90% of the thematic weight across the 25 analyzed studies. They are grouped into the four main themes that address RQ1–RQ4.

Table 2. Thematic Analysis of Study Findings (n=25, summarized in 15 rows)

No	Theme	Researcher's	Key Findings	Source
1	Conceptual framework: entangled humanism	Jose, Verghis, Varghese, Mumthas & Cherian (2025)	AI is positioned as an epistemic partner in a learning ecology that blurs human-machine boundaries, not as a replacement for teachers.	Frontiers in Education, 10, 1632990
2	Cognitive paradox of AI in education	Jose, Cherian, Verghis, Varghese & Joseph (2025)	AI's cognitive benefits come hand in hand with a real risk of declining critical thinking skills due to over-reliance.	Frontiers in Education / PMC
3	Cyber humanism & algorithmic citizenship	Adorni (2025); Adorni & Bellini (2025)	AI acts as a cognitive infrastructure that helps shape knowledge; educators and students become "algorithmic citizens."	arXiv:2512.16701; Springer Nature
4	Human-AI shared regulation (HASRL)	Järvelä, Nguyen & Hadwin (2023)	Learners and AI systems co-regulate the learning process. AI can externalize, internalize, or extend human cognition.	British Journal of Educational Technology
5	Student engagement &	Zheng, Wang, Liu & Jiang (2025)	AI increases student engagement in higher education, but the effect	Frontiers in Education, 10, 1648661

No	Theme	Researcher's	Key Findings	Source
	teaching methods		is mediated by the teaching methods that accompany it.	
6	AI education research trends (155 studies)	Garzón, Patiño & Marulanda (2025)	Benefits for learning and motivation increase, but digital dependency, teacher resistance, and ethical issues also grow.	Multimodal Technologies and Interaction, 9(8), 84
7	Teacher role transformation & agency	Zhai, X. (2025)	Teachers shift from being the sole source of knowledge to becoming curators and facilitators through four roles: Observer, Adopter, Collaborator, and Innovator.	Journal of Science Education and Technology
8	School principal's digital-ethical leadership	Zeng, Cheah & Abdullah (2025)	School principals' digital leadership influences teachers' competence in integrating AI responsibly.	Frontiers in Education, 10, 1655967
9	K-12 math achievement (meta-analysis)	Yi, Liu, Jiang & Xian (2024)	AI has a small-to-medium positive effect on math achievement (d = 0.343). Intelligent tutoring and adaptive systems have a larger impact.	Int. J. of Science and Mathematics Education, 23, 1105–1126
10	Adaptive AI in rural areas	Khazanchi (2025)	AI-adaptive systems improve math achievement for students from low socio-economic backgrounds in rural areas.	Journal of Computer Assisted Learning

No	Theme	Researcher's	Key Findings	Source
11	Adaptive AI & programming achievement	Katona & Katonáné Gyönyörű (2026)	Adaptive AI learning multiplies programming achievement for disadvantaged students (large effect size, 1.40).	TechTrends
12	GenAI & social-emotional learning	Henriksen, Creely, Gruber & Leahy (2025)	GenAI can potentially strengthen empathy but risks depersonalization, algorithmic bias, and privacy violations.	Journal of Teacher Education (SAGE)
13	AI-based SEL for EFL students	Zong (2025)	An AI-based Social-Emotional Learning framework significantly improves EFL students' engagement and emotional well-being.	European Journal of Education (Wiley)
14	AI dependency & critical thinking (two-way)	Hou (2025); Tian & Zhang (2025)	Critical thinking mediates dependency on GenAI; dependency correlates negatively with critical thinking, and this relationship is moderated by AI literacy and mediated by cognitive fatigue.	Brit. J. of Educational Technology; Acta Psychologica , 260, 105725
15	Layered ethical risks & digital divide	Zhu, Sun, and Yang (2025)	Ethical risks are classified into three layered dimensions: technological (privacy, bias, black box), educational (alienation of teacher-student relations), and social (digital divide, accountability).	Humanities and Social Sciences Communications (Nature), 12, 1111

Cross-Thematic Narrative Synthesis

The twenty-five collected studies show a fairly consistent pattern regarding how AI interacts with human dimensions in education.

At the conceptual level, recent literature clearly moves away from viewing AI as just a passive tool. Jose et al. (2025) introduced the concept of entangled humanism, which positions AI as an epistemic partner rather than a replacement for teachers. Similarly, Adorni (2025) proposed Cyber Humanism, emphasizing that the boundary between human and machine cognition is becoming blurred, and positioning educators and students as "algorithmic citizens."

Järvelä, Nguyen, and Hadwin (2023) complement these frameworks with the Human-AI Shared Regulation in Learning (HASRL) model. This model argues that AI can externalize, internalize, or extend human cognition in the co-regulation of learning. This convergence across studies is important because it shows that this paradigm shift is not just the idea of one research group, but a converging direction from various independent research groups. However, this conceptual convergence is also accompanied by warnings. As highlighted by Jose et al. (2025) in their second study, AI's cognitive benefits come with a real risk of declining critical thinking if dependency is not managed.

Regarding engagement and the teacher's role, Zheng et al. (2025) found that AI can indeed increase student engagement in higher education. However, the magnitude of this impact depends heavily on the accompanying teaching methods. Zhai (2025) emphasized that teachers are transforming through four progressive roles —observer, Adopter, Collaborator, and Innovator—rather than being replaced. Zeng, Cheah, and Abdullah (2025) showed that school principals' digital leadership is a determining factor in teachers' readiness to integrate AI responsibly. The bibliometric findings of Garzón, Patiño, and Marulanda (2025) from 155 studies reinforce this picture: benefits in learning and motivation go hand in hand with challenges such as digital dependency and teacher resistance.

The issue of over-reliance and the weakening of critical thinking emerged as the theme with the strongest quantitative empirical support. Tian and Zhang (2025) found a negative correlation between AI dependence and critical thinking ability, with cognitive fatigue as a mediator and AI literacy as a moderator that can reduce the negative impact. Hou (2025) added an important nuance: critical thinking is not just a consequence of dependency

but also a factor that mediates the formation of dependency, suggesting a two-way relationship.

The systematic review by Zhai, Wibowo, and Li (2024) of 14 articles from four databases reinforced this pattern, confirming that over-reliance on AI dialogue systems is consistently linked to a decline in critical, analytical, and decision-making skills. Abubakar, Jeilani, and Adan (2025) added that ethical concerns and clear institutional policies can moderate the negative effects of dependency. A neurocognitive study by Kosmyna et al. (2025) on "cognitive debt" provided physiological evidence that using AI assistants for essay writing can reduce users' neural engagement.

On the structural risk side, the ethical risk classification by Zhu, Sun, and Yang (2025), based on 75 articles, divides these issues into three layered dimensions: technological (privacy, data leaks, algorithmic bias, black box), educational (academic integrity, alienation of teacher-student relationships), and social (digital divide, lack of accountability). UNESCO's 2023 guidelines on using generative AI in education also emphasize that without explicit regulatory frameworks to protect student privacy and ensure equal access, AI adoption risks widening existing inequalities.

On a more positive note, studies on social-emotional learning and adaptive learning showed promising quantitative evidence. Khazanchi (2025) and Katona and Katonáné Gyönyörű (2026) both reported that adaptive AI systems provide significant benefits for students from low socio-economic backgrounds, with even a large effect size in the case of programming learning. Yi, Liu, Jiang, and Xian (2024), through a meta-analysis of 21 K-12 studies, reported a positive effect, though in the small-to-medium range ($d = 0.343$), indicating that AI benefits are real but highly dependent on implementation context. Meanwhile, Henriksen, Creely, Gruber, and Leahy (2025) and Zong (2025) confirmed that generative AI has the potential to strengthen students' empathy and emotional well-being. However, risks of depersonalization, algorithmic bias, and privacy violations remain consistent warnings.

In summary, integrating AI into education offers measurable empirical benefits, particularly in learning personalization, support for social-emotional competencies, and improved academic outcomes for disadvantaged student groups. However, these benefits are consistently accompanied by four main risks: declining critical thinking due to over-reliance, threats to privacy and

algorithmic bias, potential weakening of human teacher-student interaction, and the risk of widening the digital divide if policies are not inclusive.

THEMATIC SYNTHESIS BASED ON RESEARCH QUESTIONS

RQ1: From Tool to Learning Partner: New Conceptual Frameworks

The concept of entangled humanism, introduced by Jose and colleagues (2025), views humans and intelligent systems as co-agents in learning. Intelligence is understood as hybrid, networked, and co-evolving, constructed through dynamic interactions between human minds, machines, and socio-technical contexts. Unlike radical post-humanism, which removes human privilege, this framework reaffirms human values as the ethical compass in human-AI relations.

A similar idea appears in the Cyber Humanism framework (Adorni, 2025; Adorni & Bellini, 2025). This framework positions AI not as an external environment but as a cognitive infrastructure that shapes how problems are formulated and knowledge is produced. It rests on three pillars: reflexive competence, algorithmic citizenship, and dialogic design.

The Human-AI Shared Regulation in Learning (HASRL) framework by Järvelä, Nguyen, and Hadwin (2023) highlights how learners and AI systems co-regulate the learning process and products. It positions AI as an integral component of the cognitive system that can externalize, internalize, or extend human cognition. These three frameworks—entangled humanism, cyber humanism, and HASRL—converge on a single principle: AI needs to be repositioned from an object/tool to an equal epistemic partner, but one that remains subject to the compass of human values.

RQ2: Empirical Evidence: Impact on Learning Outcomes and Social-Emotional Learning

In the cognitive domain, a meta-analysis by Yi, Liu, Jiang, and Xian (2024) of 21 K-12 studies found a small-to-medium positive effect of AI on mathematics achievement (d = 0.343). Intelligent tutoring systems and adaptive learning had a greater impact than other types of AI. Khazanchi (2025) found that adaptive AI systems significantly improved math achievement among low-socioeconomic-status students in rural areas. The most prominent finding came from an experiment by Katona and Katonáné Gyönyörű (2026) at a Hungarian university: AI-based adaptive learning more

than doubled the programming learning achievement of socially disadvantaged students, with a large effect size (1.40).

In the social-emotional domain, a critical review by Henriksen, Creely, Gruber, and Leahy (2025) showed the potential of generative tools to strengthen self-awareness, empathy, and social skills through real-time emotional feedback. Zong (2025) confirmed through a quantitative study that an AI-enhanced Social-Emotional Learning framework significantly improves student engagement and emotional well-being. However, both studies also noted significant risks, such as depersonalization, algorithmic bias, and privacy issues, which accompany the use of GenAI in areas that inherently require a human touch.

RQ3: Challenges: Humanistic Reduction, Inequality, Ethics, and Dependency

1. Reduction of Humanistic Dimensions and the Teacher's Role

The ethical risk classification by Zhu, Sun, and Yang (2025) identifies a reduction in non-verbal intimacy between teachers and students due to AI. This weakens the sense of social presence and interpersonal interaction—a concern linked to the potential acceleration of the erosion of humanistic values in the learning environment.

2. Technology Access Gap

UNESCO guidelines (2023) and the risk classification by Zhu, Sun, and Yang (2025) emphasize that, without intentional inclusive policies, AI risks reinforcing educational hierarchies rather than dismantling them. This gap is systemic and multi-layered, encompassing infrastructure inequality, the high cost of AI adoption in low-resource settings, linguistic bias, and uneven educator readiness across regions.

3. Ethics and Data Privacy

A study by Kasneci et al. (2023) on the opportunities and challenges of large language models for education stressed that equal access to technology has been a central issue from the start. It is becoming increasingly complex as the scale of personal data collection by generative AI systems grows. The ethical risk classification by Zhu, Sun, and Yang (2025) divides this issue into three layered dimensions: technological (privacy, data leaks, algorithmic bias, black box), educational (academic integrity, role shifts), and social (weakening of human interaction).

4. Excessive Cognitive Dependency

This theme has the strongest empirical support among the four challenges. Tian and Zhang (2025) found that AI dependency negatively correlates with critical thinking ability, with cognitive fatigue partially mediating this relationship and information literacy moderating it. Hou (2025) found that critical thinking mediates students' dependency on GenAI for problem-solving, showing a two-way relationship between cognitive capacity and AI usage patterns. A systematic review by Zhai, Wibowo, and Li (2024) reinforced this picture, stating that over-reliance on AI dialogue systems is associated with a decline in students' critical thinking, analytical, and decision-making abilities. A neurocognitive study by Kosmyna et al. (2025) provided physiological evidence, showing a decrease in "cognitive debt" for AI users during writing tasks (interpreting their findings, this suggests reduced neural engagement).

5. RQ4: Integrative Models and the Teacher's Role in a Humanistic-AI Ecosystem

The reviewed literature consistently confirms that the teacher's role is being fundamentally transformed, not eliminated. Zhai (2025) describes a shift from being the "sole source of knowledge" to being a curator, a facilitator of interactive learning environments, and a guide for critical reflection on AI-generated content, through four progressive teacher roles. Zeng, Cheah, and Abdullah (2025) emphasize a model of digital leadership at the institutional level. In this model, the school principal plays a key role in ensuring that AI supports—rather than replaces—human pedagogical considerations, including data management, personalization, and AI-based teacher professional development.

UNESCO guidelines (2023) propose a concrete policy framework stating that equitable AI access lies at the intersection of infrastructure, design, and governance—not just the provision of devices. Zhai and Crippen (2026) complement this with the Responsible and Ethical Principles (REP) framework, which emphasizes fairness, transparency, privacy, accountability, and respect for human values as the foundation for responsible AI integration in science education—principles that are relevant across disciplines.

DISCUSSION

The synthesis above shows strong convergence between recent findings (2023–2026) and the conceptual frameworks—entangled humanism, cyber humanism, and HASRL—that served as initial references for this study. The latest empirical evidence strengthens the argument that AI can indeed be a real force for educational equity. This is shown by the large effect size (1.40) found among disadvantaged students in Hungary (Katona & Katonáné Gyönyörű, 2026) and the consistent positive effects on K-12 mathematics achievement (Yi et al., 2024). However, equally strong evidence shows that these benefits are not automatic. They depend on proper pedagogical design, continuous human oversight, and policy frameworks that explicitly protect vulnerable groups from access gaps and cognitive dependency.

Shift in Discourse: From "Whether" to "Under What Conditions"

Interestingly, recent literature shows a shift in discourse, moving from simply asking "whether AI is beneficial" to asking "under what conditions AI is beneficial without compromising human agency." Concepts like cyber humanism and algorithmic citizenship (Adorni, 2025) signal an effort within the education discipline to build critical AI literacy as an integral part of education's goals. In this context, various ethical frameworks—including religious perspectives on collective moral responsibility for technology—remain relevant as one value compass among the various normative frameworks used in the mainstream academic literature reviewed here, without being the sole reference.

The four identified challenges—reduction of the teacher's role, access gaps, ethical and privacy issues, and cognitive dependency—are consistent with the findings of major reviews (Garzón et al., 2025, n=155; Zhu et al., 2025, n=75) included in this corpus. This shows that these challenges are not just speculative concerns but recurring patterns across geographical contexts and educational levels.

6. IMPLICATIONS AND RECOMMENDATIONS

Practical Implications

For educators, these findings imply a need for critical AI literacy. This literacy should include not only technical skills but also the capacity to question the assumptions, biases, and ethical implications of the AI systems they use. For institutions, there is a need for systematic investment in

equitable infrastructure, ongoing teacher training (including the development of digital leadership in school principals, as emphasized by Zeng et al., 2025), and human-in-the-loop oversight mechanisms to ensure that final pedagogical decisions remain with humans.

Policy Recommendations

1) Develop a regulatory framework that protects student data privacy and ensures algorithmic transparency in educational AI systems, in line with UNESCO guidelines (2023).
2) Allocate resources specifically to address access gaps—including infrastructure, cost, and digital literacy—prioritizing disadvantaged groups.
3) Revise teacher competency standards to include both digital-pedagogical capacity and ethical understanding of AI, following the four-role teacher framework (Zhai, 2025).
4) Establish institutional mechanisms to prevent excessive cognitive dependency. Examples include AI-free learning sessions, mandatory reflection points, and process-based authentic assessments (Zhai, Wibowo, & Li, 2024).
5) Encourage interdisciplinary research and cross-cultural replication, given that the majority of strong quantitative evidence in this review still comes from contexts in China, the United States, and Europe.

RESEARCH LIMITATIONS

This review has several methodological limitations that need to be openly acknowledged. First, the literature search was conducted using a single combined academic search engine rather than formal Boolean queries across separate databases such as Scopus or Web of Science. Therefore, this process is more accurately categorized as a rapid systematic review rather than a full PRISMA-style SLR across multiple databases. Second, the number of records at the identification stage (approximately 65) was relatively small compared to large-scale SLRs, which typically screen hundreds or thousands of records.

CONCLUSION

This systematic review of 25 recent articles with fully verified authorship confirms that integrating humanistic values into AI-based education is not just a normative discussion. It is supported by a growing body of empirical evidence. When properly designed and overseen, AI has been

shown to improve learning outcomes and social-emotional well-being, especially for disadvantaged students. However, these benefits are accompanied by real risks: reduced teacher role, access gaps, ethical and privacy issues, and excessive cognitive dependency. These four challenges consistently appear across studies. Conceptual frameworks like entangled humanism, cyber humanism, and HASRL offer a middle path: positioning AI as an equal epistemic partner that remains subject to human values, with the teacher playing the irreplaceable role of an ethical facilitator. Future research needs to expand the scope of search databases, strengthen the representation of Global South contexts, and test the long-term impact of intensive AI interaction longitudinally.

REFERENSI

Abubakar, S., Jeilani, A., & Adan, M. Y. (2025). The role of over-reliance on AI in the negative consequences of student learning: The moderating effects of ethical concerns and institutional policies. Cogent Education, 12(1), 2591503. https://doi.org/10.1080/2331186X.2025.2591503

Adorni, G. (2025). Cyber humanism in education: Reclaiming agency through AI and learning sciences. arXiv:2512.16701. https://doi.org/10.48550/arXiv.2512.16701

Adorni, G., & Bellini, E. (2025). Reclaiming agency through cyber humanism: A European agenda for AI, education and culture. In Proceedings of the International Conference on Cyber-Humanities. Springer Nature. https://link.springer.com/chapter/10.1007/978-3-032-11108-1_31

Garzón, J., Patiño, E., & Marulanda, C. (2025). Systematic review of artificial intelligence in education: Trends, benefits, and challenges. Multimodal Technologies and Interaction, 9(8), 84. https://doi.org/10.3390/mti9080084

Henriksen, D., Creely, E., Gruber, N., & Leahy, S. (2025). Social-emotional learning and generative AI: A critical literature review and framework for teacher education. Journal of Teacher Education. https://doi.org/10.1177/00224871251325058

Hou, X. (2025). The role of critical thinking on undergraduates' reliance behaviours on generative AI in problem-solving. British Journal of Educational Technology. https://doi.org/10.1111/bjet.13613

Järvelä, S., Nguyen, A., & Hadwin, A. (2023). Human-artificial intelligence collaboration for socially shared regulation in learning. British Journal of Educational Technology, 54(5), 1057–1076. https://doi.org/10.1111/bjet.13325

Jose, B., Cherian, J., Verghis, A. M., Varghese, S., & Joseph, S. (2025). The cognitive paradox of AI in education: Between enhancement and erosion. Frontiers in Education. https://pmc.ncbi.nlm.nih.gov/articles/PMC12036037/

Jose, B., Verghis, A. M., Varghese, S. M., Mumthas, S., & Cherian, J. (2025). From tools to co-learners: Entangled humanism and the co-evolution of intelligence in AI education. Frontiers in Education, 10, 1632990. https://doi.org/10.3389/feduc.2025.1632990

Kasneci, E., Seßler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., … Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. Learning and Individual Differences, 103, 102274. https://doi.org/10.1016/j.lindif.2023.102274

Katona, J., & Katonáné Gyönyörű, K. I. (2026). AI-based adaptive learning and programming achievement gains among socioeconomically disadvantaged university students. TechTrends.

Khazanchi, R. (2025). The effect of AI-based systems on mathematics achievement in a rural context: A quantitative study. Journal of Computer Assisted Learning. https://doi.org/10.1111/jcal.13098

Kosmyna, N., Hauptmann, E., Yuan, Y. T., Situ, J., Liao, X.-H., Beresnitzky, A. V., Braunstein, I., & Maes, P. (2025). Your brain on ChatGPT: Accumulation of cognitive debt when using an AI assistant for an essay writing task. arXiv:2506.08872. https://doi.org/10.48550/arXiv.2506.08872

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., … Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. BMJ, 372, n71. https://doi.org/10.1136/bmj.n71

Tian, J., & Zhang, R. (2025). Learners' AI dependence and critical thinking: The psychological mechanism of fatigue and the social buffering role of AI literacy. Acta Psychologica, 260, 105725. https://doi.org/10.1016/j.actpsy.2025.105725

UNESCO. (2023). Guidance for generative AI in education and research. UNESCO Publishing.

Yi, L., Liu, D., Jiang, T., & Xian, Y. (2024). The effectiveness of AI on K-12 students' mathematics learning: A systematic review and meta-analysis.

International Journal of Science and Mathematics Education, 23, 1105–1126. https://doi.org/10.1007/s10763-024-10499-7

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

Zeng, M., Cheah, K. S. L., & Abdullah, Z. (2025). The influence of school principals' digital leadership on teachers' competency in integrating artificial intelligence: A systematic thematic review. Frontiers in Education, 10, 1655967. https://doi.org/10.3389/feduc.2025.1655967

Zhai, C., Wibowo, S., & Li, L. D. (2024). The effects of over-reliance on AI dialogue systems on students' cognitive abilities: A systematic review. Smart Learning Environments, 11(1), 28. https://doi.org/10.1186/s40561-024-00316-7

Zhai, X. (2025). Transforming teachers' roles and agencies in the era of generative AI: Perceptions, acceptance, knowledge, and practices. Journal of Science Education and Technology. https://doi.org/10.1007/s10956-024-10174-0

Zhai, X., & Crippen, K. (2026). Charting the future of AI-supported science education: A human-centered vision. In X. Zhai & K. Crippen (Eds.), Advancing AI in science education: Envisioning responsible and ethical practice. Springer.

Zheng, Y., Wang, Y., Liu, K. S. X., & Jiang, M. Y. C. (2025). Artificial intelligence in higher education: Systematic review of student engagement and teaching methods. Frontiers in Education, 10, 1648661. https://doi.org/10.3389/feduc.2025.1648661

Zhu, H., Sun, Y., & Yang, J. (2025). Towards responsible artificial intelligence in education: A systematic review on identifying and mitigating ethical risks. Humanities and Social Sciences Communications, 12, 1111. https://doi.org/10.1057/s41599-025-05252-6

Zong, Y. (2025). How an AI-enhanced social-emotional learning framework transforms EFL students' engagement and emotional well-being. European Journal of Education. https://doi.org/10.1111/ejed.12925

Author's biography 1



Wilfridus Vinsentius Sarah, Pr., S.Fil., M.Pd., was born in Timor, East Nusa Tenggara. He completed his philosophical studies at the Faculty of Philosophy of Saint Michael, Widya Mandira University in Kupang-Timor, and theological studies at the School of Philosophy-Theology of Saint John Pematang Siantar, North Sumatra. Moreover, he earned his Master's in Educational Administration from Satya Wacana University in Salatiga, Central Java. He is an Associate Pastor at Saint Therese of Lisieux Parish in Sibolga City and a member of the board of the Budi Bakti Educational Foundation under the Diocese of Sibolga, North Sumatra.

Author's biography 2



Paulinus Kanisius Ndoa was born in Mbae-Keo Tengah, Nagekeo, Flores, East Nusa Tenggara, on June 22, 1979. In 2007, he completed his bachelor's degree in Philosophy and Theology at STFT St. Yohanes Pematangsiantar, affiliated with the University of St. Thomas Medan. In 2011, he was ordained as a Catholic priest and has since served the Catholic community of the Diocese of Sibolga. In 2016, he earned his Master's degree in Educational Administration from the Indonesian University of Education in Bandung. Since 2016, he has been serving as a lecturer at the Dian Mandala Pastoral College in Gunungsitoli, Nias, where he teaches several education courses, including Foundations of Education, Didactics and Methodology, Educational Media, and Learning Evaluation. In 2021, he began his doctoral studies in Educational Administration at the Indonesian University of Education in Bandung and completed them in January 2026.

CHAPTER XIV

Adaptive Educational Ecosystems in the Era of Artificial Intelligence: A Systematic Literature Review on the Synergy of Educational Institutions, Families, and Government Policies in Christian Higher Education

Lidiawati Lidiawati, Immanuela Deru
Kerusso Indonesia Theological Seminary
Coresponden Author : lp3m@sttkerussoindonesia.ac.id

ABSTRACT

This study aims to synthesize the existing body of literature on the integration of Artificial Intelligence in higher education and to propose an adaptive educational ecosystem model that connects educational institutions, families, and public policy within the context of Christian higher education. The rapid development of Artificial Intelligence has fundamentally transformed global education systems, including the emergence of adaptive learning, generative technologies, and large-scale data analytics that reshape pedagogical practice and institutional responsibility. Within Christian higher education, this transformation raises pedagogical, ethical, and theological concerns, particularly regarding the preservation of holistic human formation. A systematic literature review was conducted following the PRISMA 2020 guideline and the systematic review procedure of Kitchenham and Charters. Literature was retrieved from Scopus, Web of Science, ERIC, ScienceDirect, Google Scholar, and official grey literature published between 2015 and 2025. After identification, deduplication, title and abstract screening, and full text eligibility assessment, thirty sources met the inclusion criteria and were analyzed through thematic synthesis. The synthesis identifies four interrelated themes, namely the shifting authority of knowledge production and the redefinition of the educator role, the risk of pedagogical relation reduction associated with cognitive offloading, the position of the family as a digital ethics partner, and the role of public policy in ensuring equitable and ethical AI governance. These themes converge into an adaptive educational ecosystem model that places the student at the center of a dynamic relationship among institution, family, and government, oriented toward the formation of the whole person.

Keywords: Artificial Intelligence in Education; Systematic Liter ature Review; Christian Higher Education; Adaptive Learning Ecosystem; Digital Ethics

INTRODUCTION:

The rapid advancement of artificial intelligence (AI) has fundamentally transformed the global educational landscape. Artificial intelligence no longer functions merely as a supplementary learning tool; rather, it shapes a new ecosystem characterized by algorithm-based personalization, learning process automation, and the large-scale utilization of data analytics. This transformation impacts the ways in which learners acquire knowledge, educators design instructional strategies, and educational institutions formulate curricula and assessment frameworks (Zawacki-Richter et al., 2019). Education is now progressing toward an increasingly adaptive, responsive, and digitalized system.

At the global level, the emergence of generative AI is recognized as a milestone in the history of education, given that this technology enables the instantaneous generation of learning materials, automated feedback mechanisms, and learning simulations that closely mirror human interaction (Chan & Hu, 2023). However, this transformative potential does not come without risks. Beneath the surface of efficiency and ease of access lie growing concerns regarding the reduction of pedagogical relationships, overreliance on algorithms, and the erosion of the reflective dimension within the learning process (Selwyn, 2019). Numerous international bodies emphasize that the integration of artificial intelligence must be anchored within a clear ethical framework to ensure that human values and the dignity of learners are not compromised (UNESCO, 2023).

Concurrently, the boundaries of the learning space have undergone a fundamental shift. Education no longer takes place exclusively within formal classrooms; instead, it expands into domestic spaces and the global digital sphere. Students engage in learning through online platforms at home, in public venues, and across cross-border virtual networks. While this shift fosters a more flexible learning experience, it simultaneously poses challenges to quality assurance mechanisms, depth of understanding, and character formation. Consequently, education can no longer be conceptualized as the sole responsibility of formal institutions, as families and governments have now emerged as strategic actors who significantly influence the quality of the learning experience (Bond et al., 2022).

These global transformations resonate strongly within the Indonesian context. The digitalization of education presents substantial opportunities to

expand access to learning, particularly for students in remote regions. Nevertheless, social realities reveal persistent structural challenges. Disparities in internet access across regions remain stark, as do inequalities in the ownership of digital learning devices. Educators' readiness to integrate technology pedagogically remains uneven, while family digital literacy varies significantly (Badan Pusat Statistik, 2023). This situation creates a paradox: technology is readily available, yet the human capacity to use it meaningfully is not fully established.

The implementation of technology-based educational policies in Indonesia also encounters practical hurdles. The Kurikulum Merdeka (Independent Curriculum) highlights differentiated learning and the strategic use of technology to strengthen student competencies (Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi, 2022). In practice, however, many institutions remain preoccupied with the technical aspects of utilizing digital platforms, without a corresponding emphasis on critical and reflective pedagogy. Online learning is frequently reduced to the mere transfer of instructional material from physical classrooms to screens, leaving the dialogic dimension and character formation without proportionate attention.

This dynamic becomes increasingly complex when education is understood not merely as an academic exercise, but also as a domain for value formation. For Christian higher education institutions, the integration of artificial intelligence introduces an additional dimension that is inherently spiritual and theological. Christian education aims not only to transmit knowledge and practical skills, but also to shape character, spirituality, integrity, and a sense of vocation. Christian campuses are called to produce graduates who are not only intellectually competent, but also morally and spiritually mature. As learning systems become increasingly mediated by algorithms, there is a risk that the educational process may shift from personal relationships toward impersonal, machine-driven interactions—potentially weakening the values of community, empathy, and pastoral mentorship that have historically characterized Christian education (Bearman et al., 2023).

Regrettably, much of the discourse surrounding artificial intelligence in education remains fragmented to date. A significant portion of existing studies centers heavily on technological dimensions, others focus primarily on policy frameworks, while the family dynamic and spiritual values are frequently marginalized. This partial approach makes it difficult for

educational institutions—particularly Christian colleges and universities—to formulate a holistic adaptation strategy. This fragmentation underscores the urgent need for a systematic literature review capable of mapping, evaluating, and synthesizing previous research findings in a structured and methodologically rigorous manner, rather than merely compiling references narratively.

Based on the background articulated above, this study is formulated to address four key research questions: Research Question 1 (RQ1): How does artificial intelligence influence the authority of knowledge production and pedagogical relationships in higher education according to the existing literature? Research Question 2 (RQ2): How can the family be positioned as a digital ethics partner within an AI-driven learning ecosystem? Research Question 3 (RQ3): How do public policies, particularly within the Indonesian context, support or hinder the equitable and ethical integration of artificial intelligence in higher education?Research Question 4 (RQ4): How can an adaptive educational ecosystem model be formulated for Christian higher education institutions so that AI integration remains oriented toward whole-person formation?

Accordingly, this study aims to synthesize literature findings regarding the integration of artificial intelligence in higher education, identify core themes connecting educational institutions, families, and public policy, and articulate a conceptual model of an adaptive educational ecosystem relevant to Christian higher education in Indonesia.

Beyond addressing scholarly fragmentation, the urgency of this study is driven by the reality that the majority of literature on AI in higher education remains centered on developed countries with relatively mature digital infrastructure. In contrast, the challenges faced by Christian higher education institutions in Indonesia—particularly those located in semi-urban and rural areas—possess distinct characteristics, both in terms of infrastructural limitations and human resource readiness. Therefore, a critical synthesis of global literature must be grounded within the Indonesian context so that the resulting recommendations are genuinely actionable, rather than serving merely as normative adaptations of practices from other nations.

The novelty of this study lies in the application of a systematic literature review methodology following the PRISMA 2020 guidelines (Page et al., 2021) to synthesize three domains that have historically developed in

isolation: educational technology integration, family involvement in digital learning, and AI-based educational policy governance—subsequently bridging them with the theological perspectives of Christian education. Consequently, the primary contribution of this study is the development of an adaptive educational ecosystem model that is methodologically validated through transparent literature selection and synthesis procedures, while simultaneously enriching the discourse on Christian higher education with a humanistic and theological framework that has often been detached from educational technology studies.

RESEARCH METHODS

This study employs a systematic literature review (SLR) design adhering to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses 2020 (PRISMA 2020) guidelines (Page et al., 2021; Moher et al., 2009) as well as the systematic review procedures developed by Kitchenham and Charters (2007). The selection of this design is grounded in the necessity to map, evaluate, and synthesize evidence from diverse sources in a transparent, replicable, and methodologically sound manner, as highlighted by Snyder (2019) regarding literature reviews as a research methodology. The PRISMA 2020 framework was chosen because it provides an explicit reporting structure—ranging from record identification, screening, and eligibility assessment to the inclusion of studies in the final synthesis—thereby ensuring that the literature selection process does not rely solely on the subjective judgment of the researchers.

The literature search was conducted across five primary databases: Scopus, Web of Science, Education Resources Information Center (ERIC), ScienceDirect, and the academic search engine Google Scholar. In addition to these sources, this study also retrieved grey literature in the form of official reports from international organizations and governmental bodies, including reports from the Organisation for Economic Co-operation and Development (OECD, 2021), the United Nations Educational, Scientific and Cultural Organization (UNESCO, 2023), the World Bank (2020), the World Economic Forum (2023), as well as data and policy documents from Statistics Indonesia (Badan Pusat Statistik, 2023) and the Ministry of Education, Culture, Research, and Technology of the Republic of Indonesia (Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi, 2022).

The search strategy was constructed using a combination of keywords and Boolean operators as follows: ("artificial intelligence" OR "generative AI" OR "adaptive learning") AND ("higher education" OR "university") AND ("Christian education" OR "faith-based education" OR "family involvement" OR "digital ethics" OR "education policy"). For Indonesian-language literature, equivalent Indonesian keywords were used, including kecerdasan buatan (artificial intelligence), pendidikan tinggi (higher education), pendidikan Kristen (Christian education), literasi digital keluarga (family digital literacy), and kebijakan pendidikan (educational policy). The search was limited to publications released between 2015 and 2025 to ensure the relevance of the findings to the rapid advancements in generative AI technology over the past decade. Inclusion and exclusion criteria were established prior to the screening process to ensure consistency throughout the selection process.

Table 1: Summary of Inclusion and Exclusion Criteria

Aspect	Inclusion Criteria	Exclusion Criteria
Year of Publication	Published between 2015 and 2025	Published prior to 2015
Source Type	Reputable indexed journal articles, indexed proceedings, academic books, and official reports from international and governmental institutions	Popular opinion articles, blogs, and non-peer-reviewed sources lacking clear institutional authority
Thematic Focus	Discusses artificial intelligence in higher education, adaptive learning, digital ethics, family involvement, or educational policy	Does not address artificial intelligence or is irrelevant to the higher education context
Language	Written in English or Indonesian	Written in languages other than English or Indonesian
Accessibility	Full-text officially accessible through academic databases or institutional repositories	Only available as abstracts or lacking full-text access

The literature selection process was executed in four stages in accordance with the PRISMA 2020 framework: identification, title and abstract screening, full-text eligibility assessment, and final inclusion. In the identification stage, searching across the five databases and grey literature sources yielded 268 records. After removing duplicates using matching titles, authors, and publication years, 231 unique records remained.

The title and abstract screening subsequently excluded 152 records that were irrelevant to the research questions, leaving 79 records for full-text eligibility assessment. At this stage, 49 records were excluded due to an unaligned thematic focus, inaccessible full texts, or methodological quality that failed to meet the established quality assessment standards. This selection process resulted in 30 studies that met all criteria and were included in the final synthesis. A summary of the selection stages is presented in the following table.

Table 2: Summary of Literature Selection Stages

Stage	Number of Records	Number Excluded
Initial Identification from Databases (Scopus: 96, Web of Science: 54, ERIC: 41, ScienceDirect: 37, Google Scholar and Grey Literature: 40)	268	—
Duplicate Removal	231	37
Title and Abstract Screening	79	152
Full-Text Eligibility Assessment	30	49
Studies Included in Final Synthesis	30	—

A quality assessment of the 79 studies that passed the title and abstract screening was performed using a checklist instrument adapted from the quality assessment principles of Kitchenham and Charters (2007), tailored to accommodate conceptual literature and policy documents. The instrument evaluated five aspects: clarity of research objectives, appropriateness of the approach or method relative to the stated objectives, clarity and consistency of findings, contextual relevance to higher education or value-based education, and source credibility demonstrated through peer-review status or institutional publishing authority. Each aspect was scored on a three-point scale: *fulfilled*, *partially fulfilled*, and *unfulfilled*. Studies scoring fulfilled or partially fulfilled across all aspects were deemed eligible for full-text

assessment, whereas studies scoring unfulfilled in two or more aspects were excluded from further evaluation.

Data extraction was carried out manually by two researchers using a standardized extraction form that recorded author names, publication years, publication types, regional study foci, research objectives, key findings, and relevance to the four formulated research questions. Each included study was subsequently analyzed using a thematic synthesis approach, which proceeded in three steps.

The first step was descriptive coding, which involved identifying key concepts and findings from each study. The second step entailed grouping these codes into broader categories based on shared conceptual meanings. The third step was the formulation of analytical themes, synthesizing these categories into main themes that directly addressed the research questions. The coding process was conducted independently by both researchers and subsequently discussed collectively to reach consensus, thereby ensuring the consistency and credibility of the interpretations. Through this process, four primary themes were established: the shift in knowledge authority, the risk of reducing pedagogical relationships, the family as a partner in digital ethics, and public policy within the Indonesian context, which were ultimately synthesized into an adaptive educational ecosystem model

RESULTS

This section presents the results of the literature selection process, along with the characteristics of the included studies, prior to an analytical discussion of these findings. The presentation of results proceeds sequentially, beginning with a general overview of the study characteristics and a summary of the thematic synthesis, which serves as the foundation for the subsequent discussion.

Characteristics of Included Studies

A total of 30 studies met the inclusion criteria and were analyzed in the final synthesis. These studies comprise articles from reputable, indexed journals, official reports from international organizations, and national policy documents and data. The majority of the studies were published between 2020 and 2025, reflecting a sharp increase in academic attention to the integration of artificial intelligence in education since the widespread emergence of generative AI. In terms of geographical scope, most studies address global or cross-national contexts. In contrast, others specifically examine the

Indonesian context, including data on internet penetration, the Kurikulum Merdeka policy, and digital literacy disparities. The complete characteristics of the included studies are presented in the table below.

Table 3: Summary of Included Study Characteristics

Characteristic	Category	Number of Studies
Publication Type	Indexed journal articles	21
	Reports from international organizations	6
	National policy documents and data	3
Publication Period	2015 to 2019	9
	2020 to 2025	21
Geographical Scope	Global or cross-national studies	18
	Indonesian context studies	12
Dominant Theme	Knowledge, authority and the role of educators	9
	Pedagogical relationships and affective dimensions	7
	Family and digital literacy	6
	Public policy and equitable access	8

To provide transparency regarding the data extraction process, the following table summarizes the source, context, primary focus, and theme of each study included in the final synthesis.

Table 4: Extraction Summary of Included Studies

Source	Context	Primary Focus	Theme
Zawacki-Richter et al. (2019)	Global	Systematic review of AI applications in higher education	Knowledge authority
Holmes et al. (2019)	Global	Role of AI and the redefinition of educator functions	Knowledge authority
Chan & Hu (2023)	Global	Student perceptions of generative AI	Knowledge authority
Popenici & Kerr (2017)	Global	Impact of AI on teaching and learning	Knowledge authority
Luckin (2017)	United Kingdom	AI-based assessment systems	Knowledge authority
Luckin et al. (2018)	United Kingdom	Arguments for AI integration in education	Knowledge authority

Source	Context	Primary Focus	Theme
Woolf (2019)	United States	Interaction-based intelligent tutoring systems	Knowledge authority
Holstein et al. (2019)	United States	Intelligent tutoring systems and human roles	Knowledge authority
Ouyang & Jiao (2023)	Global	Three paradigms of AI in education	Knowledge authority
Selwyn (2019)	United Kingdom	Critique of educational automation	Pedagogical relationships
Bearman et al. (2023)	Australia	Discourse on AI in higher education	Pedagogical relationships
Williamson (2021)	United Kingdom	Governance of AI in education	Pedagogical relationships
Barr et al. (2015)	Canada	The phenomenon of cognitive offloading	Pedagogical relationships
Fischer & Kiili (2020)	United States	Sociotechnical environments for collaborative inquiry	Pedagogical relationships
Brown & Green (2018)	United States	Challenges in adaptive learning systems	Pedagogical relationships
Kizilcec & Hoadley (2021)	United States	Impact of AI on student engagement	Pedagogical relationships
Boulton & Hramiak (2020)	United Kingdom	Teachers' professional learning for AI	Pedagogical relationships
OECD (2021)	Cross-national	Digital literacy and family engagement	Family
World Bank (2020)	Global	Educational reform and social actor integration	Family
Kementerian Pendidikan, Kebudayaan, Riset, dan	Indonesia	*Kurikulum Merdeka* policy framework	Family

Source	Context	Primary Focus	Theme
Teknologi (2022)			
Woolf (2019)	United States	Intelligent tutor interaction and mentoring support	Family
Badan Pusat Statistik (2023)	Indonesia	Statistics on internet penetration and digital literacy	Public policy
UNESCO (2023)	Global	Ethical guidelines for generative AI in education	Public policy
World Economic Forum (2023)	Global	Policy collaboration for digital transformation	Public policy
Li & Lalani (2020)	Global	Pandemic impacts on educational transformation	Public policy
Crawford et al. (2023).	Australia	Ethical leadership in ChatGPT utilization	Public policy
Bond et al. (2022).	Global	Systematic review of higher education digital transformation	Public policy
Page et al. (2021)	Global	PRISMA 2020 methodological guidelines	Methodology
Kitchenham & Charters (2007)	United Kingdom	Systematic literature review guidelines	Methodology
Snyder (2019)	Sweden	Literature review as a research methodology	Methodology

THEMATIC SYNTHESIS

The thematic synthesis of the 30 included studies yielded four interrelated primary themes. The first theme is the shift in knowledge production authority and the redefinition of the educator's role, which emerged consistently across 9 studies and affirms that the presence of artificial intelligence alters relational patterns among educators, students, and knowledge sources. The second theme is the risk of reducing pedagogical relationships, addressed in 7 studies, which closely relates to the phenomenon

of cognitive offloading and the diminishing depth of affective interactions within technology-mediated learning.

The third theme is the family as a digital ethics partner, highlighted in 6 studies, emphasizing the family's role in shaping learning habits and integrity in technology use. The fourth theme, public policy and equitable access, is discussed across 8 studies and stresses the critical importance of regulation, infrastructure, and data governance in supporting equitable AI integration. These four themes are elaborated in detail in the discussion section and synthesized into an adaptive educational ecosystem model that places the student at the center of a dynamic relationship between educational institutions, families, and the government.

DISCUSSION

The transformation of education driven by artificial intelligence cannot be understood merely as a technical alteration; rather, it represents a structural shift in pedagogical relationships, knowledge production, and the orientation of human formation. Within Christian higher education, this shift influences not only how students learn but also how institutions interpret their educational mission. Christian education, which has historically been rooted in the formation of faith, character, and integrity, now confronts technologies that accelerate cognitive processes while simultaneously threatening to diminish the depth of reflection. This section elaborates on the synthesized findings across the four identified primary themes, subsequently integrating them into a conceptual model and examining its strategic implications for Christian higher education institutions.

The Shift in Knowledge Authority and the Redefinition of the Educator's Role

One of the most significant shifts identified in the literature synthesis is the reallocation of authority in knowledge production. Whereas faculty members were previously viewed as the primary repositories of academic information, algorithms now participate in delivering instructional content. Zawacki-Richter et al. (2019) demonstrated that the application of artificial intelligence in higher education has altered instructional interaction patterns from a transmission model of knowledge to a data-driven exploration model. This finding is reinforced by Chan and Hu (2023), who observed that students

increasingly view generative AI as a partner in knowledge exploration rather than merely a tool for administrative assistance.

This reallocation of knowledge authority does not automatically eliminate the educator's role. On the contrary, the literature synthesis indicates that this shift demands a redefinition of the lecturer's role as a facilitator of reflection, an ethical guide, and a mentor of academic integrity (Holmes et al., 2019). Faculty members no longer focus solely on information delivery, but rather on mentoring thought processes, refining argumentation, and fostering student character development. A similar perspective was articulated by Popenici and Kerr (2017), who asserted that the successful integration of artificial intelligence in higher education depends heavily on educators' pedagogical readiness rather than solely on the sophistication of the systems deployed.

Within the context of Christian education, this shift carries vital theological implications. Theological and humanistic learning requires a contemplative process involving reflection, dialogue, and the integration of faith values. When artificial intelligence provides instantaneous answers, there is a risk that students may bypass the reflective process essential for wisdom formation. Barr, Pennycook, Stolz, and Fugelsang (2015) termed this phenomenon *cognitive offloading*—the tendency to surrender cognitive processing to digital systems. Consequently, the literature synthesis affirms that Christian campuses must position artificial intelligence as a supportive tool for intellectual exploration rather than a substitute for contemplative processing, thereby guiding students to utilize such technologies to broaden their perspectives while retaining intellectual responsibility in interpreting and evaluating information.

The Risk of Reducing Pedagogical Relationships and the Phenomenon of Cognitive Offloading

The second theme identified from the literature synthesis is the risk of reducing pedagogical relationships. Digital platforms enable learning interactions without physical presence; however, personal relationships between faculty and students often become superficial. Selwyn (2019) contended that technology-driven education tends to emphasize efficiency, while affective and relational dimensions receive insufficient attention. This finding aligns with the critical study by Bearman, Ryan, and Ajjawi (2023),

which demonstrated that the discourse surrounding artificial intelligence in higher education remains dominated by narratives of efficiency and optimization. In contrast, questions concerning the intrinsic meaning of pedagogical relationships have received limited academic focus.

In Christian education, personal relationships constitute the core of character formation. Informal conversations, pastoral mentoring, and the lived example of faculty play essential roles in instilling faith values. When screens predominantly mediate interactions, the depth of empathy and reflective dialogue can be attenuated. To address this challenge, the literature synthesis recommends that Christian higher education institutions develop a reflective blended learning approach—one that thoughtfully integrates digital learning with meaningful face-to-face encounters, ensuring that technology enhances rather than replaces educational relationships. Williamson (2021) emphasized that the governance of artificial intelligence in education must maintain a balance between automation and human relationality so that learning is not reduced to performance optimization. Accordingly, policies governing AI integration on Christian campuses must be intentionally designed to safeguard spaces for dialogue, academic mentorship, and spiritual formation, positioning technology as a catalyst for relational support rather than its replacement.

The Family as a Digital Ethics Partner

The third theme emerging from the literature synthesis is the positioning of the family as a digital ethics partner within the learning ecosystem. In the Indonesian context, family involvement in higher education is frequently assumed to be minimal. However, the digital era challenges this assumption, as students remain embedded within domestic social environments that shape their attitudes toward academic integrity and responsibility. The OECD (2021) noted that family support significantly contributes to the success of digital learning and the cultivation of healthy study habits.

The family serves as a foundational locus for the formation of values regarding technology use. When families instill honesty and discipline, students are more inclined to utilize artificial intelligence ethically. Conversely, without value reinforcement, students may adopt a pragmatic approach to technology, such as fully outsourcing assignments to AI systems

without personal reflection. In Christian education, the family holds a vital spiritual function as the primary faith community. Therefore, the literature synthesis recommends that Christian campuses establish regular communication channels with families through digital literacy seminars, ethical guidelines for AI usage, and faith-based academic integrity forums. Such collaboration extends the educational space from the campus into the domestic domain, ensuring consistency between formal education and daily life. This ecosystem approach aligns with the World Bank's (2020) perspective, which noted that the success of educational reforms depends on the cohesion of all social actors rather than on formal institutions alone.

Public Policy and the Indonesian Context

The fourth identified theme concerns public policy in the Indonesian national context. The digital transformation of education in Indonesia is driven by various national initiatives, including the implementation of the *Kurikulum Merdeka*, which emphasizes differentiated learning and technological integration (Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi, 2022). However, macro-level policies must be translated into context-specific institutional policies to avoid remaining merely normative statements.

Data from Statistics Indonesia (Badan Pusat Statistik, 2023) indicate that despite increases in national internet penetration, the quality of access and digital literacy levels remain uneven across regions. Christian higher education institutions located in semi-urban or rural areas face distinct infrastructural limitations and faculty capacity constraints compared to those in major urban centers. This reality underscores that AI implementation cannot be standardized; instead, it must be adapted to each institution's readiness. UNESCO (2023) emphasized that integrating generative AI into education must be accompanied by data protection regulations, equitable access, and clear ethical frameworks. Crawford, Cowling, and Allen (2023) noted that institutional leadership plays a crucial role in ensuring the ethical use of artificial intelligence, particularly in assessment and academic integrity.

Within the Indonesian context, the government holds a strategic responsibility to ensure student data security and to provide pedagogical—rather than purely technical—training for educators utilizing artificial intelligence. The World Economic Forum (2023) similarly highlighted the

importance of collaboration between educational institutions and policymakers to forge an equitable digital learning ecosystem. Christian campuses can contribute by offering ethical and humanistic perspectives to national AI policy formulation, ensuring that technological transformation remains centered on human dignity.

Furthermore, synthesis reveals that the gap between macro policy and micro classroom practice is not solely due to budgetary constraints, but also to weak mechanisms for translating policies into operational guidelines at the academic department level. National policies such as *Kurikulum Merdeka* provide general direction on differentiated learning. However, they do not always provide technical guidance on managing artificial intelligence in assessments, assignments, or thesis supervision within Christian higher education environments.

This vacuum generates substantial variance in institutional practices, leading some campuses to enforce total bans on AI usage while others permit its unguided deployment without adequate ethical guardrails. The literature synthesis indicates that both extremes carry inherent risks, ranging from the stagnation of students' digital competencies to the uncritical misuse of technology. Consequently, a measured middle ground is required—one in which AI use is transparently permitted within explicit ethical guidelines and boundaries.

The Synthesized Adaptive Educational Ecosystem Model

Based on these four themes, this study articulates a conceptual model of an adaptive educational ecosystem that centers on a dynamic, reciprocal interaction among three primary pillars: educational institutions, families, and the government. Educational institutions function to guide pedagogy, curriculum design, and the ethical use of artificial intelligence. Families serve to instill foundational values, discipline, and daily technological oversight. The government enacts regulations, infrastructure, and policies to ensure equitable access and data privacy. These three pillars interact continuously, converging toward a single ultimate orientation: the whole-person formation of the student encompassing intellectual, moral, spiritual, and social dimensions.

This model asserts three fundamental principles. First, the student occupies the center of a dynamic relational network rather than acting as a

passive recipient of technology. Second, the three pillars of the ecosystem are interdependent, meaning that weakness in one pillar adversely affects the entire system. When an institution acts without family support, learning risks losing the reinforcement it receives within the home. When families lack pedagogical guidance from the university, domestic educational support becomes unstructured. Similarly, without public policies favoring equitable access and data protection, AI integration risks widening existing educational disparities. Third, the primary objective of the ecosystem is not technological efficiency, but whole-person formation, positioning artificial intelligence as a supportive instrument rather than a substitute for personal relationships and character mentorship.

This conceptual model can also be operationalized through four core characteristics that distinguish it from being consistently evaluated based on its impact on the quality of relationships among students, faculty, families, and the broader community. The second characteristic is a balance between autonomy and guidance, in which students are given space to use artificial intelligence independently while remaining anchored in clear academic and spiritual mentorship structures. The third characteristic is tiered accountability, wherein ethical responsibility is not borne by a single actor but is distributed proportionally among the institution, the family, and the government according to their respective roles. The fourth characteristic is a long-term orientation toward character formation, signifying that the success of AI integration is measured not merely by short-term academic efficiency, but by its long-term contribution to the intellectual, moral, and spiritual maturity of the student.

The literature synthesis further demonstrates that this three-pillar `ecosystem model aligns with the findings of Fischer and Kiili (2020) regarding the necessity of sociotechnical environments that foster collaborative inquiry, as well as the perspective of Holstein, McLaren, and Aleven (2019) that intelligent tutoring systems achieve optimal effectiveness when designed to augment, rather than replace, human roles in the learning process. Thus, the adaptive educational ecosystem model proposed in this study serves not merely as a normative framework but also as a synthesis of empirical and conceptual evidence that consistently shows that successful AI integration in higher education relies on the cohesion of the human actors surrounding the technology.

Strategic Implications for Christian Higher Education

The literature synthesis yields five strategic implications for Christian higher education institutions seeking to manage the integration of artificial intelligence responsibly:

First, Christian campuses must systematically integrate AI literacy into the curriculum as a core twenty-first-century competency. This literacy encompasses not only the technical skills required to operate AI tools, but also the capacity to evaluate information quality, recognize technological limits, and cultivate critical reflection regarding algorithmically generated outputs (OECD, 2021; UNESCO, 2023). Given that the Merdeka curriculum emphasizes differentiated learning and the use of technology to support students' potential, embedding AI literacy aligns directly with national policy frameworks.

Second, reinforcing faith-based academic integrity must become an ongoing institutional priority. The phenomenon of cognitive offloading illustrates that unmonitored technology usage can erode reflective thinking and intellectual accountability (Barr et al., 2015). Consequently, Christian higher education institutions should update their academic codes of conduct to include explicit guidelines on AI use in coursework and research, while grounding academic practice in Christian values such as honesty, stewardship, and neighborly love.

Third, enhancing faculty capacity through training in reflective pedagogy represents a vital prerequisite within an adaptive educational ecosystem. Holmes et al. (2019) emphasized that the educator's role in the AI era shifts from information provider to learning facilitator. Faculty development should focus on authentic assessment design, project-based learning, and reflective discussion to ensure that artificial intelligence functions as an aid to intellectual exploration rather than a substitute for critical thinking.

Fourth, Christian campuses should foster closer partnerships with families as integral components of the educational ecosystem. The OECD (2021) demonstrated that family engagement contributes significantly to successful digital learning. Through family communication forums, digital literacy workshops, and ethical guidelines for AI usage, institutions can assist parents in navigating digital education challenges while reinforcing value formation in the home.

Fifth, Christian higher education institutions have a moral responsibility to contribute to public policy discourse on artificial intelligence in education. The World Economic Forum (2023) and UNESCO (2023) underscored the necessity of institutional participation in policy formation to ensure that digital transformations remain aligned with social justice and human dignity. Through research, academic forums, and strategic partnerships, Christian campuses can offer ethical and theological perspectives that enrich national policy directions.

Taken together, these five strategic implications demonstrate that adapting to artificial intelligence requires a holistic framework uniting pedagogical innovation, faith-value reinforcement, faculty development, family engagement, and public policy participation. This approach enables Christian campuses to leverage artificial intelligence as an instrument for educational transformation that is both technologically adept and faithful to the calling of whole-person formation.

Limitations of the Study

This study presents several limitations that should be considered when interpreting its findings. First, this systematic literature review is conceptual and has not been empirically tested in a specific Christian campus context; thus, the resulting ecosystem model requires field validation through subsequent quantitative or qualitative research. Second, although the search encompassed five major databases and official grey literature, relevant studies published in languages other than English or Indonesian, or those not indexed within the searched databases, may have been omitted. Third, the majority of included studies originated in global or developed-country contexts, warranting caution when generalizing the findings to Christian higher education institutions in semi-urban or rural areas of Indonesia. Fourth, despite being conducted independently by two researchers and resolved through consensus, the quality assessment and thematic coding processes inherently involve interpretive elements that may reflect the researchers' theological and pedagogical perspectives. These limitations present valuable opportunities for future empirical studies to test and refine the proposed adaptive educational ecosystem model.

In addition to these methodological limitations, it should be recognized that the thematic synthesis approach employed here is interpretive.

Consequently, thematic groupings might vary when conducted by research teams from different disciplinary backgrounds, such as those oriented solely to computer science or public policy. Therefore, the adaptive educational ecosystem model developed in this study should be understood as an open, preliminary framework subject to further testing and refinement, rather than a final, universally applicable model across all Christian higher education contexts.

CONCLUSION AND RECOMMENDATIONS

The emergence of artificial intelligence has transformed education not merely at a technological level, but also within the structural dynamics of pedagogical relationships and the intrinsic meaning of learning itself. Through a systematic literature review of the 30 included studies, this study demonstrates that artificial intelligence presents profound opportunities through learning personalization, administrative efficiency, and expanded access to knowledge resources. Concurrently, however, this technology introduces risks of diminishing educational relationships, overreliance on algorithms, and the erosion of reflective processes if it is not integrated judiciously. Within the context of Christian higher education, this transformation becomes even more significant, given that education is aimed not solely at mastering academic competencies but also at forming character, spirituality, and integrity.

The synthesized findings affirm that the integration of artificial intelligence cannot be conceptualized as an isolated technical project; rather, it must be embedded within an adaptive educational ecosystem framework involving the synergy of three primary pillars: educational institutions, families, and the government. Christian educational institutions serve as pedagogical leaders, designing curricula grounded in AI literacy, establishing reflective assessments, and embedding faith values into the learning process. The family serves as a strategic partner that cultivates study habits, guides students' digital practices, and reinforces integrity in daily life. The government acts as a regulator and facilitator, providing policy frameworks, infrastructure, and systemic support to ensure the use of artificial intelligence remains equitable and ethical.

The dynamic interaction among these three actors shapes the trajectory of Christian education in the digital era. The success of an adaptive

educational ecosystem depends on the quality of cross-sectoral collaboration that centers on human dignity, with technology serving as a supporting instrument. Furthermore, this study emphasizes that the success of Christian education in the age of artificial intelligence is determined not by the sophistication of digital systems but by the reflective maturity with which it integrates technology with faith values and moral responsibility.

Based on these findings, this study offers three key recommendations for action. First, Christian higher education institutions are advised to formulate internal policies regarding the use of artificial intelligence that explicitly integrate technical literacy, academic ethics, and faith values into both curricula and codes of conduct. Second, further empirical research grounded in field data across various Christian campus contexts in Indonesia is needed to validate the proposed adaptive educational ecosystem model and identify concrete challenges encountered by institutions, families, and policymakers. Third, Christian campuses are encouraged to expand cross-sectoral partnerships with families and the government through regular dialogue forums, ensuring that digital transformations in education remain steadfastly oriented toward whole-person formation—producing individuals who are intellectually adept, emotionally mature, socially conscious, and spiritually grounded.

REFERENCES

Badan Pusat Statistik. (2023). Statistik telekomunikasi Indonesia 2023. BPS.

Barr, N., Pennycook, G., Stolz, J. A., & Fugelsang, J. A. (2015). The brain in your pocket: Evidence that smartphones are used to supplant thinking. Computers in Human Behavior, 48, 473- 480.

Bearman, M., Ryan, J., & Ajjawi, R. (2023). Discourses of artificial intelligence in higher education: A critical literature review. Higher Education, 86(2), 369- 385.

Bond, M., Marín, V. I., Dolch, C., Bedenlier, S., & Zawacki Richter, O. (2022). Digital transformation in higher education: A systematic review. Educational Technology & Society, 25(2), 1 16.

Boulton, H., & Hramiak, A. (2020). Teacher and professional learning for sustainable AI in education. Journal of Learning for Development, 7(1), 66- 78.

Brown, M., & Green, T. (2018). Issues and challenges in artificial intelligence-based adaptive learning systems. Educational Technology Research and Development, 66(5), 1123- 1142.

Chan, C. K. Y., & Hu, W. (2023). Students' voices on generative AI: Perceptions, benefits, and challenges in higher education. International Journal of Educational Technology in Higher Education, 20(1), 43.

Crawford, J., Cowling, M., & Allen, K. A. (2023). Leadership is needed for ethical ChatGPT: Character, assessment, and learning using artificial intelligence—Journal of University Teaching and Learning Practice, 20(3).

Fischer, G., & Kiili, C. (2020). Socio-technical environments supporting inquiry learning with AI tools. IEEE Transactions on Learning Technologies, 13(4), 725- 738.

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

Holstein, K., McLaren, B. M., & Aleven, V. (2019). Intelligent tutoring systems. AI Magazine, 40(3), 32- 45.

Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi. (2022). Kurikulum Merdeka dan implementasinya dalam pendidikan tinggi. Kemendikbudristek.

Kitchenham, B., & Charters, S. (2007). Guidelines for performing systematic literature reviews in software engineering (EBSE 2007 001): Keele University and University of Durham.

Kizilcec, R. F., & Hoadley, C. M. (2021). The impact of AI-assisted learning on student engagement and outcomes. Journal of Educational Computing Research, 59(3), 432- 456.

Li, C., & Lalani, F. (2020). The COVID-19 pandemic has changed education forever: Here is how. World Economic Forum.

Luckin, R. (2017). Towards artificial intelligence-based assessment systems. UCL Knowledge Lab.

Luckin, R., Holmes, W., Griffiths, M., & Forcier, L. B. (2018). Intelligence unleashed: An argument for AI in education. Pearson.

Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. PLoS Medicine, 6(7), e1000097.

OECD. (2021). 21st-century readers: Developing literacy skills in a digital world. OECD Publishing.

Ouyang, F., & Jiao, P. (2023). Artificial intelligence in education: The three paradigms. Computers and Education: Artificial Intelligence, 4, 100123.

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo Wilson, E., McDonald, S., ... Moher, D. (2021). The

PRISMA 2020 statement: An updated guideline for reporting systematic reviews. BMJ, 372, n71.

Popenici, S. A. D., & Kerr, S. (2017). Exploring the impact of artificial intelligence on teaching and learning in higher education. Research and Practice in Technology Enhanced Learning, 12(1), 22.

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

Snyder, H. (2019). Literature review as a research methodology: An overview and guidelines. Journal of Business Research, 104, 333- 339.

Tricco, A. C., Lillie, E., Zarin, W., O'Brien, K. K., Colquhoun, H., Levac, D., Moher, D., Peters, M. D. J., Horsley, T., Weeks, L., Hempel, S., Akl, E. A., Chang, C., McGowan, J., Stewart, L., Hartling, L., Aldcroft, A., Wilson, M. G., Garritty, C., ... Straus, S. E. (2018). PRISMA extension for scoping reviews (PRISMA ScR): Checklist and explanation—Annals of Internal Medicine, 169(7), 467- 473.

UNESCO. (2023). Guidance for generative AI in education and research. UNESCO Publishing.

Williamson, B. (2021). Governing artificial intelligence in education: Policy networks, governance and platformization. Educational Technology Research and Development, 69, 1 24.

World Bank. (2020). Realizing the future of learning: From learning poverty to learning for everyone. World Bank.

World Economic Forum. (2023). Shaping the future of learning: The role of AI and digital transformation. WEF.

Woolf, B. P. (2019). Building intelligent interactive tutors: Student-centered strategies for revolutionizing e-learning. Morgan Kaufmann.

Zawacki Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education. International Journal of Educational Technology in Higher Education, 16(39).

AUTHOR PROFILE



Lidiawati, M.Psi., M.Pd. I am an author, academic, and psychology practitioner devoted to exploring the dynamics of the human psyche and the shaping of character. As a lecturer at STT Kerusso Indonesia, I draw on my dual expertise in psychology and educational science to craft insights that speak to both mind and heart. In my writing, I bridge the psychological realities of modern life with profound theological truths, translating complex subjects—mental health, parenting, spiritual and character formation, and

counseling—into warm, reflective, and accessible narratives. My academic grounding never makes my prose rigid; instead, it gives my arguments vitality, rooted in real experience and genuine care for others. For me, writing is more than sharing knowledge—it is a space of encounter, meant to nurture the soul, inspire growth, and guide readers toward deeper self-awareness, resilience, and a meaningful impact on the world around them.



Dr. Immanuela Deru. I am an academic, researcher, and senior lecturer who has devoted my career to STT Kerusso Indonesia. Holding a doctorate in education and a master's degree in theology, I bring a deep understanding of how academic theological principles intersect with the practical work of Christian Religious Education. In my teaching, I cover instructional methodology, Christian education curricula, and spiritual formation, while my research centers on contextual discipleship strategies, youth character mentorship, and Christian leadership in the digital era. I regularly speak at theological seminars and training sessions for Christian educators, offering scholarly insight to strengthen the roles of church and school in nurturing faith within the community. Through my holistic approach to teaching, I strive to inspire students preparing for ministry, equipping them to serve with both wisdom and conviction.

CHAPTER XV

Teachers as Moral Guardians of Artificial Intelligence: A Systematic Literature Review on Ethics, Agency, and Pedagogical Responsibility

Benediktus Bani and Karolina Lero
SMAK Santa Maria Monte Carmelo Maumere

Corresponding author: beniedetto78@gmail.com

Abstract

Generative artificial intelligence is now widely present in classrooms around the world. Its presence raises urgent ethical questions for teachers. This article presents a systematic literature review on teachers as moral gatekeepers of artificial intelligence. The review focuses on three dimensions, namely ethics, teacher agency, and pedagogical responsibility. The PRISMA protocol was used to screen 114 Scopus-indexed articles published between 2023 and 2026, yielding 43 relevant articles for review. The findings reveal five major themes, namely ethical mediation by teachers, teachers' AI literacy, school leadership, student agency, and academic integrity. The thematic discussion presents a novel contribution in the form of a framework positioning teachers as moral gatekeepers. This framework is contextualised within religio-cultural values, using the case of a Catholic secondary school. This novelty complements the global literature, which remains largely centred on technology acceptance theory. The study recommends strengthening teacher training in AI ethics **based on local values.**

Keywords: artificial intelligence; educational ethics; teacher agency; ethical mediation; systematic literature review

1. Introduction

Generative artificial intelligence has rapidly spread into classrooms around the world. UNESCO has promoted an AI competency framework for teachers and students (Hsu & Lao, 2025). Tools such as ChatGPT are now widely used by students (Granström & Oppi, 2025). A national survey in Estonia found that most students had already used AI tools (Granström & Oppi, 2025). This phenomenon has rapidly and fundamentally transformed the learning landscape.

Behind this enthusiasm, the adoption of artificial intelligence is accompanied by ethical challenges. A systematic review found that

educational ethics is often overlooked (Mouta et al., 2024). Issues of cultural diversity and the emotional dimensions of students remain insufficiently discussed (Mouta et al., 2024). Ethical frameworks for artificial intelligence in secondary education also continue to have significant gaps (Ofosu-Asare, 2026). Privacy, algorithmic bias, and accountability have become the main concerns (Ofosu-Asare, 2026).

Teachers play a key role in these ethical issues, expected to guide students in the responsible use of artificial intelligence (Pardede et al., 2026). However, AI also challenges teacher agency, as a study in Azerbaijan found that AI can diminish teachers' instructional authority (Alasgarova & Rzayev, 2025). Teachers' ethical preparedness hinges on AI literacy and professional development (Alasgarova & Rzayev, 2025), yet many countries still have limited teacher knowledge of AI content (Kim & Kwon, 2023).

Institutional settings influence teachers' ethical abilities. Leadership with a digital mindset promotes effective AI integration (Pietsch & Mah, 2025). Transformative leadership alongside the TPACK framework enhances ethical AI governance (Alshamsi, 2025). Many school principals worldwide worry that AI may compromise academic integrity (Eryilmaz et al., 2025). However, the specific values rooted in institutional beliefs are seldom directly analyzed. Most current research primarily focus on technology acceptance theories that are value-neutral.

The research gap is clearly evident. While the global literature mainly measures adoption intentions, it seldom explores the moral aspects of AI. Teachers' roles as moral gatekeepers are rarely examined comprehensively. Religio-cultural settings, like faith-based schools, are underrepresented in systematic studies. However, religious values could enhance ethical frameworks for AI. This gap highlights the necessity of this systematic literature review.

This article presents a systematic literature review to address this gap. The review focuses on teachers' ethics, agency, and pedagogical responsibility. The review also presents a thematic discussion that offers a novel contribution from the Indonesian context. Three research questions were formulated to guide this systematic literature review.

Based on the above discussion, this study formulates the following three research questions.

RQ1: How has research on teachers as moral gatekeepers of artificial intelligence evolved between 2023 and 2026?

RQ2: What themes shape teachers' ethics, agency, and pedagogical responsibility?

RQ3: How do local religious and cultural contexts enrich universal ethical frameworks for artificial intelligence?

2. Literature Review

This section offers a thematic overview of prior research, focusing on four key sub-themes aligned with this study's conceptual framework: teacher mediation, ethical filtering, TPACK and AI literacy, and school leadership.

Teacher Mediation from the Perspective of Cultural-Historical Activity Theory

Teacher mediation is based on Cultural-Historical Activity Theory, where teachers act as mediators between students and cultural artefacts, such as technology. The Zone of Proximal Development describes how teachers guide students to reach their potential. Applying this framework to AI highlights that teachers' active involvement continues to be vital (Engeness, 2025). AI feedback works best when integrated with social interaction and teacher guidance (Engeness, 2025).

This mediation principle applies to Indonesia's secondary schools, where teachers assist students in verifying ChatGPT-generated content (Pardede et al., 2026) and encourage them to rephrase information in their own words (Pardede et al., 2026). This approach exemplifies mediation as a cognitive bridge, extending beyond simple supervision. Consequently, teacher mediation influences students' critical thinking regarding artificial intelligence.

Educational Ethics and Ethical Filtering in the Use of ChatGPT

Recent educational research highlights the growing concern over the ethical aspects of AI use. A systematic review reveals that enthusiasm for technology often hides its ethical issues (Mouta et al., 2024). Topics such as culture, inclusion, and students' emotions are frequently neglected in the existing literature (Mouta et al., 2024). The ability of teachers plays a crucial role in the ethical deployment of AI in classrooms (Mouta et al., 2024). Therefore, it is essential to systematically enhance teacher training in AI ethics (Mouta et al., 2024).

The emphasis on ethical approaches that focus solely on adult perspectives is increasingly being questioned. Incorporating students' voices when setting ethical standards leads to a more practical and relatable understanding (Burriss et al., 2024). Ideally, teachers should engage in a dialogical process of ethical filtering rather than just delivering one-way instructions. Teachers' ethical commitment and pedagogical skills are strong indicators of students' learning outcomes, even surpassing the impact of their reported teaching practices (Rajabi, 2026). Therefore, the ethical role of teachers is crucial for the success of AI-supported learning.

TPACK and Teachers' Artificial Intelligence Literacy as Prerequisites for Ethical Mediation

Teachers' ability to facilitate ethical mediation relies on their mastery of the TPACK framework. In South Korea, primary school teachers show low confidence in AI content knowledge (Kim & Kwon, 2023). They identified twenty-two AI competencies that primary teachers should possess. Similarly, in Sri Lanka, a lack of content knowledge was the most notable issue (Rajapakse, 2026), surpassing gaps in technological and pedagogical skills (Rajapakse, 2026).

Teachers' competence in teaching artificial intelligence is also dynamic and continues to develop. Experienced teachers integrate TPACK with more mature pedagogical considerations (Yau et al., 2025). In contrast, novice teachers tend to focus mainly on content aspects (Yau et al., 2025). Another systematic review identified six dimensions of AI literacy for teachers (Feng & Carolus, 2026). These dimensions include contextual knowledge and continuous professional development (Feng & Carolus, 2026). The affective and ethical dimensions of AI literacy are shared across both teacher and student groups (Feng & Carolus, 2026).

School Leadership and Institutional Values in Technology Integration

School leadership significantly influences how teachers adopt artificial intelligence. The willingness of school principals to embrace change affects AI implementation success (Pietsch & Mah, 2025). Leaders' proactive agility and empathy are key supporting factors (Pietsch & Mah, 2025). Combining the transformative leadership framework with TPACK effectively explains the embedding of institutional values (Alshamsi, 2025). This approach promotes equitable access and effective change management (Alshamsi, 2025). Consequently, digital leadership is essential for ethical governance of AI.

Teachers' acceptance of AI depends heavily on how useful and easy to use they perceive it to be. These two factors are key predictors of whether teachers intend to adopt AI (Hamdan & Kamsin, 2026). However, the study did not examine the influence of religious values specifically. Teachers' views on AI ethics might be influenced by their institutional beliefs. This gap in research offers a valuable basis for the new thematic insights discussed in the next section.

Synthesis and Research Gap

The literature review synthesizes four interconnected themes: teacher mediation, ethical filtering, TPACK, and school leadership, which collectively constitute the AI ethics ecosystem. Nonetheless, these themes have seldom been analyzed together within a unified conceptual framework. Additionally,

religio-cultural aspects are scarcely featured in the existing literature. This omission will be explored further in the upcoming thematic discussion, which offers a new contribution rooted in Indonesia's religio-cultural context.

3. Research Method

This study used a systematic literature review (SLR) design. To ensure transparency in the selection process, the PRISMA 2020 protocol was followed. This method was selected because it provides a comprehensive overview of research trends and themes. Additionally, an SLR facilitates the synthesis of studies from various national contexts and educational levels.

3.1 Data Source and Search Strategy

The data were gathered from the Scopus database, accessed on 4 August 2026. Keywords used in the search included combinations of *"artificial intelligence*," *"teacher*," *"ethic**", and *"agency*." The search was limited to articles published from 2023 to 2026, a period chosen to reflect research trends after the rise of ChatGPT's popularity. The initial search yielded 114 documents, encompassing research articles and editorials.

3.2 Inclusion and Exclusion Criteria

The inclusion criteria comprised four key requirements. Firstly, the articles needed to focus on the role of teachers concerning artificial intelligence in education. Secondly, they had to address aspects of ethics, agency, or pedagogical responsibility. Thirdly, the selected articles were required to be empirical studies, systematic literature reviews, or conceptual works. Lastly, the articles had to be published in English and accessible in full text. Exclusion criteria were set to keep the review focused. Articles not addressing teachers' ethical roles or teacher agency were excluded. Technical studies on algorithms lacking pedagogical aspects were also omitted. Short opinion papers without clear methodology descriptions were excluded. Duplicates or articles missing complete abstracts were likewise filtered out.

3.3 PRISMA Selection Process

The selection process adhered to the four PRISMA stages: identification, screening, eligibility, and inclusion. Initially, 114 records were collected from the Scopus database. Since only one database was used, no duplicates were found. These records were then screened based on their titles and abstracts. Out of these, 43 were excluded for not aligning with the focus on ethics or teacher agency. As a result, 71 records moved on to the initial screening phase.

The eligibility of the remaining 71 articles was then assessed through full-text reading. Twenty-eight articles were excluded at this stage for three

reasons. Fifteen articles did not explicitly discuss teachers' ethical or moral roles. Six articles were conceptual studies without sufficient supporting empirical evidence. Seven articles were excluded because they focused exclusively on technology acceptance theory. As a result, 43 articles met all the eligibility criteria and were included in the thematic synthesis.

Table 1. Summary of the PRISMA Selection Process

PRISMA Stage	Description	n
Identification	Records identified from the Scopus database (2023–2026)	114
Identification	Duplicate records removed	0
Screening	Records screened based on title and abstract	114
Screening	Records excluded (not relevant to ethics/teacher agency)	43
Eligibility	Full-text articles assessed for eligibility	71
Eligibility	Articles excluded (ethical role not explicit; conceptual without empirical evidence; focused only on technology acceptance theory)	28
Inclusion	Articles included in the thematic synthesis	43

3.4 Data Analysis and Quality Assessment

The data were analyzed using a thematic synthesis method applied to 43 selected articles. Each article was manually coded based on publication year, country, educational level, research method, and main theme, through repeated readings of abstracts and findings. The identified themes were then organized into broader categories, leading to five main themes which are detailed in the findings section. The quality assessment evaluated how clear the research methodology was and how representative the sample was. Articles from Scopus-indexed journals were assumed to have been peer-reviewed. The assessment was qualitative and did not use a standardized appraisal tool. This limitation is clearly acknowledged in the study's limitations section.

4. Findings

This section outlines the characteristics and key findings of the 43 selected articles. It covers publication years, geographical distribution, educational levels, and research methods. Additionally, five main themes are discussed in relation to RQ1 and RQ2.

4.1 Characteristics of the Selected Studies

The distribution by publication year shows a continuing increase in research. The number of publications in 2023 was the lowest compared with

the following years. The number of publications increased substantially in 2025 and 2026. This increase is consistent with the widespread adoption of ChatGPT since the end of 2022. This trend confirms the growing importance of AI ethics as an emerging research issue. The studies are geographically distributed across over twenty countries, with East Asia and Southeast Asia representing the majority. Research from Indonesia, including North Sumatra (Pardede et al., 2026), was also noted. Europe and the Middle East also make notable contributions to the topic of school leadership. However, the database lacked any studies from Eastern Indonesia.

Most studies utilized quantitative methods, mainly surveys and statistical models, with the Partial Least Squares Structural Equation Modelling (PLS-SEM) approach frequently used to analyze the link between ethical AI use and learning outcomes (Rajabi, 2026). Additionally, qualitative research through interviews and case studies was common. A few studies used mixed-methods designs and systematic reviews. This variety of methods broadens the understanding of AI ethics in education.

4.2 Major Themes of the Thematic Synthesis

The thematic synthesis of the 43 selected articles identified five key themes: teachers' ethical mediation, teachers' AI literacy and competence, school leadership, student agency, and academic integrity. Table 2 provides a summary of these themes, along with the number of studies and representative supporting sources.

Table 2. Major Themes Identified from the Thematic Synthesis of the 43 Selected Articles

Theme	Number of Studies	Representative Sources
Teachers' ethical mediation	12	Pardede et al. (2026); Alasgarova & Rzayev (2025); Engeness (2025)
AI literacy and competence (TPACK)	11	Kim & Kwon (2023); Rajapakse (2026); Yau et al. (2025); Feng & Carolus (2026)
School leadership	9	Pietsch & Mah (2025); Alshamsi (2025); DeMatthews et al. (2026); Eryilmaz et al. (2025)
Student agency in AI ethics	6	Burriss et al. (2024); Granström & Oppi (2025); Kajan et al. (2026)
Academic integrity	5	Gonsalves (2025); Ofosu-Asare (2026); Mouta et al. (2024)

The first theme concerns teachers' ethical mediation in the use of generative artificial intelligence. Teachers play an important role in verifying AI-generated content and guiding students' reflection (Pardede et al., 2026). This mediation bridges students' cognitive abilities and the complexity of AI-generated outputs. The same theme also emerged in the Technology Acceptance Model (TAM) study conducted in Azerbaijan (Alasgarova & Rzayev, 2025). Ethical mediation therefore constitutes the core of teachers' professional agency in the era of artificial intelligence.

The second theme addresses teachers' AI literacy and skills within the TPACK framework. In various countries, teachers show gaps in AI content knowledge (Kim & Kwon, 2023; Rajapakse, 2026). These skills are built through experience and ongoing professional development (Yau et al., 2025). Six aspects of AI literacy were also identified across different national settings (Feng & Carolus, 2026). This theme suggests that AI literacy is a continuous process of growth rather than a fixed state.

The third theme addresses school leadership as a key element in ethical AI governance. School principals who possess a digital mindset promote effective AI integration (Pietsch & Mah, 2025). Additionally, a six-domain AI leadership framework has been developed specifically for schools (DeMatthews et al., 2026). Concerns over academic integrity continue to be a significant issue for school leaders (Eryilmaz et al., 2025). Furthermore, transformative leadership enhances the consistent application of the TPACK framework (Alshamsi, 2025).

The fourth theme highlights the role of student agency in shaping AI ethics. Co-design methods include students in revising ethical principles related to artificial intelligence (Burriss et al., 2024). A large-scale survey in Estonia uncovered gaps in how these principles are taught (Granström & Oppi, 2025). Additionally, research in STEM education noted students' social ambivalence towards AI chatbots (Kajan et al., 2026). Consequently, student perspectives are essential to participatory ethical frameworks.

The fifth theme addresses academic integrity and adherence to AI-use declarations. Most university students do not report their use of AI in academic work (Gonsalves, 2025). Vague institutional policies heighten the likelihood of integrity breaches. Implementing a clear ethical framework for AI could help mitigate these risks (Ofosu-Asare, 2026). This theme reinforces teachers' role as moral gatekeepers in upholding integrity in AI-assisted learning.

The five themes thoroughly address RQ1 and RQ2. Since 2025, research has grown significantly, especially regarding AI literacy and school leadership. The themes of ethics, teacher agency, and pedagogical responsibility are closely linked. However, religio-cultural contexts are missing from these

themes. This omission is examined in the thematic discussion, which highlights the unique contribution of this study in the next section.

5. Discussion

This section responds to RQ3 with a thematic discussion that provides a unique contribution. The five themes from the synthesis are combined into one conceptual framework, called '**teachers as moral gatekeepers of artificial intelligence**.' The study's novelty lies in its context within Indonesia's religio-cultural values.

5.1 Teachers as Moral Gatekeepers: A Conceptual Synthesis

The term '*moral gatekeepers*' highlights teachers' unique role. They filter, verify, and guide students' use of artificial intelligence (Pardede et al., 2026). This aligns with the idea of mediation in Cultural-Historical Activity Theory (Engeness, 2025). Teachers go beyond supervision to become students' dialogical partners (Burriss et al., 2024). To fulfill this role effectively, teachers need sufficient AI literacy (Feng & Carolus, 2026).

However, the role of moral gatekeeping becomes fragile without institutional backing. If artificial intelligence is seen as the only authority, teacher agency may diminish (Alasgarova & Rzayev, 2025). To prevent this, school leadership must bolster teachers' ethical skills (Pietsch & Mah, 2025). This synthesis highlights that moral gatekeeping is fundamentally relational, involving interactions among teachers, students, school leaders, and the institution's core values.

5.2 Novelty: Religio-Cultural Contextualisation in Catholic Secondary Schools in East Nusa Tenggara

This study's uniqueness lies in its focus on Catholic secondary schools in East Nusa Tenggara, where these institutions prioritize moral and spiritual character education rooted in Christian values over purely academic achievement. This offers a unique intersection between AI adoption and faith integrity, an area not explicitly explored in existing global literature. Teachers in these schools mediate ethics through four mechanisms: verifying ChatGPT-generated content, engaging in reflective dialogues, providing moral and spiritual guidance aligned with honesty and responsibility, and applying their TPACK and AI literacy skills. These approaches extend beyond mere factual verification, addressing broader ethical considerations. Catholic values — honesty, responsibility, and human dignity — serve as core ethical principles, expanding AI ethics beyond common concerns like privacy and bias. The spiritual dimension enriches pedagogical responsibility, yet has been overlooked in secular AI ethics frameworks. Additionally, limited digital

infrastructure in the region influences ethical mediation strategies, prompting teachers to depend on creativity and local values rather than technology alone. This approach contrasts with schools with better resources and highlights the importance of considering regional disparities in AI ethics discussions.

5.3 A Value-Based Conceptual Model of Moral Gatekeeping in Artificial Intelligence

The above synthesis presents a value-driven conceptual model of moral gatekeeping, placing students' interaction with ChatGPT at the core of the phenomenon. This interaction is mediated by teachers through verification, dialogue, spiritual guidance, and TPACK skills. The mediation is shaped by infrastructure, local culture, and institutional readiness.

The process aims to achieve three key outcomes: fostering strong student ethical character, enabling meaningful learning, and promoting responsible AI use. Teachers are positioned as central figures, not just technical enablers, continuously balancing AI integration with faith-based educational values. Although centered on Catholic contexts, the model can be adapted by any values-based school using these four ethical mediation mechanisms. This adaptability is a key distinction from previous generic reviews of AI ethics.

5.4 Thematic Implications Across Contexts

These thematic insights impact global AI ethics frameworks. UNESCO and secular approaches should allow more room for local values (Hsu & Lao, 2025). AI ethics training for teachers should not be a one-size-fits-all program but tailored to institutional values, including religious aspects when relevant. This aligns with findings that ethical commitment goes beyond technical practice alone (Rajabi, 2026).

Another implication relates to educational policy in regions with limited infrastructure. AI policies must recognize teachers' adaptive creativity in under-resourced areas, where they often act as key ethical filters. This role is seldom acknowledged in AI frameworks focused on urban settings. Recognizing this is crucial for making educational AI policies more inclusive and equitable.

6. Conclusion

This systematic review analyzed 43 articles focusing on teachers as moral gatekeepers of artificial intelligence. It identified five key themes: ethical mediation, AI literacy, school leadership, student agency, and academic integrity. These themes were combined into a value-based moral gatekeeping framework. The study's novelty is exemplified by the case of

Catholic secondary schools in East Nusa Tenggara, illustrating how religio-cultural values can enhance universal AI ethics.

6.1 Implications

This study's theoretical contribution is expanding secular AI ethics frameworks. It highlights that moral gatekeepers are better viewed as relational rather than solely individual actors. The Cultural-Historical Activity Theory framework effectively explains the mediation process among teachers, AI, and students. Additionally, religio-cultural values can act as an important and legitimate ethical layer. These results suggest new avenues for future research into AI ethics based on diverse value systems.

The practical implications focus on creating more context-aware teacher training programs. AI literacy education should combine TPACK skills with reinforcement of institutional values. Values-driven schools might craft unique modules for ethical mediation. School leaders need to implement policies that reinforce teachers' responsibilities as ethical mediators. Educational policies in under-resourced areas should acknowledge and foster the adaptive creativity of local teachers.

6.2 Limitations

This study has some limitations worth noting. First, the literature search was limited to only the Scopus database; incorporating other sources like Web of Science could offer a more comprehensive collection of relevant articles. Second, the assessment of article quality was qualitative and lacked a standardized evaluation tool. Third, the unique contribution of this study is based on a single case within eastern Indonesia.

The discussion of Catholic secondary schools in East Nusa Tenggara is based on an initial conceptual proposal that has not been confirmed through direct empirical research. As a result, the contextual findings should be interpreted cautiously. Additional studies involving empirical field data are necessary to confirm this proposition.

6.3 Recommendations

Future research should empirically explore the five proposed aspects of ethical mediation: content verification, Catholic values, TPACK, infrastructure context, and character development. Conducting field studies using interviews and classroom observations in Catholic secondary schools in East Nusa Tenggara is highly encouraged. Additionally, comparative studies across schools founded on various value systems would be valuable for further understanding.

Policymakers should prioritize enhancing AI ethics training rooted in local values, involving both school leadership and teachers. Researchers are encouraged to broaden their literature searches beyond the Scopus database. Additionally, fostering cross-national collaborations can enhance the generalisability of the moral gatekeeper framework.

References

Adewale, S., & Ndwandwe, N. D. (2025). Artificial intelligence-driven leadership in higher education: A bibliometric analysis of research trends and development. *International Journal of Learning, Teaching and Educational Research, 24(7)*. https://doi.org/10.26803/ijlter.24.7.31

Alasgarova, R., & Rzayev, J. (2024). The role of artificial intelligence in shaping high school students' motivation. *International Journal of Technology in Education and Science, 8*. https://doi.org/10.46328/ijtes.553

Alasgarova, R., & Rzayev, J. (2025). The implications of artificial intelligence for teacher agency and teacher-student relationships through the Technology Acceptance Model. *International Journal of Technology in Education and Science, 9*. https://doi.org/10.46328/ijtes.645

Alé, J., Ávalos, B., & Araya, R. (2025). Chilean teachers' knowledge of and experience with artificial intelligence as a pedagogical tool. *Education Sciences, 15(10)*. https://doi.org/10.3390/educsci15101268

Alshamsi, A. S. (2025). Integration of transformative leadership, artificial intelligence, and the TPACK framework for efficient pedagogy: A documentary analysis. *International Journal of Learning, Teaching and Educational Research, 24(9)*. https://doi.org/10.26803/ijlter.24.9.47

Beták, N., & Szőköl, I. (2026). Primary school teachers' perceived AI-related professional knowledge in Slovakia: A city-wide ITPACK study. *Frontiers in Education, 11*. https://doi.org/10.3389/feduc.2026.1813449

Berkovich, I. (2025). The rise of AI-assisted instructional leadership: Empirical survey of generative AI integration in school leadership and management work. *Frontiers in Education, 10*. https://doi.org/10.3389/feduc.2025.1643023

Burriss, S. K., Hutchins, N., Conley, Z., Deweese, M. M., Doe, Y. J., Eeds, A., Villanueva, A., Ziegler, H., & Oliver, K. (2024). Redesigning an AI bill of rights with/for young people: Principles for exploring AI ethics with middle and high school students. *Computers and Education: Artificial Intelligence, 6*. https://doi.org/10.1016/j.caeai.2024.100317

Chang, J., Park, J., & Sim, J. Y. (2026). Harnessing generative AI to facilitate epistemic understanding of students in scientific inquiry. *Journal of Science Education and Technology*. https://doi.org/10.1007/s10956-026-10298-5

Chassida, Y. (2026). Transformational leadership as a contextual enabler of teachers' AI use. *Education Sciences, 16(4)*. https://doi.org/10.3390/educsci16040572

DeMatthews, D., Reyes, P., Hart, T. D., & James, L., III (2026). Leadership for artificial intelligence use in schools: A six-domain framework for ethical, equitable, and effective integration. *Educational Management Administration & Leadership*. https://doi.org/10.1177/17411432261418940

Dimeli, M., & Kostas, A. (2025). The role of ChatGPT in education: Applications, challenges: Insights from a systematic review. *Interdisciplinary Journal of Information, Knowledge, and Management, 20*. https://doi.org/10.28945/5422

Elkhodr, M., & Gide, E. (2025). The SAGE framework for developing critical thinking and responsible generative AI use in cybersecurity education. *Discover Education, 4*. https://doi.org/10.1007/s44217-025-00935-3

Engeness, I. (2025). Cultural-historical perspective to design pedagogical AI for enhancing student writing. *Technology, Knowledge and Learning*. https://doi.org/10.1007/s10758-025-09876-0

Eryilmaz, N., Mah, D.-K., Bellibas, M. S., & Pietsch, M. (2025). School leaders' expectations of AI's impact on students and teachers: Insights from ICILS 2023. *Educational Management Administration & Leadership*. https://doi.org/10.1177/17411432251401487

Feng, S., & Carolus, A. (2026). Artificial intelligence literacy at school: A systematic review with a focus on psychological foundations. *Computers and Education: Artificial Intelligence, 10*. https://doi.org/10.1016/j.caeai.2026.100551

Flogie, A., Hari, D., Skrbinjek, V., & Brejc, M. (2026). Use and aspects of generative artificial intelligence in the educational system among leadership personnel. *Education and Information Technologies*. https://doi.org/10.1007/s10639-026-13986-2

Frick, L., & Tshuma, N. (2026). Reconsidering researcher independence in an era of AI enabled learning. *Teaching in Higher Education*. https://doi.org/10.1080/14703297.2026.2631580

Gonsalves, C. (2025). Addressing student non-compliance in AI use declarations: Implications for academic integrity and assessment in higher education. *Assessment & Evaluation in Higher Education*. https://doi.org/10.1080/02602938.2024.2415654

Granström, M., & Oppi, P. (2025). Student engagement with AI tools in learning: Evidence from a large-scale Estonian survey. *Frontiers in Education, 10*. https://doi.org/10.3389/feduc.2025.1688092

Hamdan, H. H., & Kamsin, I. F. (2026). Secondary teachers' acceptance of generative artificial intelligence in teaching based on the TAM.

International Journal of Information and Education Technology, 16(7). https://doi.org/10.18178/ijiet.2026.16.7.2638

Hsu, T.-C., & Lao, N. (2025). Editorial: From policy to pedagogy - Building human-centered AI literacy across educational contexts. *Informatics in Education*. https://doi.org/10.15388/infedu.2025.00

Kajan, K., Shi, W., Wanatowski, D., & Ryan, M. (2026). Navigating the dual-view phenomenon: Social ambivalence, ambivalence literacy, and lecturer role transformation in AI-integrated transnational STEM education. *Education Sciences, 16(4)*. https://doi.org/10.3390/educsci16040554

Kahila, J., Vartiainen, H., Tedre, M., Arkko, E., Lin, A., Pope, N., Jormanainen, I., & Valtonen, T. (2024). Pedagogical framework for cultivating children's data agency and creative abilities in the age of AI. *Informatics in Education, 23*. https://doi.org/10.15388/infedu.2024.15

Kim, K., & Kwon, K. (2023). Exploring the AI competencies of elementary school teachers in South Korea. *Computers and Education: Artificial Intelligence, 4*. https://doi.org/10.1016/j.caeai.2023.100137

Kong, S. C., & Hu, W. (2026). Unleashing human potential: An artificial intelligence competency framework for K-12 education. *Computers and Education: Artificial Intelligence, 10*. https://doi.org/10.1016/j.caeai.2026.100556

Marsalek, R., & Teplá, M. (2026). "Is AI inevitable?" Development of attitudes and practices of Czech teachers between 2023 and 2025. *Education Sciences, 16(2)*. https://doi.org/10.3390/educsci16020335

Mešková, L. (2026). Exploring the IntelligentTPACK gap: A qualitative analysis of teachers' AI competencies and AI self-reported classroom uses. *Frontiers in Education, 11*. https://doi.org/10.3389/feduc.2026.1790642

Mouta, A., Pinto-Llorente, A. M., & Torrecilla-Sánchez, E. M. (2024). Uncovering blind spots in education ethics: Insights from a systematic literature review on artificial intelligence in education. *International Journal of Artificial Intelligence in Education*. https://doi.org/10.1007/s40593-023-00384-9

Ofosu-Asare, Y. (2026). Guiding the future: Developing an ethical framework for generative AI use in education. *The International Journal of Information and Learning Technology*. https://doi.org/10.1108/IJILT-06-2024-0113

Pardede, S., Simamora, S., & Sinaga, D. (2026). ChatGPT in secondary education: A multi-theoretical case study from a developing country. *Africa Education Review*. https://doi.org/10.25159/1947-9417/20308

Pietsch, M., & Mah, D.-K. (2025). Leading the AI transformation in schools: It starts with a digital mindset. *Educational Technology Research and Development*. https://doi.org/10.1007/s11423-024-10439-w

Rajabi, K. I. (2026). Modeling school teachers' ethical use of AI tools in English instruction and its impact on student learning: A PLS-SEM approach. *International Journal of Information and Education Technology, 16(1)*. https://doi.org/10.18178/ijiet.2026.16.1.2490

Rajapakse, C. P. (2026). A TPACK-assisted descriptive analysis of the Sri Lankan ICT teachers' readiness to teach AI in schools. *International Journal of Information and Education Technology, 16(2)*. https://doi.org/10.18178/ijiet.2026.16.2.2527

Relmasira, S. C., & Donaldson, J. P. (2025). Deep learning pedagogies enhance AI literacy in elementary students: A five-cycle implementation study. *Online Learning and Educational Review, 5(1)*. https://doi.org/10.58524/oler.v5i1.720

Sun, Z. L., & Chan, Y. Y. (2025). AI literacy development in middle school: Profiles and demographic predictors from an ABCE perspective. *Computers and Education: Artificial Intelligence, 8*. https://doi.org/10.1016/j.caeai.2025.100480

Taktak, M., Bellibaş, M. Ş., & Özgenel, M. (2024). Use of ChatGPT in education: Future strategic road map with SWOT analysis. *Educational Process: International Journal, 13(3)*. https://doi.org/10.22521/edupij.2024.133.1

Tang, Y., & Zhong, L. (2026). K-12 teachers' adoption of generative AI for teaching: An extended TAM perspective. *Education Sciences, 16(1)*. https://doi.org/10.3390/educsci16010136

Tripathi, T., Sharma, S. R., Singh, V., Bhargava, P., & Raj, C. (2025). Teaching and learning with AI: A qualitative study on K-12 teachers' use and engagement with artificial intelligence. *Frontiers in Education, 10*. https://doi.org/10.3389/feduc.2025.1651217

Uğur, S. (2026). Beyond tool use: An analysis of prompt literacy and educators' readiness for artificial intelligence in schools. *Interdisciplinary Journal of Information, Knowledge, and Management, 21*. https://doi.org/10.28945/5751

Wu, D., Chen, M., Chen, X., & Liu, X. (2024). Analyzing K-12 AI education: A large language model study of classroom instruction on learning theories, pedagogy, tools, and AI literacy. *Computers and Education: Artificial Intelligence, 7*. https://doi.org/10.1016/j.caeai.2024.100295

BOOK CHAPTER

REHUMANIZING EDUCATION IN THE AI ERA:

Reimagining School Relevance, Teacher Roles, and Assessment Integrity for Critical Learners



UNTAG PRESS

ISBN 978-634-264-014-2 (PDF)

9 786342 640142