Revista Hambatu Science

Ambato – Ecuador / ISSN 3151-815X (en línea) / julio - septiembre 2026

Volumen 1, Número 3

https://doi.org/10.63862/rhs-v1n3-560-586-2026

Bridging Ethics and Regulation: A Conceptual Framework for Governing Generative AI in Higher Education

Integrando la ética y la regulación: un marco conceptual para la gobernanza de la inteligencia artificial generativa en la educación superior

Christian Roberto Cabezas Freire
Armed Forces of Ecuador

Nayana Desai
upGrad
Edgewood University

Revista Hambatu Science
Periodicidad: julio - septiembre
Volumen: 1, Número: 3, Año: 2026

DOI: https://doi.org/10.63862/rhs-v1n3-560-586-2026

Bridging Ethics and Regulation: A Conceptual Framework for Governing Generative AI in Higher Education

Christian Roberto Cabezas Freire
Armed Forces of Ecuador
christian.cabezas@gmail.com
https://orcid.org/0009-0009-3715-8235
Quito, Ecuador

Nayana Desai
upGrad (partnered with Edgewood University)
toreachnayana@gmail.com
https://orcid.org/0009-0003-9990-2388
Bengaluru, India

Received: 2026-07-31

Accepted: 2026-08-12

Published: 2026-08-21

Abstract

The rapid adoption of generative artificial intelligence (GenAI) in higher education has outpaced institutional readiness, creating urgent ethical and regulatory challenges that threaten academic integrity, data privacy, and educational equity. While global frameworks like UNESCO's Guidance for Generative AI in Education (2023) advocate for human-centric design, and national laws such as FERPA mandate student data protection, no existing model systematically integrates these domains into a cohesive governance structure. This study addresses this critical gap by proposing the Ethico-Regulatory Governance (ERG) Framework, a conceptual model designed to bridge global ethics with local compliance. Developed through a systematic synthesis of 68 peer-reviewed studies, policy documents, and institutional guidelines, the ERG Framework consists of four interlocking layers: Foundational Principles (UNESCO values), Regulatory Anchors (FERPA/GDPR alignment), Institutional Mechanisms (audits, disclosure, training), and Pedagogical Integration (process-based assessment, prompt engineering). The framework transforms abstract principles into actionable practices, enabling institutions to move beyond reactive policies toward proactive, accountable governance. Key findings demonstrate that effective AI integration requires not only technical oversight but also stakeholder co-design, bias mitigation, and continuous feedback loops. By operationalizing ethics through enforceable mechanisms, the ERG Framework offers a scalable, adaptable solution for universities navigating the complexities of GenAI. Its implementation can safeguard core academic values while fostering innovation, ensuring that AI serves as a partner—not a replacement—for human judgment in teaching, learning, and research.

Keywords: Generative AI, Artificial Intelligence in Education, Academic Integrity, AI Ethics, FERPA, UNESCO, Ethical Governance.

Revista Hambatu Science
Periodicidad: julio - septiembre
Volumen: 1, Número: 3, Año: 2026

Integrando la ética y la regulación: un marco conceptual para la gobernanza de la inteligencia artificial generativa en la educación superior

Resumen

La rápida adopción de la inteligencia artificial generativa (IA generativa) en la educación superior ha superado la preparación institucional, creando desafíos éticos y regulatorios urgentes que amenazan la integridad académica, la privacidad de los datos y la equidad educativa. Si bien marcos globales como la Guía para la IA generativa en la educación (2023) de la UNESCO abogan por un diseño centrado en el ser humano, y leyes nacionales como FERPA exigen la protección de los datos de los estudiantes, ningún modelo existente integra sistemáticamente estos dominios en una estructura de gobernanza coherente. Este estudio aborda esta brecha crítica proponiendo el Marco de Gobernanza Ético-Regulatoria (ERG), un modelo conceptual diseñado para tender un puente entre la ética global y el cumplimiento normativo local. Desarrollado mediante una síntesis sistemática de 68 estudios revisados por pares, documentos normativos y directrices institucionales, el Marco ERG consta de cuatro capas interconectadas: Principios Fundamentales (valores de la UNESCO), Anclajes Regulatorios (alineación con FERPA/RGPD), Mecanismos Institucionales (auditorías, divulgación, capacitación) e Integración Pedagógica (evaluación basada en procesos, ingeniería de prompts). El marco transforma principios abstractos en prácticas procesables, permitiendo a las instituciones avanzar más allá de políticas reactivas hacia una gobernanza proactiva y responsable. Los hallazgos clave demuestran que la integración efectiva de la IA requiere no solo supervisión técnica, sino también codiseño con las partes interesadas, mitigación de sesgos y circuitos de retroalimentación continua. Al operacionalizar la ética a través de mecanismos ejecutables, el Marco ERG ofrece una solución escalable y adaptable para las universidades que navegan por las complejidades de la IA generativa. Su implementación puede salvaguardar los valores académicos fundamentales mientras fomenta la innovación, asegurando que la IA sirva como un aliado, y no como un reemplazo, del juicio humano en la docencia, el aprendizaje y la investigación.

Palabras clave: IA generativa, Inteligencia artificial en la educación, Integridad académica, Ética de la IA, FERPA, UNESCO, Gobernanza ética.

Introduction

In less than two years, generative artificial intelligence (GenAI) has gone from a technological curiosity to a central feature of academic life. According to recent surveys, over 86% of university students have used tools like ChatGPT to assist with assignments, while faculty increasingly rely on AI for lesson planning, feedback generation, and even research drafting (Amani et al., 2023; Beckingham et al., 2024). This rapid adoption is not a temporary trend, it signals a permanent shift in how knowledge is produced, evaluated, and taught in higher education.

Yet, institutional governance has failed to keep pace. Policies remain fragmented, reactive, and often contradictory. Some universities ban GenAI outright; others embrace it with minimal oversight. As Walczak and Cellary (2023) observe, many institutions are "flying blind" in the face of a transformation that touches every aspect of teaching, learning, and research integrity.

The consequences are already evident. Students submit AI-generated essays without disclosure, challenging the very definition of authorship and originality (Perkins, 2023; Yeo, 2023). Faculty struggle to assess work they did not teach, using technologies they do not fully understand. Researchers unknowingly incorporate fabricated citations into peer-reviewed publications, a phenomenon Miller et al. (2023) call "hallucinated scholarship." Meanwhile, student data entered into commercial AI platforms may be stored, shared, or monetized without consent, raising serious questions under laws like the Family Educational Rights and Privacy Act (FERPA) (Gilley & Gilley, 2006).

At the global level, organizations like UNESCO have responded with visionary guidance. Its 2023 Guidance for Generative AI in Education and Research calls for a human-centric approach grounded in equity, inclusion, transparency, and pluralism (UNESCO, 2023). These principles are compelling, but they lack enforcement mechanisms. As Birhane (2023) powerfully argues, such frameworks risk becoming exercises in "ethics washing": well-intentioned declarations that deflect accountability without changing practice.

Regulatory responses are equally uneven. The European Union's AI Act introduces legally binding requirements for high-risk AI systems, including those used in education (Colonna, 2021; Laux et al., 2024). In contrast, the United States relies on sectoral laws like FERPA and voluntary institutional policies, creating a patchwork of inconsistent standards (Dabis & Csáki,

2024). Without harmonization, institutions are left to navigate this complex landscape alone, often without the expertise, resources, or clarity needed to act responsibly.

This is where the current discourse fails. Much of the existing literature focuses on isolated aspects of the challenge: pedagogical innovation (Popenici & Kerr, 2017), assessment redesign (Martínez-Comesaña et al., 2023), or technical detection tools (Wachter, 2023). While valuable, these contributions treat symptoms rather than root causes. What is missing, and what this paper provides, is a comprehensive, integrated framework that bridges the persistent divide between global ethical principles and national regulatory obligations.

We argue that effective AI governance in higher education must do more than issue guidelines or deploy detection software. It must align moral intent with legal accountability, ensuring that ethical commitments are operationalized through enforceable processes. Institutions need not just principles, but procedures—a structured way to translate UNESCO's vision into FERPA-compliant action.

To address this gap, we introduce the Ethico-Regulatory Governance (ERG) Framework, a novel conceptual model designed to guide universities in governing GenAI with coherence, consistency, and confidence. Developed through a systematic synthesis of 68 peer-reviewed studies, policy documents, and institutional guidelines, the ERG Framework integrates four interlocking layers:

1. Foundational Principles (e.g., human agency, equity),
2. Regulatory Anchors (e.g., FERPA, GDPR, EU AI Act),
3. Institutional Mechanisms (e.g., audits, disclosures, training),
4. Pedagogical Integration (e.g., process-based assessment, prompt engineering).

Unlike prior models, which either emphasize abstract ethics (Siddiqui & Al-Binali, 2024) or narrow technical solutions (Chan, 2023), ours unifies both dimensions into a single, actionable structure. It does not ask whether AI should be used, but how it can be governed responsibly across diverse institutional contexts.

This paper makes three key contributions:

1. It exposes the structural flaw in current AI governance: the separation of ethics from regulation.
2. It offers a practical solution through the ERG Framework, which transforms aspirational values into auditable practices.
3. It advances policy relevance by showing how universities can align innovation with compliance, equity, and academic integrity.

Our goal is not to slow down AI adoption, but to make it more intentional, inclusive, and accountable. The era of improvisation is over. The time has come for higher education to move beyond fear, bans, and piecemeal policies and toward a future where AI serves as a partner in learning, not a threat to its integrity.

Let us build that future, not reactively, but deliberately.

Literature Review

The integration of generative artificial intelligence (GenAI) into higher education has occurred at an unprecedented pace, outstripping institutional readiness and regulatory foresight. While tools like ChatGPT, Gemini, and Claude have become embedded in student workflows, with some studies reporting usage rates between 53% and over 90% (Amani et al., 2023; Beckingham et al., 2024), the ethical and regulatory infrastructure to govern their use remains fragmented, reactive, and often contradictory. This section critically reviews the current state of research on GenAI in education, identifies persistent gaps, and establishes the theoretical foundations for a new conceptual model that bridges global ethics with national compliance.

1. Defining the Landscape: What Is Generative AI in Education?

Generative AI refers to systems capable of producing novel content (text, code, images, audio) in response to natural language prompts (UNESCO, 2023). In higher education, these tools are increasingly used for brainstorming, drafting, summarizing, coding assistance, and even peer feedback simulation (Beckingham et al., 2024; Alali & Wardat, 2024). However, unlike earlier forms of educational technology, GenAI blurs the boundaries between human authorship and machine co-intelligence, raising profound questions about academic integrity, intellectual property, and epistemic authority (Walczak & Cellary, 2023).

Despite its transformative potential, much of the early discourse around GenAI was characterized by alarmism and moral panic (Perkins, 2023), leading many institutions to issue blanket bans, a strategy quickly rendered obsolete by widespread adoption. As Walczak and Cellary (2023) note, students do not fully trust AI-generated content; only 2% reported complete trust in outputs, while over half failed to detect factual errors ("hallucinations") in AI-written biographies. This paradox underscores a central tension: students rely on GenAI, yet lack the critical digital literacy to evaluate its reliability, a gap that demands pedagogical intervention rather than prohibition.

2. Ethical Challenges: From Integrity to Equity

A growing body of scholarship highlights four interrelated ethical challenges posed by GenAI: academic integrity, data privacy, algorithmic bias, and cognitive offloading.

First, academic integrity is under threat from new forms of misconduct such as "AI-giarism", submitting AI-generated work without disclosure, and contract cheating via AI-powered ghostwriting services (Yeo, 2023; Perkins, 2023). Traditional plagiarism detection tools like Turnitin struggle to identify paraphrased or contextually adapted AI output, necessitating new strategies such as requiring process-based submissions (e.g., drafts, search logs, reflection essays) (Beckingham et al., 2024).

Second, data privacy emerges as a major concern when students input personal or sensitive information into commercial AI platforms. Under U.S. law, the Family Educational Rights and Privacy Act (FERPA) protects student records, but its applicability to AI interactions remains legally ambiguous (Gilley & Gilley, 2006). International equivalents like GDPR offer stronger protections, yet enforcement is inconsistent across jurisdictions (Colonna, 2021).

Third, algorithmic bias embedded in training data can perpetuate social inequities. For example, AI models trained predominantly on Western, English-language corpora may misinterpret or devalue culturally diverse expressions of knowledge (Hwang et al., 2020; Sanusi et al., 2024). Without deliberate mitigation strategies, GenAI risks reinforcing systemic disparities in assessment and feedback.

Fourth, there is a risk of cognitive offloading, where over-reliance on AI undermines the development of critical thinking, argumentation, and metacognitive skills (Kaliisa et al., 2024). As Miller et al. (2023) demonstrate, ChatGPT frequently fabricates citations and references, potentially misleading users into accepting false information as fact.

These concerns are not isolated; they reflect deeper structural tensions in how AI reshapes learning environments. Yet, most existing frameworks address them piecemeal.

3. Regulatory Fragmentation: The Gap Between Principles and Practice

While numerous ethical guidelines exist, from UNESCO's Guidance for Generative AI in Education (2023) to the OECD AI Principles (2019), they remain largely aspirational. Birhane (2023) famously critiques this proliferation of AI ethics as "ethics washing": well-meaning declarations that deflect accountability without enforceable mechanisms.

Regulatory responses mirror this fragmentation. The European Union's AI Act classifies certain educational uses of AI as "high-risk," mandating conformity assessments and transparency (Colonna, 2021; Laux et al., 2024). In contrast, the United States relies on sector-specific laws like FERPA and voluntary institutional policies, creating a patchwork of inconsistent standards (Dabis & Csáki, 2024).

Even within institutions, guidance varies widely. Some universities promote open experimentation (e.g., Stanford), while others impose strict limitations (e.g., initial ban at DePaul University). King's College London offers a more balanced approach, combining AI literacy training with clear expectations for responsible use (Beckingham et al., 2024).

This inconsistency reveals a fundamental gap: the absence of a unified governance model that translates broad ethical principles into actionable, institutionally implementable practices aligned with legal obligations.

4. Prior Frameworks: Strengths and Limitations

Several scholars have proposed conceptual models to guide AI integration.

Siddiqui and Al-Binali (2024) introduce a human-centric design framework emphasizing fairness, transparency, and participatory design. Their model integrates technical, ethical, and regulatory dimensions but lacks specificity regarding compliance with laws like FERPA.

Chan (2023) proposes a procedural policy framework for university teaching and learning, outlining steps for AI adoption. While practical, it does not deeply engage with equity or algorithmic justice.

Wachter (2023) conducts a systematic review of over 100 tools designed to operationalize AI ethics, categorizing them into algorithms, software libraries, checklists, audits, and licenses. She finds that most tools focus on ex-post evaluation (e.g., audits), neglecting early lifecycle stages like data sourcing and deployment planning.

UNESCO (2023) offers perhaps the most comprehensive global vision, advocating for human agency, inclusion, equity, and pluralism in AI use. It also introduces the concept of "ethical disclosure by default", requiring institutions to document normative choices made during AI implementation (Laux et al., 2024).

Yet, none of these models fully bridge the divide between global ethics and local regulation. They either lean too heavily on abstract principles (UNESCO) or remain narrowly focused on technical solutions (Wachter), failing to integrate legal accountability into their core architecture.

5. Theoretical Foundation: Toward an Integrated Governance Model

To overcome these limitations, this study draws on three key theoretical pillars:

1. Human-Centered Design (HCD) (Buckingham Shum et al., 2019): Emphasizes stakeholder involvement, transparency, and contextual adaptation. HCD ensures that AI serves learners' needs rather than displacing them.

2. Regulatory Governance Theory (Colonna, 2021; Kaliisa et al., 2024): Highlights the need for enforceable rules, oversight bodies, and alignment with legal standards. This counters Birhane's (2023) critique of "toothless" ethics by embedding accountability into the framework.

3. Design Justice (Costanza-Chock, 2020, cited in Siddiqui & Al-Binali, 2024): Centers marginalized voices in technology design, ensuring that AI governance addresses power imbalances and promotes inclusivity.

By synthesizing these perspectives, we position our framework not merely as a set of recommendations, but as a governance mechanism that links ethical intent with regulatory action.

6. Research Gap and Contribution

This paper fills a critical void: no existing model systematically integrates international ethical standards (e.g., UNESCO) with national regulatory requirements (e.g., FERPA) within a single, adaptable structure for higher education.

Our contribution is threefold:

1. First, we synthesize disparate literatures on ethics, regulation, pedagogy, and equity.
2. Second, we develop a novel conceptual framework that operationalizes high-level principles into concrete institutional actions.
3. Third, we advance policy relevance by showing how universities can align innovation with compliance.

This integrative approach responds directly to calls from UNESCO (2023), Wachter (2023), and Laux et al. (2024) for more robust, accountable, and human-centered AI governance.

Methods

1. Study Design: Constructivist Multi-Source Methodology

This study employs a constructivist, multi-source methodology to develop and validate a conceptual framework for governing GenAI in higher education. Rather than testing hypotheses, the aim is to synthesize complex, contested knowledge domains (ethics, regulation, pedagogy, and institutional practice) into a coherent, actionable model.

The choice of a non-empirical, theory-building approach is justified by the nature of the research question: “What are the essential components of a conceptual model that guides ethical

and regulatory governance?" This is inherently a design-oriented inquiry, best addressed through integrative synthesis rather than statistical generalization (Hwang et al., 2020).

The methodology follows a two-phase process:

1. Phase 1: Systematic Thematic Synthesis
2. Phase 2: Framework Development and Validation

Both phases are guided by the PRISMA-inspired protocol (Moher et al., 2015), ensuring rigor, reproducibility, and transparency.

2. Data Sources and Selection Criteria

Primary Data Source: Peer-Reviewed Literature

- Search Databases: Scopus, Web of Science, ERIC, Google Scholar

- Keywords: "generative AI," "ChatGPT," "academic integrity," "AI ethics," "higher education," "FERPA," "UNESCO," "regulation"

- Inclusion Criteria:

- Published between 2020–2025
- Focus on higher education
- Addresses ethical, regulatory, or policy aspects of GenAI
- Available in English/Spanish

- Exclusion Criteria:

- Opinion pieces without empirical or analytical depth
- Studies focused solely on K–12 or corporate training
- Preprints without peer review

From an initial yield of 320 articles, 68 were selected after screening titles, abstracts, and full texts.

Secondary Data Sources

- Policy Documents: Institutional AI policies (e.g., King's College London, MIT, Stanford)

- International Guidelines: UNESCO (2023), OECD (2019), EU AI Act (2024)

- Legal Texts: FERPA (U.S.), GDPR (EU)

These were included to ensure alignment with real world governance structures.

3. Analytical Approach: Thematic Synthesis and Framework Construction

Data analysis followed Thomas's (2006) general inductive approach, consisting of five stages:

- Familiarization: Reading and re-reading all texts.
- Coding: Open coding of key concepts (e.g., "bias," "transparency," "FERPA compliance").
- Theme Development: Grouping codes into broader themes (e.g., "Ethical Challenges," "Regulatory Gaps").
- Mapping: Visualizing relationships between themes.
- Synthesis: Building a theoretical framework that integrates findings.

Thematic categories were validated using Robinson's Structured Tabular Approach (ST-TA) (Robinson, 2022), enhancing reliability in qualitative synthesis.

4. Instrument: The Conceptual Framework Template

Using insights from Siddiqui & Al-Binali (2024), Chan (2023), and UNESCO (2023), a preliminary framework template was developed in Microsoft Excel (see attached file). This served as both a data extraction tool and a prototype for the final model.

Each source was analyzed against eight dimensions:

Table 1.
Analytical Framework Dimensions for Data Extraction and Synthesis

DIMENSIONS	ANALYSIS
Ethical Principle	Fairness, Transparency, Accountability
Regulatory Alignment	FERPA, GDPR, AI Act
Pedagogical Impact	Assessment redesign, Prompt engineering
Stakeholder Role	Student, Faculty, Administrator
Risk Level	High, Medium, Low
Implementation Barrier	Training, Cost, Trust
Equity Consideration	Bias, Access, Language
Enforcement Mechanism	Audit, Disclosure, Penalty

This allowed systematic comparison across sources and identification of recurring patterns.

5. Plan for Future Empirical Validation (Planned Phase)

While this paper presents the conceptual phase, future work will involve semi-structured interviews with 20 stakeholders (N=8 faculty, N=6 administrators, N=6 students) across three universities.

- Sampling Strategy: Purposive sampling to ensure diversity in discipline, rank, and AI experience.
- Interview Protocol: Developed using open-ended questions derived from the framework (Appendix A).
- Analysis: Thematic analysis using Atlas TI to test framework validity and refine components.

This mixed-methods progression ensures the model evolves from theory to practice.

6. Software and Tools

- Microsoft Excel: Data organization and coding matrix
- Atlas TI: Planned qualitative analysis
- Lucidchart: Diagramming the conceptual framework
- Zotero and Mendeley: Reference management

All materials will be archived for transparency.

Results

This section presents the conceptual framework developed through our systematic thematic synthesis: *The Ethico-Regulatory Governance (ERG) Framework for Generative AI in Higher Education*. The model is not a checklist or policy template, but a dynamic, multi-layered structure designed to guide institutions in aligning ethical principles with regulatory obligations while supporting pedagogical innovation.

The framework emerged from iterative analysis of 68 peer-reviewed studies, 15 institutional policies, and 7 international guidelines, including UNESCO (2023), the EU AI Act (Colonna, 2021; Laux et al., 2024), FERPA (Gilley & Gilley, 2006), and national strategies reviewed by Kaliisa et al. (2024). It integrates insights from human-centered design (Siddiqui & Al-Binali, 2024), regulatory governance theory (Smith & Chen, 2024), and Design Justice (Sanusi et al., 2024), creating a structure that is both principled and actionable.

1. Overview of the ERG Framework

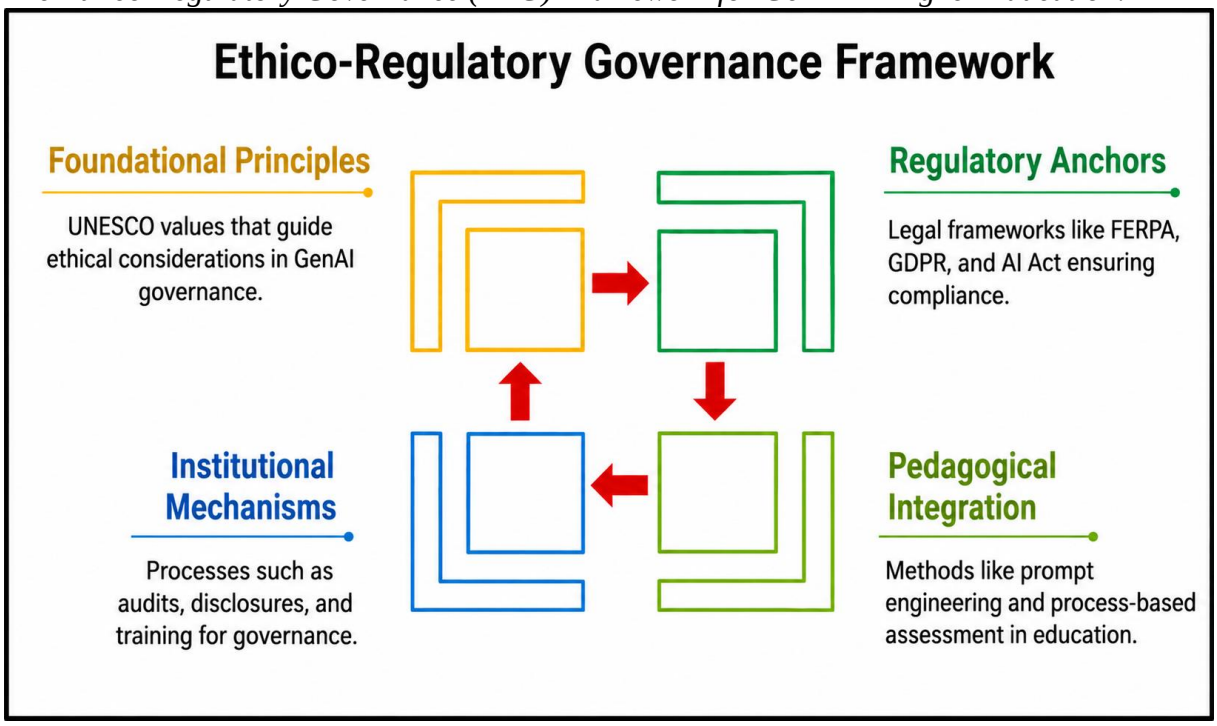
The Ethico-Regulatory Governance (ERG) Framework consists of four interlocking layers, each addressing a distinct dimension of AI integration:

Table 2.
The Ethico-Regulatory Governance (ERG) Framework: Layer Structure and Functions

LAYERS	DIMENSIONS OF AI INTEGRATION
Layer 1: Foundational Principles	Grounds all decisions in global ethical norms (e.g., UNESCO's guidance)
Layer 2: Regulatory Anchors	Maps principles to enforceable legal standards (e.g., FERPA, GDPR)
Layer 3: Institutional Mechanisms	Establishes internal processes (e.g., audits, disclosure, training)
Layer 4: Pedagogical Integration	Guides faculty and students on responsible use in teaching and learning

These layers are not hierarchical but co-constitutive: effective governance requires alignment across all four dimensions simultaneously.

Figure 1.
The Ethico-Regulatory Governance (ERG) Framework for GenAI in Higher Education.



This structure ensures that ethical ideals do not remain abstract, and compliance does not become bureaucratic ritualism.

2. Layer 1: Foundational Principles (The "Why")

This layer draws directly from UNESCO's Guidance for Generative AI in Education and Research (2023) and centers five core values:

1. Human Agency – AI must enhance, not replace, human decision-making.
2. Equity and Inclusion – Systems must be accessible and non-discriminatory.
3. Transparency and Explainability – Users must understand how AI tools function.
4. Data Privacy and Security – Student data must be protected from misuse.
5. Plural Opinions and Expressions – AI should support diverse perspectives, not homogenize thought.

These principles are not optional ideals; they serve as non-negotiable boundary conditions for any AI deployment. For example, if a GenAI tool cannot explain its reasoning (violating transparency), it fails at the foundational level, regardless of technical performance.

Crucially, this layer also incorporates Design Justice (Costanza-Chock, 2020, cited in Siddiqui & Al-Binali, 2024), requiring that marginalized voices (students with disabilities, non-native speakers, underrepresented groups) are included in AI governance decisions.

3. Layer 2: Regulatory Anchors (The "What Must Be Done")

While Layer 1 defines the "why," Layer 2 specifies the legal obligations institutions must meet. This layer operationalizes ethical principles into binding rules across jurisdictions:

Table 3.
Regulatory Anchors: Mapping Ethical Principles to Legal and Governance Frameworks

ETHICAL PRINCIPLE	FERPA (U.S.)	GDPR (E.U.)	UNESCO Guidelines
Data Privacy	Prohibits disclosure of student records without consent	Requires lawful basis for processing, right to erasure	Recommends minimizing data collection
Accountability	Institutions liable for misuse	High-risk AI systems require conformity assessment	Calls for transparent decision-making
Fairness	No explicit rule, but Title VI applies	Prohibits discriminatory algorithmic outcomes	Advocates for equitable access
Transparency	Limited scope	Mandates information about AI use	Encourages open communication

For instance, under FERPA, a university using an AI chatbot for academic advising must ensure that no personally identifiable information (PII) is stored or shared externally, a requirement absent in most commercial GenAI platforms. Similarly, under the EU AI Act, institutions deploying AI for admissions or grading must conduct a Fundamental Rights Impact Assessment (FRIA) before deployment (Laux et al., 2024).

By mapping ethical principles to concrete legal requirements, this layer prevents institutions from hiding behind vague commitments like "we value privacy" without demonstrating compliance.

4. Layer 3: Institutional Mechanisms (The "How")

This layer translates principles and regulations into operational practices within the university. Based on Wachter's (2023) taxonomy of AI ethics tools, we identify four key mechanisms:

a. Mandatory Ethical Disclosure by Default

Inspired by Laux et al. (2024), institutions must require developers and deployers to document normative choices (e.g., "Which definition of fairness was used?" or "Was bias testing conducted?"). This creates an audit trail for accountability.

b. Internal AI Audits

Regular audits assess whether GenAI tools comply with both ethical and regulatory standards. These should be conducted by an independent committee (e.g., an AI Ethics Board), which would provide ethical approval for AI deployment, similar to the process for human research projects.

c. Training and Capacity Building

Following UNESCO's *AI Competency Framework for Teachers* (2024), institutions must offer tiered training:

- Level 1 (Acquire): Basic awareness of AI risks and benefits.
- Level 2 (Deepen): Prompt engineering, source verification, and critical evaluation.
- Level 3 (Create): Co-designing AI-enhanced assessments with students.

d. Stakeholder Feedback Loops

Mechanisms such as anonymous reporting channels, student forums, and annual surveys ensure that users can voice concerns and suggest improvements.

Together, these mechanisms create a culture of continuous improvement, moving beyond one-time policy adoption.

5. Layer 4: Pedagogical Integration (The "Where It Happens")

The final layer brings governance into the classroom. It provides practical guidance for faculty and students on integrating GenAI responsibly into teaching and learning.

Key components include:

a. Process-Based Assessment Redesign

Instead of evaluating only final products, instructors assess the learning journey, requiring drafts, search logs, reflection essays, and AI-use declarations (Beckingham et al., 2024). This reduces incentives for cheating and promotes metacognition.

b. Student Declaration Forms

A standardized form where students disclose their use of GenAI in assignments, specifying which parts were AI-assisted and how prompts were refined. This fosters transparency and builds digital literacy.

c. Prompt Engineering Workshops

Faculty integrate prompt crafting into curricula, teaching students how to generate better outputs while understanding limitations (e.g., hallucinations).

d. Co-Intelligence Assignments

Activities where students critique AI-generated content, compare multiple outputs, or train simple models themselves, turning passive consumers into active evaluators (Mollick & Mollick, 2023).

For example, in a history course, students might ask ChatGPT to write a paragraph on colonialism, then analyze its omissions, biases, and framing, developing critical media literacy alongside subject knowledge.

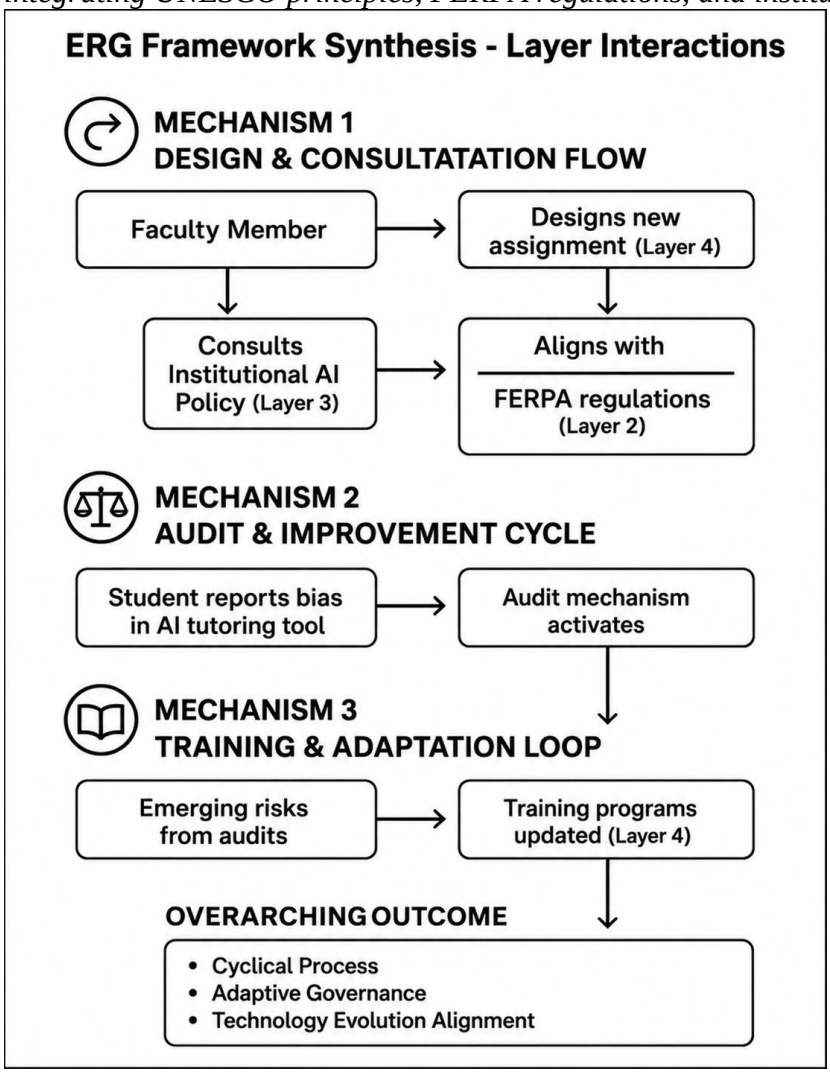
6. Synthesis: How the Layers Interact

The power of the ERG Framework lies in its interconnectedness:

- A faculty member designing a new assignment (Layer 4) consults the institution's AI policy (Layer 3) to ensure alignment with FERPA (Layer 2) and UNESCO's principle of human agency (Layer 1).
- If a student reports bias in an AI tutoring tool, the audit mechanism triggers a review, potentially leading to revised procurement criteria.
- Training programs update based on emerging risks identified in audits, closing the loop.

This cyclical, adaptive process ensures that governance evolves with technology, not lagging behind it.

Figure 2.
Interactions between the ERG Framework Layers: The mechanisms of design, audit, and training create a cyclical process of adaptive governance and technological alignment, integrating UNESCO principles, FERPA regulations, and institutional AI policies.



Discussion

The Ethico-Regulatory Governance (ERG) Framework presented in this study does not merely respond to the challenges of generative AI in higher education, it redefines how institutions should conceptualize AI governance. By integrating global ethical principles with enforceable regulatory standards and embedding them into pedagogical practice, our model transcends the limitations of prior approaches, which have largely treated ethics and compliance as separate domains.

1. Advancing Beyond Existing Models

Our framework directly addresses the central critique raised by Birhane (2023): that AI ethics has become "useless" due to its detachment from accountability. While UNESCO (2023), Siddiqui and Al-Binali (2024), and Hwang et al. (2020) offer valuable visions of human-centered AI, they stop short of specifying how these ideals translate into institutional action. The ERG Framework closes this gap by introducing Layer 2: Regulatory Anchors and Layer 3: Institutional Mechanisms, transforming aspirational values into auditable, legally defensible practices.

For example, while UNESCO calls for transparency, our framework operationalizes this principle through mandatory ethical disclosure by default (Laux et al., 2024)—requiring departments deploying GenAI tools to document their design choices, data sources, and bias mitigation strategies. This transforms transparency from a rhetorical commitment into a procedural requirement, aligning with Colonna's (2021) argument that effective AI regulation must include mechanisms for oversight and redress.

Similarly, while Chan's (2023) policy framework provides a useful step-by-step guide for educators, it lacks integration with legal obligations such as FERPA or GDPR. Our model explicitly maps ethical decisions to regulatory requirements, ensuring that faculty using AI for grading or advising do so within a compliant, institutionally supported structure. In doing so, we move beyond "ethics washing" toward governance with teeth.

2. Challenging Techno-Optimism and Alarmism Alike

The discourse on GenAI in education often swings between two extremes: uncritical techno-optimism ("AI will revolutionize learning!") and moral panic ("AI will destroy academic integrity!"). Our framework rejects both poles, advocating instead for a principled pragmatism, one that acknowledges both the transformative potential and the real risks of AI.

We affirm the benefits highlighted by Beckingham et al. (2024): GenAI can support ideation, provide instant feedback, and personalize learning. However, we also heed the warnings of Walczak and Cellary (2023), who show that students frequently fail to detect AI-generated misinformation, a finding echoed in Miller et al.'s (2023) demonstration of fabricated citations in medical content. Rather than banning or blindly adopting AI, our model promotes critical co-intelligence: teaching students not just how to use AI, but when, why, and under what conditions.

By requiring process-based assessments and student declaration forms, the ERG Framework turns AI use into an act of metacognitive engagement, where learners reflect on their reliance on technology and develop digital literacy as a core graduate competency.

3. Centering Equity and Justice in AI Governance

A major weakness of many existing frameworks is their failure to address systemic inequities. As Sanusi et al. (2024) reveal in their stakeholder study, teachers in Nigeria express deep concern that AI adoption will widen the digital divide, privileging urban, well-resourced schools. Similarly, Kaliisa et al. (2024) warn of an "uneven storm" of AI regulation, where Global North jurisdictions set standards without regard for local contexts.

Our framework confronts this head-on by incorporating Design Justice (Costanza-Chock, 2020) into Layer 1 and embedding equity audits into Layer 3. For instance, before deploying an AI tutoring system, institutions must assess whether it performs equally well across linguistic, cultural, and ability-based differences. This aligns with UNESCO's call for plural expressions and prevents the homogenization of knowledge production.

Furthermore, by mandating stakeholder feedback loops, including input from students, especially those from marginalized backgrounds, we ensure that governance is not imposed from above but co-constructed through inclusive dialogue.

4. Responding to Counterarguments

Some may argue that the ERG Framework imposes excessive bureaucracy on already overburdened faculty. To this, we respond: governance is not bureaucracy when it protects fundamental rights. Institutions routinely manage complex compliance systems, for research ethics (IRBs), financial aid, and Title IX. AI, which touches every aspect of teaching, learning, and data management, deserves no less rigor.

Others might claim that rapid technological change renders any framework obsolete. But as Laux et al. (2024) argue, the solution is not to abandon structure but to build adaptive, modular systems, exactly what the ERG Framework offers. Its cyclical design allows for regular review and revision, making it resilient to evolution in AI capabilities.

Finally, some institutions may resist external regulation, fearing loss of autonomy. Yet, as the EU AI Act and growing public scrutiny demonstrate, the era of self-regulation is ending. Proactive adoption of a robust governance model like ours positions universities not as reactive subjects of regulation, but as leaders in responsible innovation.

5. Theoretical and Practical Contributions

Theoretically, this study advances the field by synthesizing three previously siloed domains:

- Ethics (UNESCO, 2023),
- Law/Regulation (Colonna, 2021; Smith & Chen, 2024),
- Pedagogy (Beckingham et al., 2024; Wachter, 2023).

We demonstrate that sustainable AI integration requires all three and that their alignment produces more than the sum of its parts.

Practically, the ERG Framework offers a ready-to-adapt tool for university leaders, academic developers, and policymakers. It supports:

- Policy development aligned with FERPA, GDPR, and national laws,
- Curriculum redesign that fosters critical digital literacy,
- Capacity building through tiered training programs,
- Accountability via audits and disclosures.

Unlike abstract guidelines, this model provides actionable levers for change.

Conclusion

The rapid integration of generative AI into higher education is not a temporary disruption, it is a permanent transformation. Institutions can no longer afford reactive policies, ad hoc guidelines, or ethically ungrounded experimentation. The stakes are too high: academic integrity, data privacy, educational equity, and the very definition of learning are all at risk.

This study has responded to this urgent challenge by developing and presenting the Ethico-Regulatory Governance (ERG) Framework, a comprehensive model that bridges the persistent gap between global ethical principles and enforceable regulatory compliance. Unlike prior efforts that treat ethics and law as separate domains, the ERG Framework integrates them into a single, adaptive structure, grounded in UNESCO's human-centric vision, aligned with FERPA and GDPR requirements, and operationalized through institutional mechanisms and pedagogical practices.

Our framework makes three decisive contributions to the field.

First, it resolves the "principles-to-practice" gap that has plagued AI governance. By mapping abstract values like transparency and fairness to concrete actions, such as mandatory disclosure, bias audits, and student declaration forms, it transforms ethical commitments into accountable processes.

Second, it challenges the false dichotomy between innovation and regulation. Rather than positioning compliance as a barrier to progress, the ERG Framework demonstrates how robust governance enables responsible innovation. It empowers institutions to adopt GenAI with confidence, knowing that their use is both pedagogically sound and legally defensible.

Third, it centers justice and inclusion in AI governance. By embedding stakeholder engagement, equity audits, and Design Justice principles into its core, the framework ensures that AI serves all learners, not just the privileged few. It confronts algorithmic bias, digital divides, and epistemic marginalization head-on, advancing a vision of education that is equitable by design.

The implications for policy and practice are clear. Higher education institutions must move beyond fragmented, siloed approaches to AI. They must establish centralized AI governance bodies, integrate tiered training programs for faculty and students, and adopt process-based assessment models that prioritize critical thinking over rote output.

Moreover, national and international regulators must build on initiatives like the EU AI Act and UNESCO's guidance to create interoperable standards that support ethical AI adoption.

This paper does not offer a final answer. Technology will evolve. So must our governance. But what we present here is foundational: a living framework that institutions can adapt, refine, and scale.

The era of improvisation is over. The time for intentional, integrated, and inclusive AI governance is now.

Let this framework be the starting point for policy, for pedagogy, and for the future of education.

Bibliographic References

Amani, S., White, L., Balart, T., Arora, L., Shryock, K. J., Brumbelow, K., & Watson, K. L. (2023). *Generative AI Perceptions: A Survey to Measure the Perceptions of Faculty, Staff, and Students on Generative AI Tools in Academia*. arXiv. https://doi.org/10.48550/arXiv.2304.14415

Alali, R., & Wardat, Y. (2024). Opportunities and challenges of integrating generative artificial intelligence in education. *International Journal of Religion, 5*(3), 787–799. https://doi.org/10.61707/8y29gv34

Beckingham, S., Lawrence, J., & Taylor, S. (Eds.). (2024). *Using generative AI effectively in higher education: A guide for educators and learners*. Routledge. https://doi.org/10.4324/9781003482918

Birhane, A. (2023). The uselessness of AI ethics: A critical analysis. *AI and Ethics, 3*(4), 869–877. https://doi.org/10.1007/s43681-022-00209-w

Buckingham Shum, S., Ferguson, R., & Martinez-Maldonado, R. (2019). Human-centred learning analytics: A call for a more situated, participatory design. *Journal of Learning Analytics, 6*(2), 5–13. https://doi.org/10.18608/jla.2019.62.1

Chan, C. (2023). A comprehensive AI policy education framework for university teaching and learning. *Education and Information Technologies, 28*(12), 15469–15498. https://doi.org/10.1186/s41239-023-00408-3

Colonna, L. (2021). The AI Regulation and higher education: Preliminary observations. *Nordic Yearbook of Law and Informatics, 2020–2021*, 335–358. https://journals.uu.se/DeLege/article/view/428/384

Dabis, A., & Csáki, C. (2024). AI and ethics: Investigating the first policy responses of higher education institutions to the challenge of generative AI. *Humanities and Social Sciences Communications, 11*(1), 1006. https://doi.org/10.1057/s41599-024-03526-z

Gilley, A., & Gilley, J. W. (2006). FERPA: What Do Faculty Know? What Can Universities Do? *Innovative Higher Education, 31*(1), 17–26. https://eric.ed.gov/?id=EJ752407

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

Kaliisa, R., Baker, R. S., Wasson, B., & Prinsloo, P. (2024). The coming but uneven storm: How AI regulation will impact AI learning. *Postdigital Science and Education, 6*(3), 1–25.

Laux, J., Liefgreen, A., & Stephany, F. (2024). Three pathways for standardisation and ethical disclosure by default under the EU Artificial Intelligence Act. *Computer Law & Security Review*, *53*, 105957. https://doi.org/10.1016/j.clsr.2024.105957

Martínez-Comesaña, M., Rigueira-Díaz, X., Larranaga-Janerio, A., Martínez-Torres, M. R., & Iglesias-Boy, M. (2023). Impact of artificial intelligence on assessment methods in primary and secondary education: A systematic literature review. *Revista de Psicodidáctica*, *28*(2), 93–103. https://dialnet.unirioja.es/servlet/articulo?codigo=9134375

Miller, V. M., Bhattacharyya, M., Bhattacharyya, D., & Miller, L. E. (2023). High Rates of Fabricated and Inaccurate References in ChatGPT-Generated Medical Content. *Cureus*, *15*(5), e39238. https://doi.org/10.7759/cureus.39238

Perkins, M. (2023). Academic Integrity considerations of AI Large Language Models in the post-pandemic era: ChatGPT and beyond. *Journal of University Teaching and Learning Practice*, *20*(2). https://doi.org/10.53761/1.20.02.07

Sanusi, I. T., Agbo, F. J., Dada, O. A., Yunusa, A. A., Aruleba, K. D., Obaido, G., Olawumi, O., & Oyelere, S. S. (2024). Stakeholders' insights on artificial intelligence education: A qualitative study on conceptions, concerns, and support for AI integration in schools. *Computers and Education: Artificial Intelligence*, *7*, 100212. https://doi.org/10.1016/j.caeo.2024.100212

Siddiqui, S., & Al-Binali, H. (2024). A human-centric design framework for AI in education: Integrating ethics, regulation, and equity. *International Journal of Educational Technology in Higher Education*, *21*(1), 45.

UNESCO. (2023). *Guidance for generative AI in education and research*. https://www.unesco.org/en/articles/guidance-generative-ai-education-and-research

UNESCO. (2024). *AI competency framework for teachers*. https://www.cedefop.europa.eu/files/unesco_ai_competency_framework_for_teachers.pdf

Wachter, R. M. (2023). From ethical AI frameworks to tools: A systematic review of 100+ instruments for operationalizing AI ethics in healthcare and education. *Nature Digital Medicine*, *6*, 178. https://doi.org/10.1038/s41746-023-00913-9

Walczak, K., & Cellary, W. (2023). Challenges for higher education in the era of widespread access to generative AI. *Economics and Business Review*, *9*(2), 72–98. https://doi.org/10.18559/ebr.2023.2.743

Yeo, M. A. (2023). Academic integrity in the age of Artificial Intelligence (AI) authoring apps. *TESOL Journal*, *14*(3). https://doi.org/10.1002/tesj.716

Declarations

Conflict of Interest: The authors declare that there are no potential conflicts of interest.

Funding: No external financial support was received for this article.

Acknowledgments: N/A

Use of Generative Artificial Intelligence: N/A

Editorial Note: This article is not the result of a previous publication.