www.editada.org

International Journal of Combinatorial Optimization Problems and Informatics, 17(4), 2026, 10-24. ISSN: 2007-1558.
https://doi.org/10.61467/2007.1558.2026.v17i4.1298

Ethical Governance of Artificial Intelligence in Higher Education: A Systematic Review and a Proposed Socio-Technical Framework for Smart Educational Environments

Gerardo Antonio Hernández-Torres[1], Ariel Gutiérrez Valencia[2], Armando Morales Murillo[2], Antonio Becerra Hernández[3]*

[1]Dirección del Sistema Bibliotecario, Universidad Juárez Autónoma de Tabasco, Villahermosa, Tabasco, Mexico.
[2]División Académica de Educación y Artes, Universidad Juárez Autónoma de Tabasco, Villahermosa, Tabasco, Mexico.
[3]División Académica de Ciencias de la Salud, Universidad Juárez Autónoma de Tabasco, Villahermosa, Tabasco, Mexico.
Email address(es): gerardo.hernandez@ujat.mx, ariel.gutierrez@ujat.mx, armando.morales@ujat.mx, antonio.becerra@ujat.mx

✉Corresponding author: gerardo.hernandez@ujat.mx

Abstract. The growing use of artificial intelligence (AI) in higher education has intensified debates on academic integrity, authorship, and institutional governance, especially in intelligent education environments. Although international guidelines for the responsible use of AI exist, they often adopt broad normative approaches and have limitations when applied in specific institutional contexts. This study proposes a socio-technical framework for the ethical governance of AI in higher education. Through a systematic review following PRISMA guidelines in Scopus and Web of Science (2022–2025), a final corpus of 64 studies was selected from which normative gaps were identified, and five interdependent dimensions were structured. As an additional output, the study also presents an operational matrix with indicative variables and indicators to support institutional diagnostics and decision-making regarding the responsible use of AI. The findings suggest that the ethical adoption of AI requires dynamic, context-sensitive models that balance innovation, academic integrity, and human oversight.

Keywords: ethical governance of artificial intelligence; artificial intelligence in higher education; socio-technical framework; responsible artificial intelligence; academic integrity; intelligent educational environments; institutional AI governance; human oversight; PRISMA systematic review; educational technology governance; AI ethics in education; institutional diagnosis.

Article information

Received: Jan 17, 2026
Accepted: May 19, 2026

1 Introduction

The rapid growth and integration of Artificial Intelligence (AI) tools in higher education have transformed the way we teach, learn, and conduct research, offering opportunities to enhance academic output, personalize learning, and generate knowledge, while also raising ethical dilemmas regarding authorship, plagiarism, and academic integrity.

Recent studies indicate that AI can improve writing, virtual tutoring, and personalized learning, thereby increasing student efficiency and motivation (Ayoubi, 2024; Mahmud et al., 2024). In the wake of the pandemic, it has been noted that AI can contribute to teacher well-being and professional self-efficacy through more flexible teaching practices (Alhwaiti, 2023).

However, some authors argue that its adoption faces challenges related to data privacy, content errors, and limited curricular integration, especially in technical and scientific fields (Li & Qin, 2023).

In graduate programs, where original authorship, academic integrity, and intellectual responsibility are fundamental elements of student training, the use of AI is associated with streamlining research processes, but also with issues of trust, originality, and accuracy (Mohd Rahim et al., 2022; Nemt-allah et al., 2024). To address this, international organizations such as UNESCO (2023, 2025) and COPE (2023) have developed guidelines for the responsible use of AI in education, proposing transparency, disclosure of use, and regulation of assisted authorship; however, in many Latin American universities, the absence of specific protocols continues to leave regulatory and educational gaps regarding these dilemmas.

It is necessary to analyze the ethical use of AI in higher education as a socio-technical system, in which technology, pedagogical practices, and institutional governance interact dynamically. The objective of this study is to identify, through a systematic review, the ethical dilemmas and concerns reported regarding the use of AI in higher education (2022–2025) and to derive a socio-technical framework for ethical governance. This will result in a proposal for operationalization, in the form of a matrix of variables and indicators, to support institutional assessments.

This proposal facilitates institutional decision-making teams' work by translating evidence from the literature into a socio-technical framework and a set of indicators integrated into diagnostic, auditing, and continuous improvement processes within educational organizations. The framework supports decision-making and governance in educational settings that are increasingly mediated by AI.

2 Methodology: Systematic Review

To analyze the ethical issues and concerns associated with the inappropriate use of AI in higher education, a systematic review was conducted which, according to Page et al. (2021), allows for the structured synthesis and identification of relevant trends and gaps in existing research. To enhance the transparency, reliability, and reproducibility of the process, the review was conducted in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines, particularly regarding the identification, selection, and eligibility of studies.

The databases selected for document retrieval were Scopus and Web of Science (WoS), as they are internationally recognized as repositories of high-impact academic journals with high-quality editorial standards (Zhu & Liu, 2020).

The search was conducted in July 2025, restricted to English and Spanish, and limited to the period 2022–2025. A search query was developed and used in both databases; for Scopus, the search was applied to the fields “Article title, Abstract, Keywords,” and for WoS, to the “Topic” field:

(undergraduate OR postgraduate OR “higher education” OR university) AND (“artificial intelligence” OR AI OR ChatGPT) AND (“academic performance” OR adoption) AND (perception OR acceptance OR challenges OR benefits OR impact OR uses)

A total of 1,314 documents were initially retrieved, of which 519 were from WoS and 795 from Scopus; these were exported to remove duplicate records and perform the screening in accordance with the PRISMA flowchart.

Next, inclusion and exclusion criteria were applied to screen the articles, as detailed in Table 1, covering aspects such as publication between 2022 and 2025, the languages English and Spanish, and full-text availability, in addition to focusing on the topic of perceptions, uses, ethical implications, and concerns related to AI in the context of higher education.

Table 1. Inclusion and exclusion criteria used for article screening

Inclusion criteria	Exclusion criteria
Articles published between 2022 and 2025	Articles outside the study period
Not systematic reviews	Systematic reviews, books, book chapters, or theses
Available in English or Spanish	Languages other than English or Spanish
Full-text available	No full-text access
	Duplicate articles
	Retracted articles
Articles related to artificial intelligence and ethical implications and concerns regarding its use in higher education	Articles not relevant to the research topic

For the selection based on title or abstract, documents were considered that, in context, pertained to higher education (universities, undergraduate/graduate programs, or university stakeholders) and whose content addressed AI in relation to perceptions, use, impact, barriers, or ethical implications within that context.

In the next eligibility phase, the full-text documents were reviewed, during which, in addition to verifying the year, language, and document type, texts were excluded that, while mentioning AI, addressed other educational levels (e.g., high school) or non-educational settings (e.g., businesses, hospitals), or that had no direct relation to perceptions, use, impact, barriers, or ethics in higher education.

During the screening by title or abstract and the full-text review, the reasons for exclusion were recorded using codes: E1 = outside the 2022–2025 period; E2 = language other than English or Spanish; E3 = ineligible document type (systematic review, book, chapter, or thesis); E4 = no access to full text; E5 = duplicate; E6 = retracted; E7 = unrelated to the topic (not higher education or does not address perceptions, use, impact, barriers, or ethical implications of AI in higher education).

The documents retrieved from the databases were exported to Rayyan for management. Here, duplicates were removed and the documents were screened by title or abstract, applying the eligibility criteria (Table 1). Subsequently, the remaining studies underwent a full-text review. To support this phase, SciSpace was used to group and organize the documents by thematic categories and facilitate their reading; however, the final decisions regarding inclusion and exclusion were made manually. The screening and the final decision on the inclusion or exclusion of the documents were carried out by a single reviewer; ambiguous cases were resolved by verifying the full text, applying the operational criteria and reasons E1–E7.

This process resulted in 107 articles for a detailed evaluation of their content, with the aim of confirming their thematic relevance. Finally, 64 studies met all established criteria and were included in the qualitative analysis of the review. Figure 1 details the document systematization procedure according to the PRISMA methodology.

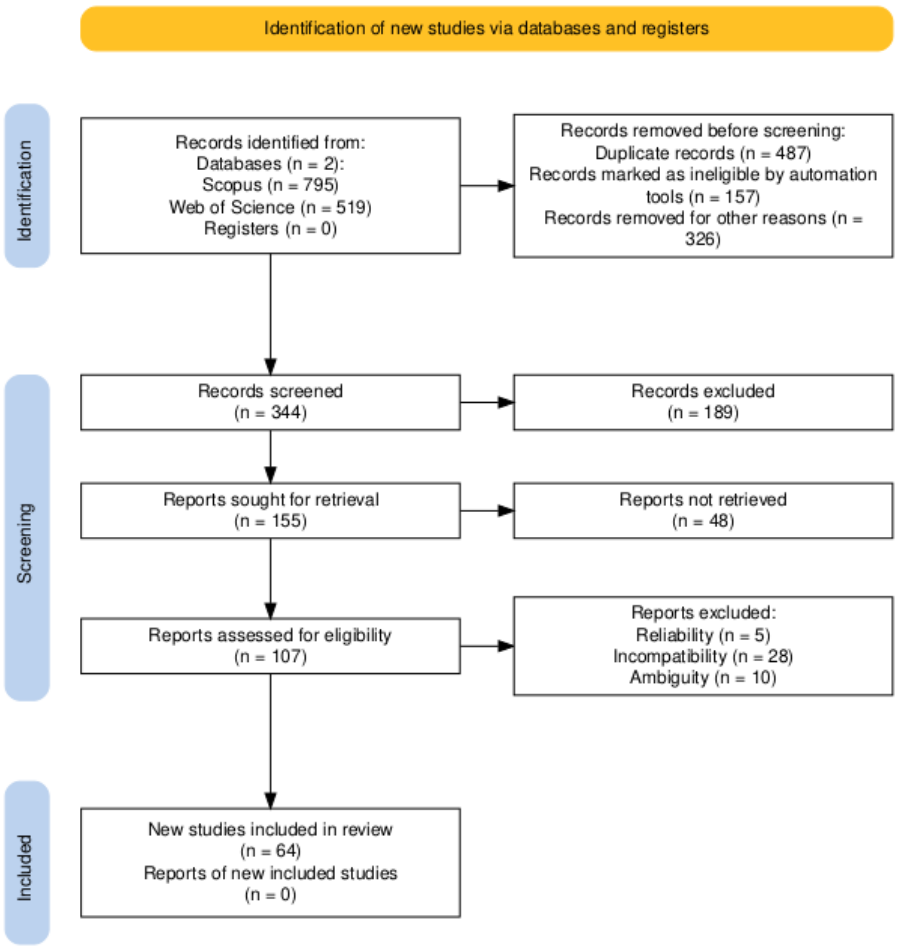

Fig. 1. PRISMA-based systematic review flow diagram.
Note: Adapted from the PRISMA Flow Diagram tool by Haddaway et al. (2022), *Campbell Systematic Reviews.*

To facilitate understanding of the final set of documents, a summary was prepared by year of publication, study region, study design, and primary ethical dimension.

Table 2. Characterization of the included corpus

Category	Count
Year of publication	
2023	11
2024	53
Study region	
Asia	32
Latin America	14
Europe	12
Africa	3
Oceania	2
Multiregional	1
Study type	
Quantitative	43
Qualitative	10
Mixed-methods	11
Primary ethical dimension	
Ethical and digital literacy for responsible use	48

Academic integrity in the face of opaque generation processes	13
Need for institutional policies regulating the use of AI	3

The ethical debate surrounding the use of AI in higher education is becoming increasingly important, as its adoption not only enhances educational efficiency and accessibility but also raises challenges related to authorship, originality, and transparency (Al-Zahrani, 2024; Gandhi & Gani, 2024). Recent literature highlights three recurring challenges:

1. Ethical and digital literacy for responsible use.
   The problem is not the use of AI, but how it is used; a lack of knowledge regarding citation, source acknowledgment, and the disclosure of writing assistant use—among both students and faculty (Nemt-allah et al., 2024; Niloy et al., 2024)—can increase the risk of technological dependence and unintentional plagiarism, thereby undermining academic credibility.
2. Academic integrity in the face of a lack of clarity in generation processes.
   The attribution of authorship and originality is compromised by the emergence of AI-generated texts that can be difficult to distinguish from those produced without assistance (Gandhi & Gani, 2024). In this regard, Cedeño Meza et al. (2024) note that explicitly declaring the use of AI is essential to ensure transparency and traceability by clarifying the tool's role and the author's responsibility.
3. The need for institutional policies regulating the use of AI.
   A regulatory gap is evident; Bouteraa et al. (2024) and Polyportis & Pahos (2024) note that many universities lack ethical protocols and institutional frameworks for the use of AI, which increases uncertainty among users, undermines trust in academic processes, and raises doubts about the role of the instructor as a facilitator of learning.

At the international level, organizations such as UNESCO and the OECD have promoted policies aimed at ensuring the ethical, inclusive, and safe use of AI. The OECD (2023) proposes incorporating ethics throughout the entire technology lifecycle, while UNESCO (2025), in its survey, indicates that two-thirds of higher education institutions have already developed guidelines for the responsible use of AI, addressing transparency, authorship, equity, and data protection. However, few universities have established clear mechanisms to implement and monitor these policies.

Table 3. Synthesis of findings from the literature review on ethics in the use of AI

Category	Key findings	Articles
Ethical training and literacy	There is a need to train students and faculty in ethics, citation practices, and the responsible use of AI.	(Nemt-allah et al., 2024; Niloy et al., 2024)
Academic integrity	There is a risk of weakening trust in postgraduate programs if the use of AI is not adequately regulated.	(Al-Zahrani, 2024; Okulich-Kazarin et al., 2024)
Plagiarism	AI introduces new forms of plagiarism that challenge traditional detection methods.	(Gandhi & Gani, 2024; Thomas et al., 2023)
Transparency and authorship	Declaring the use of AI in academic work is essential to prevent authorship conflicts.	(Cedeño Meza et al., 2024)
Institutional policies	There is an urgent need to design guidelines and protocols for responsible AI use; the literature reveals a significant regulatory gap.	(Bouteraa et al., 2024; Polyportis & Pahos, 2024)

The evidence suggests that ethics is central to the sustainable development of AI in higher education. While the need for ethical literacy and transparency in authorship is clear, there are differences of opinion regarding the boundaries between legitimate use and plagiarism, as well as over whether it is faculty, institutions, or technology developers who should ensure academic integrity. The results of this review form the basis for the design of the socio-technical framework proposed in the following sections.

3 Conceptual and Theoretical Framework for the Ethical Analysis of AI Use in Higher Education

Smart Learning Environments (SLEs) are described as intelligent educational technologies that provide adaptive support, taking into account the student's learning status, individual needs, and the online and real-world contexts in which the student learns (Alonso-Prieto et al., 2025). In this study, this concept is adopted at the institutional level and understood as university contexts in which digital technologies and data-driven systems, including AI, are integrated to support educational and management processes. Consequently, the analysis is limited to four institutional domains: teaching, assessment, research, and academic management, as these are areas where dilemmas of authorship, transparency, privacy, and decision-making associated with the use of AI are concentrated. Understanding the ethical and pedagogical dilemmas associated with the use of AI in higher education requires a transdisciplinary approach that combines different theoretical frameworks. In this context, academic integrity serves as a normative framework that guides responsible practices in the production and assessment of knowledge. According to the International Center for Academic Integrity, academic integrity is grounded in values such as honesty, trust, fairness, respect, responsibility, and courage, which today serve as practical criteria for making decisions regarding authorship, transparency, and the appropriate use of automated tools—especially in scenarios where it is difficult to precisely identify how a text was produced.

To understand why teachers and students adopt or reject these tools, despite ethical concerns, the Technology Acceptance Model (TAM) posits that the intention to use a technology is primarily explained by perceived usefulness and perceived ease of use (Davis, 1989). Along the same lines, the Unified Theory of Acceptance and Use of Technology 2 (UTAUT2) expands the framework by incorporating determinants such as social influence, enabling conditions, habit, and other drivers of intention to use (Venkatesh et al., 2012). This is useful because the adoption of AI in higher education depends not only on how good the tool is, but also on the ecosystem: established norms, institutional support, environmental pressure, and work routines. Specifically regarding generative AI, the TAME-ChatGPT model proposes measuring dimensions that typically explain attitudes and usage, incorporating—in addition to usefulness and ease of use—factors associated with interaction with the tool, such as perceived risk and tech anxiety, thereby allowing for a more accurate capture of the tension between benefit and ethical caution (Sallam et al., 2023). In this context, the risk lies not only in errors in the information but also in the loss of traceability in the writing process, ambiguous attribution of authorship, or dependence on core tasks in higher education.

Finally, coevolutionary theory provides a systemic framework for understanding how technology and academic culture (norms, assessment, teaching processes) influence and adapt to one another (Jordano, 2003; Oyama, 1986). Thus, the ethical dilemmas associated with AI do not arise solely from the tool itself, but from the way in which academic communities redefine what counts as authorship, how learning is assessed, and what is considered valid evidence when these technologies are integrated into university life.

The integration of these approaches provides a conceptual framework for analyzing the use of AI in higher education from multiple levels. Academic integrity establishes the normative framework; TAM and UTAUT2 explain individual adoption; TAME-ChatGPT extends the analysis to the emotional and perceptual dimensions; and coevolution allows us to interpret the phenomenon as part of a dynamic social system. Together, these approaches provide a conceptual foundation for the need to move toward an integrative model that allows us to understand AI not only as a technological tool, but also as an ethical, organizational, and systemic phenomenon.

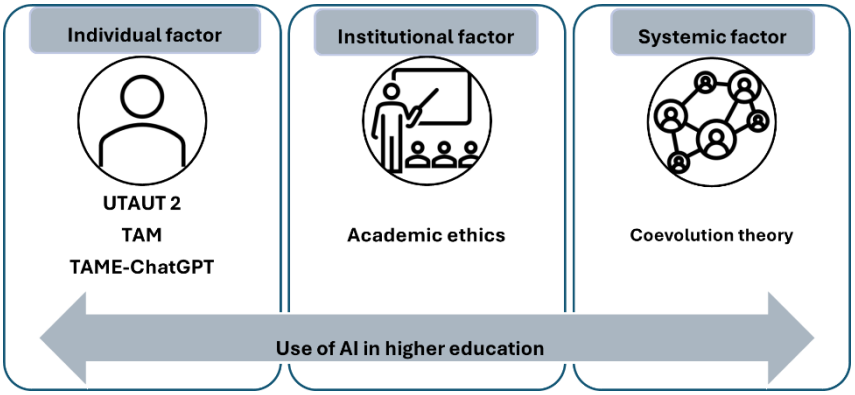

Fig. 2. Conceptual Reference Framework for the Analysis of AI Use in Higher Education.

As summarized in Figure 2, this conceptual framework articulates three levels of analysis—individual, institutional, and systemic—which serve as the basis for developing the socio-technical framework proposed in this study, aimed at the ethical governance of AI in smart education environments.

TAM and UTAUT2 were incorporated as they are widely validated frameworks for explaining the intention and use of technologies in higher education, which allowed for the analysis of utility, ease of use, social influence, and enabling conditions. TAME-ChatGPT was added due to its greater specificity regarding generative AI, as it captures the tension between perceived benefits and risk factors, as well as between anxiety and trust. Academic integrity provides the normative axis (authorship, integrity, transparency), and Coevolution Theory allows for interpreting the phenomenon as a dynamic readjustment between technology, academic culture, and governance. Together, these approaches align with the dimensions of the framework: adoption (TAM/UTAUT2), risk (TAME-ChatGPT), ethics and instructional design (academic integrity), and adaptive governance (coevolution).

4 Proposed Socio-Technical Framework

After reading the full text of the documents, data segments related to ethical risks, academic practices, adoption or use, instructional design or assessment, and institutional or governance conditions were extracted. These units were grouped by thematic coding into preliminary categories and subsequently transformed into five dimensions of the framework.

The integration of AI into higher education should not be viewed solely as the adoption of a new technology. In reality, a socio-technical system is needed in which digital tools, faculty, students, institutional norms, teaching practices, and academic values are interconnected. From this perspective, ethical dilemmas arise not only from the technology itself but also from the human and institutional decisions that guide its use in education and research.

This study proposes a socio-technical framework for the ethical governance of AI in higher education, geared toward smart education environments; it consists of five interdependent dimensions that enable the use of AI to be analyzed, regulated, and guided in a comprehensive manner, rather than through proposals focused solely on detection, control, or isolated regulatory measures (see Figure 3).

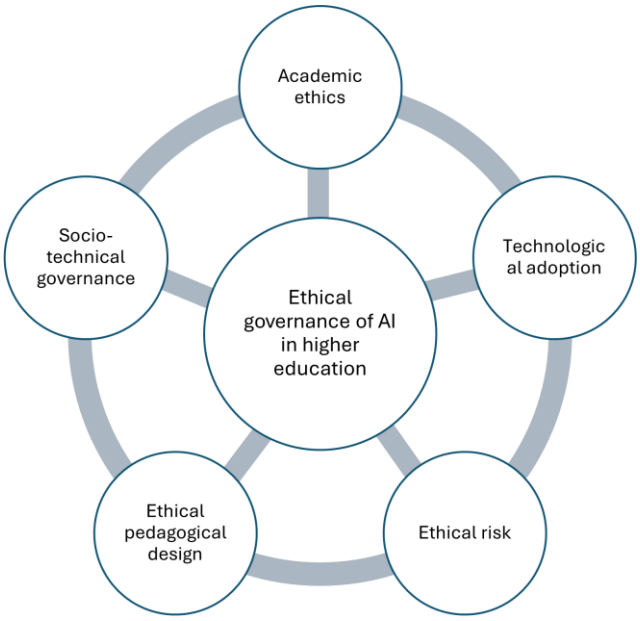

Fig. 3. Socio-Technical Framework for the Ethical Governance of AI in Higher Education.

To ensure traceability between the reviewed evidence and the framework, Table 4 summarizes, for each dimension, the core themes identified and examples of studies representative of the corpus.

Table 4. Traceability of the framework dimensions

Framework dimension	Themes	Representative evidence	Count
Academic integrity	Integrity, authorship, plagiarism and undisclosed use, citation and transparency, originality	(Acosta-Enriquez et al., 2024; Bouteraa et al., 2024; Niloy et al., 2024)	13
Technological adoption	Acceptance and intention to use, perceived usefulness and ease of use, trust, motivations and barriers to adoption	(Ayoubi, 2024; Jo & Bang, 2023; Polyportis & Pahos, 2024)	18
Ethical risk	Bias, privacy and security, reliability and hallucinations, dependency, impact on critical thinking	(Al-Zahrani, 2024; Grájeda et al., 2024; Sobaih, 2024)	27
Ethical pedagogical design	Pedagogical and assessment redesign, curricular integration, training and literacy for responsible use	(Chai et al., 2024; Dallari et al., 2024; Okulich-Kazarin et al., 2024)	3
Socio-technical governance	Institutional policies, guidelines and protocols, roles and accountability, auditing and monitoring	(Alanzi et al., 2023; Cedeño Meza et al., 2024; Ivanov et al., 2024)	3

4.1 Dimensions of the Framework

The framework is structured around five dimensions, each with a specific function within the socio-technical system:

- Dimension 1. Academic integrity.
  This constitutes the normative and cultural core of the framework. It incorporates principles of academic honesty, transparency, and responsible authorship, as well as formal and informal norms that define what constitutes legitimate use of AI in academic activity. Its primary function is to establish the ethical and cultural foundations that shape both the adoption and the ways in which AI is used.
- Dimension 2. Technological adoption.
  Based on the TAM and UTAUT2 models, it explains why and how academic actors incorporate AI into their routines. It considers factors such as perceived usefulness, ease of use, social influence, and enabling conditions, recognizing that AI adoption occurs even in the absence of clear institutional regulations. This dimension allows for the description and prediction of usage behavior based on individual and contextual factors.
- Dimension 3. Ethical risk.
  Derived from the TAME-ChatGPT model, this dimension analyzes perceptions of risk, behavior, and technological anxiety associated with the use of AI. It includes concerns such as a decline in critical thinking, ambiguity regarding authorship, technological dependence, and the reliability of generated results. The goal is not to stigmatize AI, but to model why, even with high perceived utility, users may experience resistance, covert use, or instrumental use. Its function will be to capture the tension between efficiency benefits versus ethical and academic costs.
- Dimension 4. Ethical pedagogical design.
  It promotes teaching and assessment methods that prioritize evidence of the learning process, critical reflection, and the application of knowledge. It also incorporates information and ethical literacy, respect for copyright, and ongoing verification of AI-generated content. Its purpose is to prevent technology from taking center stage and, instead, to strengthen meaningful learning. It also seeks to translate ethical principles into concrete, observable, and assessable teaching practices.
- Dimension 5. Socio-technical governance.
  This encompasses the formulation of usage policies, integration rules, and institutional mechanisms for monitoring and continuous adjustment. This includes the creation of interdisciplinary committees, periodic policy reviews, incident analysis, and the updating of criteria, with the recognition that technology, academic culture, and regulations evolve together. In this sense, its purpose is not only to regulate but also to strengthen the institutional capacity for learning, adaptation, and updating of the ethical and pedagogical framework.

The five dimensions of the framework operate across three levels of analysis: individual, institutional, and systemic. This structure makes the framework a flexible tool that can guide policy-making, instructional design, and the ethical evaluation of AI use in higher education.

Taken together, the proposed socio-technical framework offers an integrative and operational structure for the ethical governance of AI in higher education, in line with current demands for transparency, accountability, and continuous adaptation.

5 Operationalization of the Framework

The socio-technical framework for the ethical governance of AI presented in this study is an analytical and normative tool. Its aim is to help understand how AI is integrated into education, assess the state of ethical governance, and guide institutional decisions toward responsible use. It does this by constructing an analytical matrix that identifies patterns of practice, levels of institutional maturity, pedagogical tensions, and areas of ethical risk.

Here, operationalization involves converting each dimension of the framework into observable elements, such as analytical variables, criteria, and empirical evidence. These can be collected using quantitative, qualitative, or mixed methods. This approach aligns with the analysis of policies, governance, and socio-technical systems, where evaluation focuses on diagnosis and decision-making rather than strict psychometric measurement (Salazar Vargas, 2009).

The objective of this operationalization is for the framework to serve as a diagnostic and guidance tool. Thus, institutions can identify strengths, weaknesses, and tensions in the ethical governance of AI, without imposing a single model and allowing for adaptation according to educational level, discipline, institutional culture, and regulatory frameworks. It can also be used to audit institutional policies and as the basis for a maturity model by dimension.

5.1 Variables and Indicators by Dimension

We propose an initial set of analytical variables based on the dimensions of the framework, covering normative, cultural, organizational, and pedagogical aspects relevant to the ethical governance of AI. These variables are accompanied by indicators—defined as observable or declarative evidence—that allow for an assessment of the state of development of each dimension at different levels of analysis (individual, institutional, and systemic).

Table 5 summarizes this analytical matrix, conceived as a flexible and open structure designed to adapt to diverse educational contexts and used as a basis for the development of specific tools, such as surveys, interview guides, checklists for policy documents, institutional evaluation rubrics, or participatory assessments.

Table 5. Analytical Matrix for the Evaluation of Ethical Governance of AI in Higher Education

Dimension	Analytical variables	Empirical indicators	Level of analysis	Recommended instrument
Academic integrity	Academic integrity; transparency; responsible authorship; normative coherence	Existence of ethical AI guidelines; clarity in authorship criteria; perception of academic fairness; coherence between formal rules and actual practices	Institutional; systemic	Document audit (regulations and guidelines), survey (perceived fairness and transparency), interviews with faculty and committees
Technological adoption	Perceived usefulness; ease of use; social influence; facilitating conditions	Self-reported frequency of AI use; institutional access to tools; availability of technical and pedagogical support; perceived academic usefulness	Individual; institutional	Survey (frequency, intention, conditions), institutional usage analytics if available, brief interviews
Ethical risk	Perceived risk; technological anxiety; trust in AI;	Perceived technological dependency; level of trust in generated	Individual	Survey (risk, trust, anxiety), interviews or focus groups for cases and tensions

	fear of loss of human authorship	outputs; concern about plagiarism or improper attribution; acceptance of AI use		
Ethical pedagogical design	Authentic assessment; ethical and digital literacy; reflective AI use	Use of process-centered assessments; students' reflective evidence; training in citation and copyright; critical understanding of AI-generated outputs	Individual; institutional	Review of assessment instruments and assignments (documentary), faculty interviews, rubric and checklist for authentic assessment
Socio-technical governance	Institutional policies; interdisciplinary committees; normative adaptation; feedback mechanisms	Existence of AI committees; periodic policy review; protocol updates; continuous improvement mechanisms	Institutional; Systemic	Document audit (policies and protocols), interviews with administrators and committee, checklist of continuous improvement mechanisms

These indicators make it possible to assess the system's performance at different levels without imposing a single model, thereby respecting institutional and cultural diversity.

The framework can be used as a diagnostic tool, enabling institutions to identify their level of development in the ethical governance of AI. It can also be used for institutional audits and as the basis for a maturity model organized by dimension. Based on the proposed indicators, it is possible to construct institutional profiles that reflect strengths, weaknesses, and priority areas for action.

To facilitate the use of the matrix as a diagnostic tool, a simple institutional maturity scale by dimension (1–4) is proposed, in which scores are assigned based on evidence available in documents, practices, and collected perceptions (see Table 6).

Table 6. Maturity scale (Levels 1–4)

Level	Description
Level 1 – Incipient	There are no guidelines, training, or monitoring mechanisms; practices are ad hoc.
Level 2 – Emerging	Isolated actions exist (draft guidelines or one-off workshops) without institutional consistency.
Level 3 – Established	Guidelines and practices are implemented; roles are defined; verifiable evidence and basic monitoring are in place.
Level 4 – Advanced	Mature institutional governance, continuous auditing and monitoring, periodic review, incident response, and evidence-based improvement.

For example, an institution may exhibit high levels of technology use and support but low levels of ethical literacy and traditional assessments (technology adoption = level 4 and ethical instructional design = level 2), which increases the risk of misconduct. In other cases, clear guidelines may exist, but without adequate implementation or monitoring tools, leading to delays in policy implementation or their improper use. This approach enables informed ethical risk management.

6 Discussion

Ethical governance of AI in higher education has primarily been developed based on general principles and statements promoted by international organizations and academic associations (Larsson, 2020). However, these frameworks are insufficient to address the complexity of the practical use of AI in universities, where not only regulatory criteria but also technological, pedagogical, organizational, and cultural factors come into play.

In this regard, organizations such as UNESCO, the OECD, and COPE have developed principles to promote transparency, data protection, responsible authorship, and the ethical use of AI in education and research. These frameworks have helped identify the main risks associated with the use of these technologies and propose common ethical foundations at the global level.

However, these frameworks are typically applied as general principles, and thus present at least three limitations in institutional practice:

1. Operational gap: Although transparency and data protection are recommended, the guidelines rarely define minimum standards, specific procedures, required evidence, or verifiable mechanisms for putting these principles into practice (Morley et al., 2020).

2. Disconnect from pedagogy: Ethics is presented as an external layer rather than being integrated into the teaching process—such as in task design, assessment, criteria for originality, and information literacy—to ensure the proper use of tools (Luo (Jess), 2024).

3. Context dependence: In environments where regulatory gaps, inadequate infrastructure, or heterogeneous training exist, a global principle may be insufficient to guide local decisions, as there is no contextual alignment or guidance to enable situated operationalization (Fu & Weng, 2024).

The socio-technical framework proposed in this study addresses these limitations by integrating ethical, pedagogical, technological, and institutional dimensions. Its distinction lies not in repeating international principles, but in operationalizing them through a design that specifies where to intervene and with what mechanisms. Thus, the proposal goes beyond merely stating what should be done; it offers a "how" and a "where" that can be analyzed and evaluated using indicators and evidence.

6.1 Contributions of the Socio-Technical Approach

Much of the current discourse on artificial intelligence in higher education continues to focus on two things: detecting plagiarism or AI-generated content, and determining which tools can be used. The problem is that, when the discussion stops there, AI ends up being treated solely as a matter of control. And in reality, the issue is much broader. What is at stake is not just what the tool does, but how it fits into academic culture, into teaching and assessment methods, and into the decisions made by institutions.

From this perspective, ethics is not reduced to a list of rules. Rather, it becomes a matter of governance: of how responsibilities are distributed among students, faculty, administrators, and even among the systems themselves that organize the use of these technologies. This becomes especially relevant in smart education environments, where automation, personalization, and the use of data are already part of everyday life.

One of the framework's contributions is that it helps explain why AI is used even when there are no clear rules. To this end, it integrates models of technology adoption and of risk perception and trust, which allows for a better understanding of what drives a person to adopt these tools, what they expect from them, and why they sometimes use them even in ambiguous institutional contexts. Seen this way, prohibition is not always the best solution. In many cases, it is more useful to improve guidance, strengthen training, and redesign evaluation to reduce inappropriate use and promote more responsible practices.

Another important aspect is that this proposal aligns well with the Latin American context. The region is often characterized by the rapid adoption of new technologies, regulatory gaps, and institutional realities that vary greatly from one another. For this reason, an approach that is too general or imported without adaptation often falls short. The approach proposed here acknowledges that diversity and builds upon it rather than ignoring it.

Within this framework, socio-technical governance is key because it allows us to understand that the relationship between technology and the university is not fixed. As tools change, so do practices, expectations, risks, and forms of regulation. Therefore, useful ethical governance should not be conceived as a static document, but rather as a continuous process of review, learning, and adjustment.

In short, the underlying idea is simple: the ethical governance of AI cannot be resolved through isolated measures or automatic responses. It requires a perspective capable of connecting ethical culture, real-world uses of technology, pedagogical decisions, and institutional capacity to learn and adapt.

7 Conclusions

This study proposes a socio-technical framework for ethical governance to guide the integration of AI in higher education, particularly in smart education environments. Based on a systematic literature review and the combination of various theoretical approaches, this study argues that the ethical dilemmas associated with AI cannot be understood solely through regulations or the technical characteristics of the tools. Rather, these dilemmas arise from the interaction between people, academic practices, and institutional rules, within contexts that also change over time.

Both the framework and the matrix of variables and indicators are constructed based on the literature review and the synthesis of the analyzed studies. Therefore, this proposal should be understood as a tool to guide analysis and decision-making, not as a model already validated in specific universities. Its value lies in offering a conceptual and practical foundation for thinking about the problem in a broader and more structured way.

One of the study's main contributions is to shift the ethical conversation from a focus on detection and control toward a focus on governance. This allows us to move beyond a narrow perspective centered on specific tools—such as plagiarism detectors—which often reveal little about the actual context in which AI is used. Within the proposed framework, AI is understood as a complex phenomenon where perceptions of utility, trust, ethical risks, pedagogical design, critical literacy, and institutional coordination intersect.

Methodologically, the study also provides an analytical matrix with dimensions, variables, and indicators that can be used to conduct context-specific assessments and compare levels of development across different levels: individual, institutional, and systemic. This matrix can be translated into concrete instruments, such as surveys, interviews, document analysis, or process-centered rubrics, and can also support institutional self-assessment exercises. It is not intended as a one-size-fits-all framework for all universities, but rather as a flexible tool that needs to be adapted to each context.

This proposal is particularly relevant for Latin America, where the adoption of these technologies often outpaces training and regulatory frameworks. For this reason, the framework is designed to be implemented gradually. It first proposes basic requirements, such as transparency and training, and can then be expanded through protocols, interdisciplinary committees, and mechanisms for continuous updating.

Taken together, the reviewed evidence suggests that the future of AI in higher education depends not only on how sophisticated the tools are, but also on the ability of institutions to integrate them judiciously, carefully, and with an educational focus. In other words, the challenge is not only technological; it is also ethical, pedagogical, and institutional. Even so, the framework and matrix remain a proposal derived from the literature, so they still need to be tested in specific institutions. Such empirical validation will be important for assessing their real utility, adjusting their indicators, and determining to what extent they can respond to diverse university contexts.

8 Limitations and Future Research

The study has some limitations that should be taken into account when interpreting its findings. A first limitation is that the findings are not based on original fieldwork but rather on a review of published research. While this allowed us to identify trends and gaps in the literature, it also means that the results are influenced by the quality and type of studies available during the period analyzed.

Another limitation of the study is that the search was conducted exclusively in Scopus and Web of Science. While this ensures a rigorous selection of academic literature, it also means excluding non-indexed documents—such as reports, local studies, or institutional materials—that could offer a more contextualized perspective, particularly in regions underrepresented in these databases.

Furthermore, although the proposed framework was operationalized into variables and indicative indicators, these were not empirically validated in specific institutions. Therefore, the proposal should be understood as a conceptual-operational model, capable of being adapted, adjusted, and validated in different educational contexts, levels of technological maturity, and academic disciplines.

These limitations open up various avenues for future research. A priority area for future work involves validating the framework and the indicator matrix through a phased approach:

1. Institutional pilot (1–2 universities): Apply the matrix as an initial assessment at at least one public and one private institution, covering teaching, assessment, research, and administration.
2. Operationalization by indicator. Define, for each indicator, the most appropriate data collection method (survey, interview, checklist, or document audit) and minimum compliance criteria.
3. Content validity. Submit the indicators to expert review (ethics, education, AI, IT, university management) and estimate indices such as Aiken V or CVI.
4. Reliability. If Likert-type scales are constructed, assess internal consistency (α/ω) and, where feasible, stability (test–retest).
5. Construct validity and fit. Explore the structure of the instrument (EFA/CFA or another strategy appropriate to the design) and eliminate redundant or poorly discriminating indicators.
6. Stakeholder feedback. Incorporate feedback from teachers, students, and administrators to refine the clarity, feasibility, and usefulness of the assessment.

Finally, future research could explore comparative analyses across different types of institutions (public and private), educational modalities (in-person, hybrid, and online), and geographic regions, with the aim of identifying patterns of socio-technical co-evolution and evaluating the framework's adaptability in diverse contexts. Together, these lines of inquiry will advance toward a deeper and more applied understanding of the ethical governance of AI in higher education.

Acknowledgments

Declaration of AI Use
AI tools were used to assist with drafting and language editing of the manuscript and to translate the text into English. No AI tool was used to generate data, perform analyses, or interpret results. All content was reviewed, verified, and approved by the authors, who take full responsibility for the manuscript. No AI system is listed as an author.

References

Acosta-Enriquez, B. G., Arbulú Ballesteros, M. A., Arbulu Perez Vargas, C. G., Orellana Ulloa, M. N., Gutiérrez Ulloa, C. R., Pizarro Romero, J. M., Gutiérrez Jaramillo, N. D., Cuenca Orellana, H. U., Ayala Anzoátegui, D. X., & López Roca, C. (2024). Knowledge, attitudes, and perceived ethics regarding the use of ChatGPT among Generation Z university students. *International Journal for Educational Integrity, 20*, Article 10. https://doi.org/10.1007/s40979-024-00157-4

Alanzi, T., Alanazi, F., Mashhour, B., Altalhi, R., Alghamdi, A., Al Shubbar, M., Alamro, S., Alshammari, M., Almusmili, L., Alanazi, L., Alzahrani, S., Alalouni, R., Alanzi, N., & Alsharifa, A. (2023). Surveying hematologists' perceptions and readiness to embrace artificial intelligence in diagnosis and treatment decision-making. *Cureus, 15*(11), Article e49462. https://doi.org/10.7759/cureus.49462

Alhwaiti, M. (2023). Acceptance of artificial intelligence application in the post-covid era and its impact on faculty members' occupational well-being and teaching self efficacy: A path analysis using the UTAUT 2 model. *Applied Artificial Intelligence, 37*(1), 2175110. https://doi.org/10.1080/08839514.2023.2175110

Alonso-Prieto, V., Dimitriadis, Y., Villagrá-Sobrino, S. L., Ortega-Arranz, A., Topali, P., & Martínez-Monés, A. (2025). Exploring how teacher agency unfolds within the co-design of a smart learning environment-supported learning activity: A case study. *Journal of Information Technology Education: Research, 24*, Article 34. https://doi.org/10.28945/5615

Al-Zahrani, A. M. (2024). Unveiling the shadows: Beyond the hype of AI in education. *Heliyon, 10*(9), e30696. https://doi.org/10.1016/j.heliyon.2024.e30696

Ayoubi, K. (2024). Adopting ChatGPT: Pioneering a new era in learning platforms. *International Journal of Data and Network Science, 8*(2), 1341–1348. https://doi.org/10.5267/j.ijdns.2023.11.001

Bouteraa, M., Bin-Nashwan, S. A., Al-Daihani, M., Dirie, K. A., Benlahcene, A., Sadallah, M., Zaki, H. O., Lada, S., Ansar, R., Fook, L. M., & Chekima, B. (2024). Understanding the diffusion of AI-generative (ChatGPT) in higher education: Does students' integrity matter? *Computers in Human Behavior Reports, 14*, 100402. https://doi.org/10.1016/j.chbr.2024.100402

Cedeño Meza, J. G., Maitta Rosado, I. S., Vélez Zambrano, M. L., & Palomeque Zambrano, J. Y. (2024). Investigación universitaria con inteligencia artificial. *Revista Venezolana de Gerencia, 29*(106), 817–830. https://doi.org/10.52080/rvgluz.29.106.23

Chai, C. S., Yu, D., King, R. B., & Zhou, Y. (2024). Development and validation of the Artificial Intelligence Learning Intention Scale (AILIS) for university students. *Sage Open, 14*(2), 21582440241242188. https://doi.org/10.1177/21582440241242188

Committee on Publication Ethics. (2023, February 13). *Authorship and AI tools*. https://doi.org/10.24318/cCVRZBms

Dallari, V., Liberale, C., De Cecco, F., Nocini, R., Arietti, V., Monzani, D., & Sacchetto, L. (2024). The role of artificial intelligence in training ENT residents: A survey on ChatGPT, a new method of investigation. *Acta Otorhinolaryngologica Italica, 44*(3), 161–168. https://doi.org/10.14639/0392-100X-N2806

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly, 13*(3), 319–340. https://doi.org/10.2307/249008

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

Gandhi, A., & Gani, P. H. (2024). Would lecturers use AI-based software to write scientific article? A quantitative approach in Indonesia. *Ingénierie Des Systèmes d Information, 29*(3), 941–950. https://doi.org/10.18280/isi.290314

Grájeda, A., Burgos, J., Córdova, P., & Sanjinés, A. (2024). Assessing student-perceived impact of using artificial intelligence tools: Construction of a synthetic index of application in higher education. *Cogent Education, 11*(1), 2287917. https://doi.org/10.1080/2331186X.2023.2287917

Haddaway, N. R., Page, M. J., Pritchard, C. C., & McGuinness, L. A. (2022). PRISMA2020: An R package and Shiny app for producing PRISMA 2020-compliant flow diagrams, with interactivity for optimised digital transparency and Open Synthesis. *Campbell Systematic Reviews, 18*(2), e1230. https://doi.org/10.1002/cl2.1230

Ivanov, S., Soliman, M., Tuomi, A., Alkathiri, N. A., & Al-Alawi, A. N. (2024). Drivers of generative AI adoption in higher education through the lens of the Theory of Planned Behaviour. *Technology in Society, 77*, 102521. https://doi.org/10.1016/j.techsoc.2024.102521

Jo, H., & Bang, Y. (2023). Analyzing ChatGPT adoption drivers with the TOEK framework. *Scientific Reports, 13*(1), 22606. https://doi.org/10.1038/s41598-023-49710-0

Jordano, P. (n.d.). *Coevolución: Patrones y procesos* [Course notes]. Universidad de Sevilla–CSIC. http://pjordanolab.ebd.csic.es/pdfs/coevol.pdf

Larsson, S. (2020). On the Governance of Artificial Intelligence through Ethics Guidelines. *Asian Journal of Law and Society, 7*(3), 437–451. https://doi.org/10.1017/als.2020.19

Li, Q., & Qin, Y. (2023). AI in medical education: Medical student perception, curriculum recommendations and design suggestions. *BMC Medical Education, 23*(1), 852. https://doi.org/10.1186/s12909-023-04700-8

Luo, J. (2024). A critical review of GenAI policies in higher education assessment: A call to reconsider the "originality" of students' work. *Assessment & Evaluation in Higher Education, 49*(5), 651–664. https://doi.org/10.1080/02602938.2024.2309963

Mahmud, A., Sarower, A. H., Sohel, A., Assaduzzaman, M., & Bhuiyan, T. (2024). Adoption of ChatGPT by university students for academic purposes: Partial least square, artificial neural network, deep neural network and classification algorithms approach. *Array, 21*, 100339. https://doi.org/10.1016/j.array.2024.100339

Mohd Rahim, N. I., Iahad, N. A., Yusof, A. F., & Al-Sharafi, M. A. (2022). AI-based chatbots adoption model for higher-education institutions: A hybrid PLS-SEM-neural network modelling approach. *Sustainability, 14*(19), 12726. https://doi.org/10.3390/su141912726

Morley, J., Floridi, L., Kinsey, L., & Elhalal, A. (2020). From What to How: An Initial Review of Publicly Available AI Ethics Tools, Methods and Research to Translate Principles into Practices. *Science and Engineering Ethics, 26*(4), 2141–2168. https://doi.org/10.1007/s11948-019-00165-5

Nemt-allah, M., Khalifa, W., Badawy, M., Elbably, Y., & Ibrahim, A. (2024). Validating the ChatGPT usage scale: psychometric properties and factor structures among postgraduate students. *BMC Psychology, 12*(1), 497. https://doi.org/10.1186/s40359-024-01983-4

Niloy, A. C., Bari, M. A., Sultana, J., Chowdhury, R., Raisa, F. M., Islam, A., Mahmud, S., Jahan, I., Sarkar, M., Akter, S., Nishat, N., Afroz, M., Sen, A., Islam, T., Tareq, M. H., & Hossen, M. A. (2024). Why do students use ChatGPT? Answering through a triangulation approach. *Computers and Education: Artificial Intelligence, 6*, 100208. https://doi.org/10.1016/j.caeai.2024.100208

Organisation for Economic Co-operation and Development. (2019). *Recommendation of the Council on Artificial Intelligence* (OECD/LEGAL/0449; amended 2023). https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449

Okulich-Kazarin, V., Artyukhov, A., Skowron, Ł., Artyukhova, N., & Wołowiec, T. (2024). Will AI become a threat to higher education sustainability? A study of students' views. *Sustainability, 16*(11), 4596. https://doi.org/10.3390/su16114596

Oyama, K. (1986). La coevolución. *Ciencias*, (Esp), 64–73. https://www.revistacienciasunam.com/images/stories/Articles/ESP1/CNSE0109.pdf

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). Declaración PRISMA 2020: Una guía

actualizada para la publicación de revisiones sistemáticas. *Revista Española de Cardiología, 74*(9), 790–799. https://doi.org/10.1016/j.recesp.2021.06.016

Polyportis, A., & Pahos, N. (2025). Understanding students' adoption of the ChatGPT chatbot in higher education: The role of anthropomorphism, trust, design novelty and institutional policy. *Behaviour & Information Technology, 44*(2), 315–336. https://doi.org/10.1080/0144929X.2024.2317364

Salazar Vargas, C. (2009). La evaluación y el análisis de políticas públicas. *Revista Opera*, (9), 23–51.

Sallam, M., Salim, N. A., Barakat, M., Al-Mahzoum, K., Al-Tammemi, A. B., Malaeb, D., Hallit, R., & Hallit, S. (2023). Assessing health students' attitudes and usage of ChatGPT in Jordan: validation study. *JMIR Medical Education*, *9*, e48254. https://doi.org/10.2196/48254

Sobaih, A. E. E. (2024). Ethical concerns for using artificial intelligence chatbots in research and publication: Evidences from Saudi Arabia. *Journal of Applied Learning & Teaching, 7*(1), 93–103. https://doi.org/10.37074/jalt.2024.7.1.21

Thomas, R., Bhosale, U., Shukla, K., & Kapadia, A. (2023). Impact and perceived value of the revolutionary advent of artificial intelligence in research and publishing among researchers: A survey-based descriptive study. *Science Editing, 10*(1), 27–34. https://doi.org/10.6087/kcse.294

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

UNESCO. (2025, September 2). *UNESCO survey: Two-thirds of higher education institutions have or are developing guidance on AI use*. https://www.unesco.org/en/articles/unesco-survey-two-thirds-higher-education-institutions-have-or-are-developing-guidance-ai-use

Venkatesh, V., Thong, J. Y. L., & Xu, X. (2012). Consumer acceptance and use of information technology: Extending the unified theory of acceptance and use of technology. *MIS Quarterly, 36*(1), 157–178. https://doi.org/10.2307/41410412

Zhu, J., & Liu, W. (2020). A tale of two databases: The use of Web of Science and Scopus in academic papers. *Scientometrics, 123*(1), 321–335. https://doi.org/10.1007/s11192-020-03387-8