Building Trust Through Transparent AI Governance: Embedding Ethical Oversight Into Academic Curricula Development

Kelechi Ekuma
The University of Manchester, UK

kelechi.ekuma@manchester.ac.uk

Abstract: As artificial intelligence (AI) technologies become increasingly embedded in higher education, from adaptive learning platforms to algorithmic assessment tools, there is a growing imperative to examine the ethical and governance implications of such systems. While AI holds significant promise for improving efficiency and personalisation in education, its deployment also raises concerns around surveillance, algorithmic bias, data privacy, and equity. This article critically explores how transparent AI governance frameworks can be systematically integrated into academic curricula. It argues that ethical oversight and trust must be embedded not merely at the institutional policy level but within the design, delivery, and evaluation of educational programmes themselves. Drawing on interdisciplinary insights from digital governance, education policy, and data ethics, the paper develops a conceptual model for curriculum reform, articulates the ethical dimensions of AI in education, and presents illustrative case studies of institutions that have begun embedding such oversight into their teaching practices. By aligning curriculum design with emerging ethical principles, this study proposes that universities and higher education institutions can develop graduates who are not only technically proficient but ethically literate, capable of navigating and interrogating the socio-technical dimensions of AI in their future professions.

Keywords: Artificial intelligence, Higher education, Curriculum design, Data ethics, Transparent governance, Algorithmic accountability, Educational policy, Ethical pedagogy

1. Introduction

The rapid proliferation of artificial intelligence (AI) within higher education systems is transforming the fundamental dynamics of teaching, learning, and governance. AI now shapes a broad spectrum of academic functions, from automated admissions processes and AI-driven plagiarism detection to learning analytics and intelligent tutoring systems (Zawacki-Richter et al., 2019). These developments are frequently framed in terms of progress and innovation, often championed for their potential to enhance institutional efficiency, personalise student learning, and optimise resource allocation. However, such narratives frequently obscure the complex ethical, epistemological, and political implications of AI integration in educational spaces (Selwyn, 2022; Williamson, 2019).

Critical scholarship has started to challenge the techno-solutionist logic underpinning many AI deployments in education. Concerns about algorithmic bias, opaque decision-making processes, unequal access to digital infrastructure, and the commodification of student data raise urgent questions regarding justice, accountability, and institutional responsibility (Knox, 2020; Prinsloo & Slade, 2017). Furthermore, AI systems often operate within globalised commercial infrastructures, where datafied student experiences are subject to proprietary algorithms governed by external actors with limited institutional or pedagogical accountability. In this context, AI in education is not simply a technical innovation but a socio-political intervention with profound consequences for academic autonomy, student agency, and public trust.

Despite the growing literature on the ethics and governance of educational AI, there remains a significant disconnect between theoretical critique and curricular practice. While universities have increasingly adopted ethical AI policies at the institutional level, such frameworks rarely permeate the pedagogical core of academic programmes. Ethical AI principles tend to reside in high-level declarations or research ethics protocols, seldom filtering into learning outcomes, assessment structures, or instructional content (Floridi & Cowls, 2019; Williamson, 2019). This lacuna signals a deeper structural issue: ethical engagement with AI is too often marginalised or relegated to specialist disciplines, rather than being treated as a foundational element of higher education.

This article seeks to address that gap by interrogating how transparent AI governance can be embedded within the development and implementation of academic curricula. The central argument advanced here is that trust in AI systems – and the institutions that deploy them – must be cultivated through pedagogical transparency, critical inquiry, and ethical reflexivity. Curricula are not neutral repositories of knowledge; they are sites of power that shape how students conceptualise technology, authority, and social justice. Therefore, embedding AI ethics into curricula represents a necessary shift towards cultivating ethically literate graduates equipped to interrogate the socio-technical systems that will define their professional and civic lives.

The article adopts a multi-level analytical approach. It begins by establishing a theoretical and conceptual foundation that draws from critical pedagogy, curriculum theory, and digital ethics. It then maps the ethical risks associated with AI in education, highlighting the structural and epistemic challenges these technologies introduce. A comprehensive framework for curriculum reform is subsequently proposed, outlining principles and practices for embedding ethical AI governance into educational design. To illustrate these proposals, the article analyses case studies from institutions that have initiated reforms aimed at integrating AI ethics into teaching and learning. The final sections consider institutional and policy implications, offering recommendations for curriculum designers, accreditation bodies, and higher education leaders.

Overall, this paper calls for a reconfiguration of academic responsibility in the age of AI. It posits that the legitimacy and credibility of higher education depend not only on the adoption of innovative technologies but also on the cultivation of transparent, just, and democratically accountable pedagogical practices. By embedding ethical oversight into curricula, higher education institutions can reclaim a normative commitment to education as a public good – one that upholds the values of trust, transparency, and social responsibility in an increasingly datafied world.

2. Theoretical and Conceptual Foundations

A rigorous and critical exploration of AI governance in education necessitates a theoretical framework that not only interrogates power and ethics but also provides the conceptual tools to rethink curriculum design in an age of rapid technological change. This section draws from three interrelated domains – curriculum theory, digital ethics, and critical pedagogy – to formulate a multi-dimensional foundation for embedding AI governance into higher education curricula. These perspectives illuminate the socio-political dynamics that shape educational systems and provide normative guidance for aligning curriculum development with democratic values, human rights, and institutional accountability.

2.1 Curriculum Theory and the Politics of Educational Knowledge

Curriculum theory interrogates the purposes, structures, and contents of formal education, recognising that curricula are embedded within broader systems of cultural production and ideological control (Apple, 2008; Pinar, 2012). Traditional curriculum models, often informed by technocratic and neoliberal imperatives, tend to prioritise standardisation, efficiency, and market responsiveness over critical engagement and ethical inquiry (Au, 2012). Within such frameworks, the role of AI is frequently positioned as a neutral enhancer of educational delivery, abstracted from its socio-political implications.

Yet, as Bernstein (2000) notes, curricula encode specific forms of knowledge that reflect and reinforce dominant social relations. The absence of AI ethics and governance from mainstream educational programmes thus constitutes a form of epistemic exclusion – silencing critical perspectives on one of the most powerful technological transformations of the contemporary era. If curricula serve as mechanisms for social reproduction, then their redesign must become a site of contestation and transformation. Integrating AI ethics into curriculum development challenges not only what is taught but how knowledge is framed, whose voices are legitimised, and whose interests are served.

2.2 Digital Ethics, Surveillance Capitalism, and Algorithmic Governance

The rise of AI in education cannot be understood apart from the broader phenomenon of digital capitalism, where personal data becomes a primary commodity and algorithms operate as instruments of governance (Zuboff, 2019). Digital ethics emerges as a response to the normative crises generated by these transformations, offering a vocabulary for assessing the moral and political consequences of automated decision-making systems (Mittelstadt et al., 2016; Selbst et al., 2019).

In the educational domain, AI technologies are increasingly used to monitor student behaviour, predict academic performance, and personalise content delivery. These applications are frequently implemented with limited transparency or public debate, raising concerns about surveillance, consent, bias, and institutional accountability (Williamson, 2017). Ethical principles – such as fairness, explainability, non-maleficence, and data sovereignty - are often articulated in policy documents but lack operational clarity in everyday pedagogical practices.

Digital ethics thus demands more than reactive compliance with external regulations; it calls for proactive, critical, and context-sensitive engagements with the design, implementation, and consequences of AI systems. Embedding these concerns into curricula enables students to interrogate not only technical functionalities but also the value systems and power structures that underpin them.

2.3 Critical Pedagogy, Reflexivity, and Epistemic Justice

Critical pedagogy, grounded in the works of Freire (1970), Hooks (1994), and more recently scholars like Andreotti (2011), provides a transformative vision of education as a practice of freedom, justice, and democratic participation. It rejects the banking model of education – where knowledge is deposited into passive learners – and instead positions students as active agents capable of critiquing and reshaping the world around them.

In the context of AI, critical pedagogy emphasises the need for epistemic justice: the right of learners to access, produce, and challenge knowledge that affects their lives. AI technologies, when uncritically implemented, risk entrenching epistemic injustices by privileging certain data regimes, algorithmic logics, and commercial interests over lived experience, contextual knowledge, and social accountability (Fricker, 2007; Knox, 2020).

Embedding critical pedagogy into AI governance curricula involves cultivating reflexive awareness among students – encouraging them to question how technologies are constructed, whose perspectives are embedded within them, and who bears the risks of their failures. This approach shifts the educational emphasis from mere technical proficiency to critical consciousness, preparing students to participate ethically and democratically in an AI-mediated world.

2.4 Toward an Integrated Framework for Ethical Curriculum Design

These three domains - curriculum theory, digital ethics, and critical pedagogy – provide the building blocks for a holistic and ethically grounded approach to curriculum design. They challenge prevailing assumptions about neutrality, efficiency, and progress, instead advancing a normative agenda that places social justice, transparency, and human dignity at the heart of AI governance in education.

An integrated framework based on these principles calls for:

- Recognising the curriculum as a contested space where technological ideologies must be interrogated.
- Embedding ethical reasoning and digital literacy as core learning outcomes across disciplines.
- Designing assessment practices that reward critical thinking about technological impacts.
- Providing professional development for educators to navigate the ethical and pedagogical dimensions of AI.

In doing so, institutions can begin to cultivate a new generation of learners equipped not only to use AI technologies but to shape their development in ethically responsible and socially responsive ways. The following sections examine how these theoretical insights can be operationalised through curriculum reform and pedagogical innovation, addressing the structural and epistemic risks posed by AI while foregrounding principles of trust, transparency, and justice.

3. Ethical Risks and Governance Challenges in AI-Driven Education

The increasing reliance on artificial intelligence within educational systems has introduced a range of ethical risks and governance challenges that are both systemic and underexamined. While many institutions have embraced AI tools for purposes such as student monitoring, performance analytics, and personalised learning, these applications often outpace institutional capacity for critical oversight and ethical regulation. This section provides a critical analysis of the key ethical risks posed by AI in educational contexts and identifies the structural conditions that exacerbate their impact.

3.1 Algorithmic Bias and Discrimination

One of the most persistent risks associated with AI technologies is the reproduction and amplification of bias. Algorithms trained on historically biased datasets can reinforce existing inequities, particularly for students from racialised, disabled, and socioeconomically disadvantaged backgrounds (Benjamin, 2019; Noble, 2018). Predictive analytics used in academic advising, for instance, may systematically underestimate the potential of students from marginalised groups, leading to exclusionary decisions regarding access to advanced courses or scholarships.

Moreover, the opacity of proprietary algorithms used by edtech companies creates accountability gaps. Without access to the underlying data and model architecture, educational institutions and students have limited recourse to challenge or interrogate discriminatory outcomes. As Eubanks (2018) argues, such

algorithmic governance mechanisms often function as "automated inequality," embedding bias under the guise of neutral computation.

3.2 Surveillance and Datafication of Learning

Another key concern relates to the surveillance infrastructures embedded within AI-enhanced educational platforms. Learning management systems, plagiarism detection tools, and remote proctoring technologies collect vast amounts of data on student behaviour, often without meaningful consent or understanding. This expansion of datafication raises fundamental questions about autonomy, privacy, and the erosion of trust in the student-institution relationship (Williamson et al., 2020).

The normalisation of surveillance in educational settings may have chilling effects on intellectual freedom and student expression. Additionally, the repurposing of educational data for secondary uses – such as behavioural profiling or third-party marketing – increases concerns over data sovereignty and institutional complicity in digital exploitation (van Dijck et al., 2018).

3.3 Inequitable Access and the Digital Divide

Despite claims that AI can democratise education through personalisation and adaptive technologies, unequal access to digital resources remains a structural barrier. Students in underfunded institutions or regions with limited internet infrastructure often face systemic exclusion from AI-enabled learning environments. The implementation of AI tools without consideration of access differentials risks deepening existing educational inequalities (UNESCO, 2021).

In addition, many AI applications require students to have sophisticated digital literacies to engage meaningfully. The failure to support the development of these literacies further entrenches inequities, disproportionately affecting learners who are already marginalised by race, gender, disability, or class.

3.4 Institutional Inertia and Ethical Blind Spots

Even when ethical principles are articulated at the policy level, institutions often struggle to translate them into operational practices. Institutional inertia – driven by bureaucratic constraints, budgetary limitations, and commercial partnerships – frequently results in a gap between ethical aspiration and pedagogical implementation. Moreover, decision-making around AI adoption is often concentrated among senior administrators or IT professionals, with limited input from educators or students (Selwyn, 2022).

This centralisation of technological governance not only undermines democratic accountability but also reinforces a technocratic logic that prioritises efficiency over pedagogy. Without institutional mechanisms for ethical deliberation, participatory governance, and curricular integration, AI deployment remains disconnected from the values and mission of higher education.

Addressing these challenges requires a paradigm shift that centres ethical reflexivity and transparency in both policy and curriculum. The next section develops a framework for embedding these values into curriculum design, offering practical strategies for equipping students and educators to navigate the socio-technical complexities of AI-driven education.

4. Methodology and Approach

This article employs a conceptual and interpretive methodology grounded in critical analysis, documentary review, and illustrative case study synthesis. Rather than pursuing empirical generalisation, the objective is to construct a theoretically informed and practically relevant framework for embedding ethical AI governance into academic curricula. The approach reflects the growing consensus that curriculum design in the era of AI must be informed by interdisciplinary scholarship, normative principles, and institutional practice (Williamson, 2019; Floridi & Cowls, 2019).

The methodological strategy unfolds in three interconnected stages: (1) critical literature synthesis, (2) purposive case study selection and analysis, and (3) conceptual framework development.

4.1 Critical Literature Synthesis

The first stage entailed a systematic synthesis of academic and policy literature across four domains:

- Curriculum theory and educational policy (Young & Muller, 2013; Pinar, 2012).
- Digital ethics and responsible AI (Floridi & Cowls, 2019; Mittelstadt, 2016).

- AI governance and regulatory standards (European Commission, 2019; UNESCO, 2021; OECD, 2024).
- Critical pedagogy and democratic education (Freire, 1970; Giroux, 2011).

This interdisciplinary synthesis helped distil core ethical imperatives (transparency, fairness, accountability), governance dilemmas (opacity, regulatory lag), and curricular implications (integration of values, ethics-based assessments). The normative analysis drew upon international guidelines such as:

- The European Commission's Ethics Guidelines for Trustworthy AI (2019).
- UNESCO's Recommendation on the Ethics of Artificial Intelligence (2021).
- The OECD AI Principles (2024).

These documents provided essential reference points for evaluating how ethical principles can inform both curriculum content and institutional structures.

4.2 Case Study Selection and Documentary Analysis

To ground the normative analysis in real-world practice, three illustrative case studies were selected based on the following inclusion criteria:

- Demonstrated commitment to ethical AI integration within formal curricula.
- Availability of detailed curriculum documents on public domains.
- Institutional alignment with interdisciplinary and socially responsive education.

Documentary materials analysed included course syllabi, programme handbooks, institutional mission statements, and online curriculum repositories. The case studies reflect different geographical, institutional, and disciplinary contexts as highlighted in Table 1:

Table 1: Summary of Cases

Institution	Programme / Initiative	Core Features	Pedagogical Approach
University College London (UK)	MSc in Public Policy, Technology and Innovation (STEaPP) – Distance Learning	100% online PGT programme focusing on policy ethics, algorithmic accountability, digital governance.	Interdisciplinary, policy-oriented, critical inquiry.
Monash University (Australia)	ATS2292: Ethics of AI (Undergraduate course)	Mandatory ethics module in Monash Bioethics Centre; themes of bias, fairness, human-centred design.	Cross-disciplinary, case-based, values-driven.
Stanford University (USA)	CS181: Computers, Ethics, and Public Policy	Required ethics integration for CS majors; emphasis on platform governance and normative reasoning	Integrated humanities-computing model, reflexive.

These cases exemplify diverse strategies for embedding ethical reasoning, reflexive critique, and social accountability into AI-related educational programmes. The comparative analysis identified shared priorities (e.g. algorithmic fairness, civic responsibility), as well as structural and pedagogical innovations relevant to curriculum reform.

4.3 Framework Development and Normative Synthesis

The final stage involved the construction of a multi-level conceptual framework for embedding ethical AI governance into higher education curricula. This framework was iteratively developed by synthesising thematic insights from the literature and international guidelines, pedagogical strategies observed in the case studies, and theoretical perspectives from curriculum theory (Bernstein, 2000), critical digital pedagogy (Knox, 2020), and AI ethics (Mittelstadt, 2016).

The resulting framework – presented in the next section – encompasses three interrelated levels:

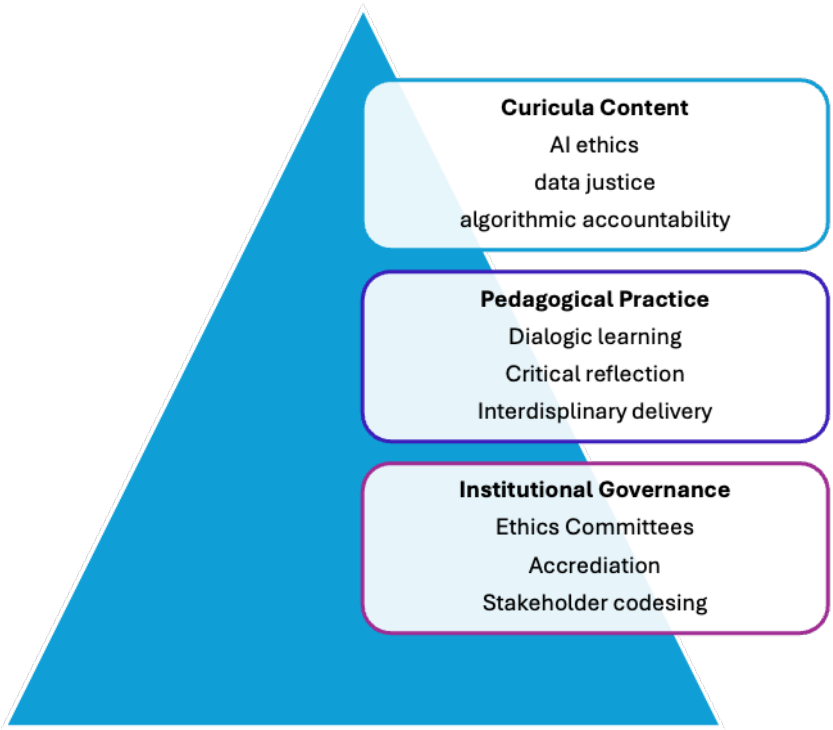
- Curricular Content: Integration of ethical principles, regulatory knowledge, and socio-technical critique into programme modules.
- Pedagogical Practice: Use of critical reflection, interdisciplinary problem-solving, and dialogic learning to foster ethical reasoning.
- Institutional Governance: Establishment of ethics boards, co-design mechanisms, and accreditation standards to reinforce ethical commitments at the systemic level.

This methodology provides the scaffolding for a normative intervention: a practical yet theoretically grounded proposal for reimagining higher education curricula in the age of AI. It privileges conceptual clarity, interpretive insight, and institutional relevance - responding to the urgent need for trust, accountability, and transparency in educational engagement with AI technologies.

5. Framework Development and Curriculum Reform

Efforts to embed ethical AI governance into higher education require far more than the retrofitting of technical syllabi with standalone ethics modules. They demand a systemic transformation of educational design, delivery, and institutional purpose. This section builds on the conceptual framework and case analysis presented earlier to articulate a robust model for curriculum reform that foregrounds trust, transparency, and accountability in the age of AI. The framework operates across three interdependent layers - curricular content, pedagogical practice, and institutional governance – each of which must be critically reconfigured if higher education is to fulfil its democratic and epistemic responsibilities in shaping AI futures.

Figure 1 presents a visual representation of a layered framework for embedding ethical AI governance into higher education curricula.



Source: Author (2025).

Figure 1: Layered Framework for Embedding Ethical AI Governance in Higher Education

The framework highlights three interdependent levels, each serving a distinct but interconnected role in fostering transparency, accountability, and ethical reflexivity. The layered structure underscores the foundational importance of curriculum design, while recognising that pedagogical strategies and institutional governance are essential for sustaining meaningful and systemic change.

It should be stressed, however, that this framework is not prescriptive but strategic. It provides a normative orientation and practical structure through which universities can diagnose gaps, build capabilities, and design ethically robust curricula responsive to the socio-technical complexity of AI. In doing so, it reasserts the public role of higher education – not as a pipeline for technical labour, but as a democratic institution tasked with preparing ethically responsible, critically informed, and socially engaged citizens.

5.1 Rethinking Curricular Content: From Technical Instrumentalism to Ethical Literacy

Contemporary AI curricula often reflect a technocratic orientation – prioritising optimisation, efficiency, and deployment over ethical reflection and public responsibility (Williamson, 2019; Knox, 2020). This instrumental approach is insufficient for addressing the profound societal challenges posed by algorithmic systems,

including surveillance, discrimination, and epistemic injustice. Ethical literacy, therefore, must become a core learning outcome – not only for computer science students, but across disciplines engaged with data and digital technologies.

A reformed curriculum should include:

- **Critical AI studies**: Modules that introduce students to the social, historical, and political dimensions of AI. Such offerings, inspired by the work of scholars like Noble (2018), Benjamin (2019), and Crawford (2021), equip learners to interrogate the embedded power relations and racialised logics within AI systems.
- **Algorithmic accountability and explainability**: Teaching students to examine the design assumptions, training datasets, and systemic risks of AI technologies. This includes practical exercises in identifying and mitigating algorithmic harms, drawing on frameworks from Binns (2018), Sandvig et al. (2014), and Mittelstadt et al. (2016).
- **Global, plural, and decolonial perspectives**: AI education must go beyond Eurocentric ethics paradigms by including epistemologies and use-cases from the Global South, Indigenous communities, and historically marginalised voices (Mohamed et al., 2020; Birhane, 2021). This enables students to appreciate the situated ethics of AI and resist normative universalism in technological development.

Importantly, these curricular elements must not be siloed in elective courses or isolated in philosophy departments. Instead, they should be embedded into core disciplinary pathways – e.g., engineering ethics as a compulsory component of machine learning modules, or critical data studies within public policy degrees. The case of Monash University illustrates this integration by mandating AI ethics training for undergraduates, thereby institutionalising ethical reasoning as a foundational competency rather than an optional enrichment.

5.2 Reimagining Pedagogical Practice: Ethical Reasoning as Transformative Learning

The pedagogy through which AI ethics is taught is just as critical as the content itself. Ethical reasoning cannot be meaningfully cultivated through didactic lectures, rote memorisation, or abstract codes of conduct. Instead, it requires a pedagogy of critical reflexivity, dialogue, and experiential engagement – an approach long championed by critical theorists such as Freire (1970), Hooks (1994), and Giroux (2011).

Ethical pedagogy should prioritise:

- **Dialogic learning**: Classrooms should function as spaces of critical dialogue, where students are encouraged to engage with ethical dilemmas through collaborative reasoning, debate, and situated reflection. Case studies involving controversial AI deployments (e.g., predictive policing, facial recognition bans, biased hiring algorithms) can serve as springboards for such dialogue.
- **Problem-based and scenario-driven approaches**: Rather than abstract ethics, students should tackle complex real-world problems that require both technical and normative reasoning. For instance, an assignment may involve auditing a content moderation algorithm for bias and then proposing mitigations based on fairness principles and legal considerations.
- **Critical self-reflection and values clarification**: Students should be supported to examine their own assumptions, professional identities, and moral frameworks. This is particularly important in technical disciplines, where students may have internalised value-neutral conceptions of design and development. Reflective essays, positionality statements, and ethical impact assessments can cultivate this awareness.

The UCL and Stanford cases both exemplify such pedagogical approaches, where students are engaged not merely as passive learners but as active moral agents - capable of scrutinising the systems they build and the social orders those systems help sustain.

Equally crucial is capacity-building among faculty. Many educators lack formal training in ethical theory, digital rights, or critical pedagogy. Universities must therefore invest in professional development programmes that support interdisciplinary team teaching, critical facilitation skills, and engagement with emerging debates in AI governance. Without this scaffolding, even well-designed curricula risk being delivered in ways that undermine their ethical ambitions.

5.3 Transforming Institutional Governance: Structural Alignment for Ethical Oversight

Curriculum and pedagogy alone cannot carry the weight of ethical transformation. Institutional governance structures must also be reconfigured to ensure that ethical commitments are institutionalised, enforced, and continuously reviewed. Ethical education in AI must be embedded not just in syllabi, but in the institutional DNA of universities.

This includes:

- **Curriculum ethics committees**: Establishing formal review bodies that assess the ethical content and implications of new and existing programmes. These committees should include interdisciplinary faculty, students, and external stakeholders to ensure legitimacy, accountability, and diverse perspectives.
- **Embedding ethics in accreditation and quality assurance**: Universities should work with national regulators and professional bodies to include ethical AI education as a core requirement for programme approval and renewal. As the OECD and UNESCO guidelines suggest, governance frameworks for trustworthy AI must include human capital development as a pillar of policy.
- **Participatory governance and stakeholder engagement**: Institutions should institutionalise co-design processes involving civil society, industry, and affected communities – especially those historically marginalised by algorithmic systems. This reflects a shift from top-down curriculum imposition to democratic knowledge production, aligning with the principles of public accountability and civic responsibility.

These bodies can also serve as bridges between academia and public discourse, helping to shape policy debates and public understanding of AI ethics.

The multi-layered conceptual framework (refer to Figure 1) underscores that ethical AI education is not merely a matter of content addition but requires systemic alignment across educational design, instructional delivery, and institutional regulation.

6. Discussion: Operationalising Ethical AI Education in Context

While the conceptual framework outlined above offers a strategic vision for embedding ethical AI governance into higher education, its practical realisation is fraught with institutional, political, and epistemological challenges. This discussion critically reflects on these tensions, exploring what it means to operationalise ethical AI curricula within diverse educational contexts shaped by uneven resources, competing imperatives, and global asymmetries in technological development.

6.1 Structural Constraints: Fragmented Governance and Institutional Inertia

A major barrier to reform lies in the fragmented and often conservative governance structures of higher education institutions. Curriculum development tends to be decentralised, slow-moving, and shaped by disciplinary silos. As such, even well-intentioned initiatives to integrate AI ethics often face resistance or dilution, especially in technically oriented faculties where ethical inquiry is seen as tangential or peripheral.

Moreover, institutional incentives remain misaligned. In many contexts, performance metrics tied to graduate employability, publication outputs, or industry funding discourage critical, values-based curriculum reform. There is often a reluctance to introduce content that could be seen as politically sensitive, ideologically divisive, or commercially disruptive. Consequently, ethical education risks being marginalised or sanitised, reduced to abstract principles rather than positioned as a foundation for critical engagement and social accountability.

This highlights the importance of institutional leadership and cross-functional governance mechanisms. Without support from senior administrators, quality assurance bodies, and academic boards, ethical AI education will remain patchy and under-resourced.

6.2 Epistemic Tensions: Whose Ethics, Whose Knowledge?

A second set of challenges arises from the epistemological assumptions embedded in dominant approaches to AI ethics. Much of the mainstream discourse reflects liberal individualist and Western moral frameworks – grounded in abstract principles like autonomy, justice, and beneficence (Floridi & Cowls, 2019). While these concepts are valuable, their universalisation often obscures cultural specificity, political context, and historical injustice.

As Mohamed et al. (2020) argue, global AI ethics must move beyond "ethics-washing" to engage with structural inequalities, colonial legacies, and systemic harms. Similarly, Birhane (2021) cautions against reductive technical fixes to ethical problems that are fundamentally social and political in nature. These raises pressing questions for curriculum designers:

- Whose ethical frameworks are being taught, and why?
- How can curricula reflect pluralistic, decolonial, and community-based approaches to AI ethics?
- What forms of knowledge and experience are being excluded in the name of technical neutrality?

Addressing these questions requires a deeper commitment to epistemic justice. It entails not only diversifying reading lists and case studies but also creating space for students to engage with contested values, lived experiences, and critical traditions from across the globe.

6.3 Resourcing and Equity: Who Has the Capacity to Reform?

The ability to implement a comprehensive ethics-focused curriculum reform is also shaped by material conditions. Well-resourced institutions – such as those included in this study (UCL, Stanford, Monash) – may have access to interdisciplinary faculty, external partnerships, and flexible governance structures. In contrast, universities in the Global South, public institutions facing austerity, or newer universities may lack the institutional bandwidth to support such reforms.

This disparity raises concerns about global equity and the reproduction of epistemic privilege. If only elite institutions can afford to "do ethics well," then ethical AI education risks becoming another site of global inequality - where critical literacy and civic agency are privileges of the few rather than rights of the many.

Thus, the framework must be adaptable and scalable. Policymakers, development agencies, and philanthropic funders have a role to play in supporting capacity-building, faculty training, and collaborative knowledge-sharing initiatives across institutions and borders.

6.4 Toward Transformative Possibilities: Ethics as Praxis

Despite these challenges, the growing visibility of AI's societal impacts presents a critical pedagogical opportunity. Students are increasingly aware of the ethical stakes of technological development – from climate surveillance to predictive policing to AI-generated misinformation. There is a rising demand for educational experiences that go beyond instrumental skills and offer critical frameworks for ethical reasoning, political engagement, and collective responsibility.

Curriculum reform can harness this momentum - but only if ethics is approached not as a static body of knowledge, but as praxis: reflective, situated, and action-oriented. This involves:

- Framing ethics as a mode of inquiry, not a checklist.
- Situating technological critique within broader questions of justice, democracy, and solidarity.
- Encouraging students to see themselves not just as future employees, but as citizens, designers, and change agents.

When embedded effectively, ethical AI education has the potential to rehumanise technological learning, reorient institutional values, and reclaim the civic mission of the university.

7. Conclusion: Ethical AI Governance as a Curriculum Imperative

As artificial intelligence continues to reshape the epistemic, economic, and political landscapes of contemporary societies, higher education institutions face a profound responsibility: to prepare students not merely to work with AI, but to critically engage with its ethical, social, and political dimensions. This paper has argued that ethical AI governance must be embedded into the core of academic curricula - not as an ancillary concern, but as a foundational imperative.

Through a multi-level framework encompassing curricular content, pedagogical practice, and institutional governance, the paper has outlined a strategy for designing ethical AI education that is theoretically grounded, practically feasible, and normatively robust. Drawing on interdisciplinary literature, international policy guidelines, and illustrative case studies from UCL, Monash, and Stanford, it has been demonstrated that systemic, cross-cutting reform is essential if higher education is to build trust, foster transparency, and contribute to a more just technological future.

The framework presented here emphasises that:

- Ethical literacy must be cultivated across disciplines and not confined to philosophy or ethics modules.
- Pedagogical methods must centre dialogue, critical reflection, and experiential engagement rather than rote compliance.
- Institutional structures must support ethics through formal governance, inclusive curriculum review processes, and alignment with accreditation standards.

However, the implementation of this agenda is not without challenge. As the discussion has shown, universities must navigate entrenched silos, resourcing disparities, competing political pressures, and deep epistemic questions about whose ethics and whose knowledge are privileged in curriculum design. These constraints cannot be ignored. Yet they do not negate the urgency or the possibility of change.

In fact, the proliferation of AI-driven harms – from algorithmic discrimination to automated misinformation – makes ethical curriculum reform more urgent than ever. In this context, higher education can no longer position itself as a neutral pipeline for technological expertise. It must instead act as a critical site of democratic formation, equipping students to ask difficult questions, recognise power asymmetries, and imagine alternatives to dominant techno-economic paradigms.

This transformation requires not only conceptual clarity and practical frameworks, but courageous institutional leadership and collective pedagogical imagination. It demands that ethics be treated not as a tick-box exercise, but as an evolving, contested, and situated form of knowledge - one that reflects the complexities of living in a world increasingly shaped by artificial intelligence.

8. Future Directions

While this paper has laid the groundwork for a theoretically and practically informed approach to embedding ethical AI governance in curricula, several areas merit further investigation:

- Empirical research on student experiences, learning outcomes, and educator perspectives in ethics-integrated programmes.
- Comparative analyses of how ethical AI education is implemented in different cultural, economic, and regulatory contexts.
- Policy research on the role of national quality assurance bodies, international standards, and funding agencies in promoting ethics integration.
- Collaborative curriculum development involving civil society, industry actors, and marginalised communities to ensure inclusive, context-responsive design.

Ultimately, building trust through transparent AI governance is not solely a technical or institutional task. It is an educational challenge – and one that higher education must meet with both intellectual rigour and moral seriousness. Indeed, the integration of transparent AI governance into academic curricula is not a peripheral concern but a strategic necessity. Failure to act risks entrenching opaque and inequitable systems that erode trust, weaken educational integrity, and leave graduates unprepared for the ethical and societal challenges of AI. Conversely, embedding governance principles into teaching and learning creates pathways for accountability, human-centred innovation, and employability, while positioning universities and higher education institutions as proactive leaders in shaping digital futures that serve the public good.

Ethics Declaration: Ethical clearance was not required for the research.

AI Declaration: I confirm that generative AI tools were utilised in the preparation of this manuscript. Specifically, ChatGPT-4.0 (OpenAI) and Grammarly (1.127.0.0) were employed for proofreading, summarising, and improving the clarity and coherence of the text. These tools were not used to produce original research content or interpret data. All outputs were reviewed and edited by the author to ensure accuracy and academic integrity.

References

Andreotti, V. (2011). Actionable postcolonial theory in education. Springer.

Apple, M. W. (2008). Content, Form, and the Politics of Accountability. The SAGE handbook of curriculum and instruction, 25.

Au, W. (2012). Critical curriculum studies: Education, consciousness, and the politics of knowing. Routledge.

Benjamin, R. (2019). Assessing risk, automating racism. Science, 366(6464), 421-422.

Bernstein, B. (2000). Pedagogy, symbolic control, and identity (Vol. 5). Bloomsbury Publishing PLC.

Binns, R. (2018). Algorithmic accountability and public reason. Philosophy & technology, 31(4), 543-556.
Birhane, A. (2021). Algorithmic injustice: A relational ethics approach. Patterns, 2(2), 100205. https://doi.org/10.1016/j.patter.2021.100205
Crawford, K. (2021). Atlas of AI: Power, politics, and the planetary costs of artificial intelligence. Yale University Press.
Eubanks, V. (2018). Automating inequality: How high-tech tools profile, police, and punish the poor. St. Martin's Press.
European Commission. (2019). Ethics guidelines for trustworthy AI. Publications Office of the European Union. https://op.europa.eu/publication-detail/-/publication/d3988569-0434-11ea-8c1f-01aa75ed71a1 [Accessed 29 July 2025].
Floridi, L., & Cowls, J. (2019). A unified framework of five principles for AI in society. Harvard Data Science Review, 1(1). https://doi.org/10.1162/99608f92.8cd550d1
Freire, P. (1970). Pedagogy of the oppressed. Continuum.
Fricker, M. (2007). Epistemic injustice: Power and the ethics of knowing. Oxford University Press.
Giroux, H. A. (2020). On critical pedagogy. Bloomsbury Academic.
Hooks, B. (1994). Teaching to transgress: Education as the practice of freedom. Routledge.
Knox, J. (2020). Artificial intelligence and education in China. Learning, Media and Technology, 45(3), 298–311. https://doi.org/10.1080/17439884.2020.1754236
Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. Big Data & Society, 3(2). https://doi.org/10.1177/2053951716679679
Mohamed, S., Png, M. T., & Isaac, W. (2020). Decolonial AI: Decolonial theory as sociotechnical foresight in artificial intelligence. Philosophy & Technology, 33, 659–684. https://doi.org/10.1007/s13347-020-00405-8
Noble, S. U. (2018). Algorithms of oppression: How search engines reinforce racism. In Algorithms of oppression. New York University Press.
OECD. (2024). OECD principles on artificial intelligence. https://oecd.ai/en/ai-principles [Accessed 12 July 2025].
Pinar, W. F. (2012). What is curriculum theory? (2nd ed.). Routledge.
Prinsloo, P., & Slade, S. (2017, March). An elephant in the learning analytics room: The obligation to act. In Proceedings of the seventh international learning analytics & knowledge conference (pp. 46-55).
Sandvig, C., Hamilton, K., Karahalios, K., & Langbort, C. (2014). Auditing algorithms: Research methods for detecting discrimination on internet platforms. Data and discrimination: converting critical concerns into productive inquiry, 22(2014), 4349-4357.
Selbst, A. D., Boyd, D., Friedler, S. A., Venkatasubramanian, S., & Vertesi, J. (2019, January). Fairness and abstraction in sociotechnical systems. In Proceedings of the conference on fairness, accountability, and transparency (pp. 59-68). https://doi.org/10.1145/3287560.3287598
Selwyn, N. (2022). The future of AI and education: Some cautionary notes. European Journal of Education, 57(4), 620-631.
UNESCO. (2021). Recommendation on the ethics of artificial intelligence. https://unesdoc.unesco.org/ark:/48223/pf0000381137 [Accessed 29 July 2025].
Van Dijck, J., de Winkel, T., & Schäfer, M. T. (2023). Deplatformization and the governance of the platform ecosystem. New Media & Society, 25(12), 3438-3454.
Williamson, B. (2017). Who owns educational theory? Big data, algorithms and the expert power of education data science. E-learning and Digital Media, 14(3), 105-122.
Williamson, B. (2019). Policy networks, performance metrics and platform markets: Charting the expanding data infrastructure of higher education. British Journal of Educational Technology, 50(6), 2794–2809. https://doi.org/10.1111/bjet.12849
Williamson, B., Bayne, S., & Shay, S. (2020). The datafication of teaching in Higher Education: critical issues and perspectives. Teaching in higher education, 25(4), 351-365.
Young, M., & Muller, J. (2013). On the powers of powerful knowledge. Review of Education, 1(3), 229–250. https://doi.org/10.1002/rev3.3017
Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education–where are the educators? International journal of educational technology in higher education, 16(1), 1-27.
Zuboff, S. (2019, January). Surveillance capitalism and the challenge of collective action. In New labor forum (Vol. 28, No. 1, pp. 10-29). Sage CA: Los Angeles, CA: Sage Publications.