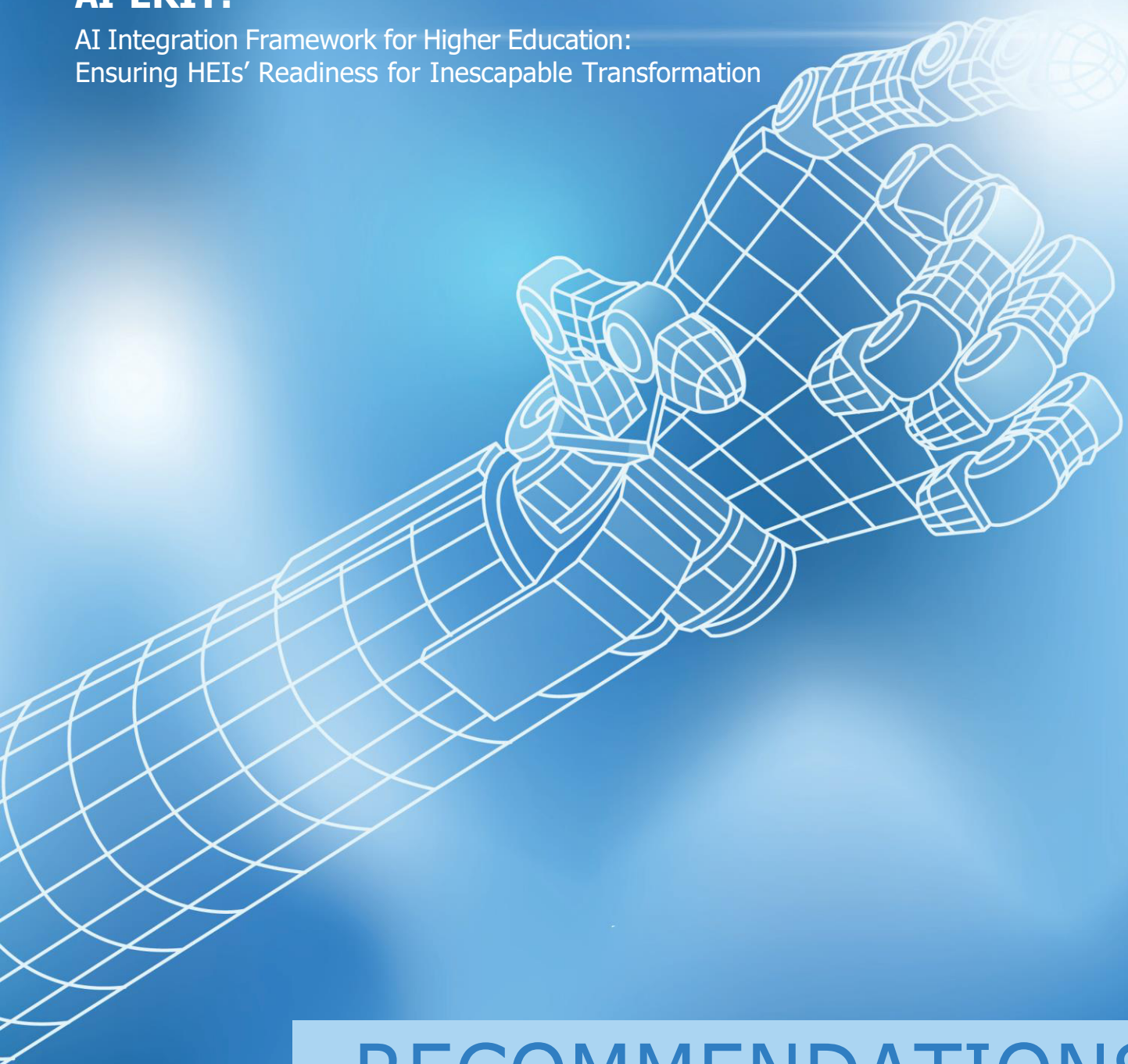


Created by:

## AI-ERIT:

AI Integration Framework for Higher Education:  
Ensuring HEIs' Readiness for Inescapable Transformation



# RECOMMENDATIONS

**on Implementing AI in Line  
with the Ethical Principles**



## Information

Dear Visitors,

Welcome to the “**Recommendations on Implementing AI in Line with the Ethical Principles**”, where you will find concrete, ethically grounded guidelines for Higher Education Institutions to implement AI in ways that ensure transparency, fairness, inclusiveness, and respect for human dignity while strengthening academic integrity and public trust.

### THE SURVEY

After reading these recommendations, we kindly ask you to complete a very short survey by providing your feedback about it.

**Please find the link to the survey [HERE](#)**

### TRANSLATION

For translating the document into your native language, you are kindly advised to use the following tool:

-  **DeepL Translator**

High-quality neural machine translation; supports **over 30 languages**. Upload your PDF or PPTX to translate.

 DeepL: <https://www.deepl.com/en/translator>

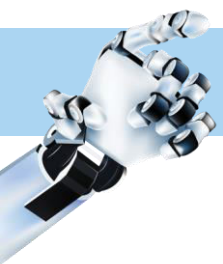
-  **Google Docs (Translate document)**

Powered by Google Translate; supports **over 130 languages**. For PDFs and PowerPoints, upload the file to Google Drive, open it in Docs or Slides, then select **Tools > Translate document**.

 <https://translate.google.com/?sl=en&tl=el&op=docs>

# CONTENT

<b>INTRODUCTION .....</b>	<b>3</b>
<b>1. EDUCATION PROCESSES FOSTERED BY AI TOOLS.....</b>	<b>4</b>
<b>2. THE ETHICAL PRINCIPLES.....</b>	<b>5</b>
21. ACCOUNTABILITY AND RESPONSIBILITY .....	6
22. BIAS AND FAIRNESS .....	6
23. HUMAN AUTONOMY AND AGENCY .....	7
24. PRIVACY AND DATA PROTECTION.....	8
25. SAFETY AND SECURITY .....	9
26. INCLUSIVITY .....	9
27. TRANSPARENCY AND EXPLAINABILITY .....	10
<b>3. IMPLEMENTING AI ETHICALLY IN HIGHER EDUCATION INSTITUTIONS:     KEY RECOMMENDATIONS.....</b>	<b>11</b>
31. ESTABLISHING AN ETHICAL GOVERNANCE STRUCTURE FOR AI .....	11
32. SAFEGUARDING PRIVACY IN AI USE: RESPONSIBLE DATA PRACTICES AND INSTITUTIONAL READINESS .....	17
33. PROMOTING FAIRNESS AND PREVENTING BIAS IN AI-SUPPORTED DECISION-MAKING .....	24
34. ALIGNING AI USE WITH INSTITUTIONAL ETHICS AND HUMAN AGENCY.....	28
35. ENSURING INCLUSIVE AND ACCESSIBLE AI USE .....	33
36. EMBEDDING AI ETHICS INTO THE CURRICULUM .....	37
37. UPDATING INSTITUTIONAL POLICIES TO REFLECT AI ETHICS.....	41
<b>CONCLUSION .....</b>	<b>50</b>
<b>REFERENCES .....</b>	<b>51</b>



## Introduction

As Artificial Intelligence (AI) technologies continue to evolve and gain ground in higher education, it is increasingly clear that their implementation cannot be guided solely by efficiency, innovation, or technological advancement. The integration of AI into educational ecosystems must be aligned with a broader set of ethical responsibilities that respect human dignity, promote inclusion, ensure transparency, and uphold the values of academic integrity, autonomy, and justice.

Higher Education Institutions (HEIs) play a pivotal role in shaping not only the knowledge and competencies of future generations but also the ethical frameworks within which digital transformation occurs. As such, HEIs are uniquely positioned to ensure that the use of AI technologies is guided by principles that reflect the mission of education in democratic societies.

The purpose of this document is to provide a set of actionable, context-sensitive recommendations to support HEIs in implementing AI systems in a way that is ethically grounded and socially responsible. These recommendations are based on internationally recognized principles, including transparency, explainability, accountability, fairness, inclusiveness, and respect for human rights. Rather than offering abstract declarations, the document focuses on concrete policy adjustments, institutional practices, and governance mechanisms that operationalize these principles within academic settings.

Ethical implementation of AI in education requires careful attention to the unique characteristics of the learning environment. Students are not simply users of AI. They are subjects of educational processes that aim to foster critical thinking, personal growth, and civic engagement. Therefore, the design and use of AI tools must protect students' rights, support their development, and avoid creating new forms of surveillance, exclusion, or dependency. Faculty and staff, likewise, must be meaningfully involved in shaping institutional decisions about AI use and provided with the necessary training and support to engage with AI in ways that enhance rather than compromise educational values.

By following these recommendations, HEIs can build an ethical foundation for AI adoption that not only mitigates risks, but also reinforces public trust, strengthens academic integrity, and positions education as a space where innovation and ethics go hand in hand.

**Keywords:** Artificial Intelligence, Ethics, Education, Accountability, Bias, Autonomy, Labor, Privacy, Safety, Transparency, Tools, Practices.



# 1. Education Processes Fostered by AI Tools

Artificial Intelligence (AI) is reshaping the landscape of education by enhancing how learning is delivered, assessed, and supported. AI tools are no longer limited to back-end systems or experimental prototypes; they are increasingly embedded into everyday educational practices across disciplines and educational levels. These tools foster personalized learning, automate administrative and pedagogical tasks, and support both educators and learners in navigating complex academic environments more efficiently and meaningfully.

Drawing from Bernard Marr's framework in "Data Strategy", several key areas are identified where AI supports and transforms education. These use cases reflect how AI can become a powerful ally for teachers and institutions:

## **Differentiated and Individualized Learning**

AI enables a shift from the traditional "one-size-fits-all" approach toward truly personalized learning experiences. Intelligent tutoring systems and adaptive learning platforms can assess each student's needs, pace, and learning style, offering targeted content and feedback. This not only enhances student engagement but also allows educators to better support diverse learners, including those with special educational needs.

## **Universal Access in Global Virtual Learning Environments**

With the growing demand for online and hybrid education, AI tools are key to ensuring equitable access to learning opportunities. Language translation, speech-to-text, and real-time transcription technologies reduce linguistic and physical barriers. AI-powered platforms can facilitate communication and collaboration across borders, helping institutions to reach wider audiences and create inclusive digital learning spaces.

## **Automation of Administrative and Routine Educational Tasks**

AI systems can streamline a variety of administrative tasks, freeing educators to focus more on pedagogy and student interaction. Examples include grading multiple-choice assessments, scheduling, attendance tracking, and even detecting plagiarism. In higher education institutions, AI can also support curriculum planning, learning analytics, and student progress monitoring – contributing to data-informed decision-making.

By leveraging these capabilities, AI becomes more than just a technological innovation; it becomes a catalyst for rethinking educational practices and improving both teaching efficiency and student outcomes. However, to harness these advantages effectively and ethically, institutions must critically reflect on implementation strategies and ensure transparency, accountability, and human oversight remain at the core of AI integration in education.

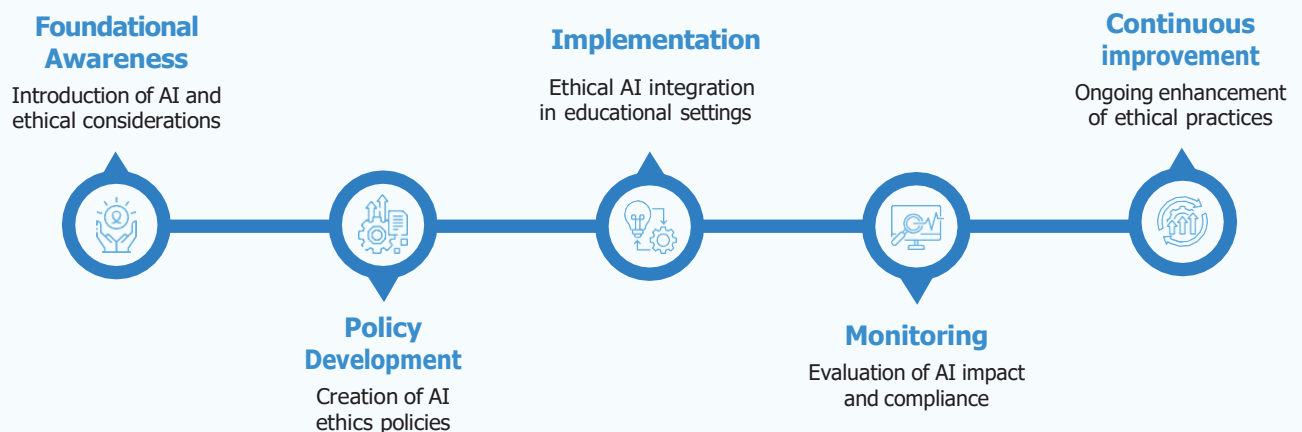


## 2. The Ethical Principles

As Wallach and Asaro (2017) point out, articulating the ethics of machines often remains a complex and ambiguous task. Ethical frameworks for artificial intelligence (AI) are still evolving, and the boundaries between human-centered values and machine-driven processes are not always clearly defined. In light of this, the recommendations provided in approach emphasize the application of AI tools and solutions in educational settings in alignment with established ethical norms. These recommendations are not focused on speculative future developments but instead aim to guide current practices through a value-driven perspective.

From a temporal standpoint, the suggested ethical framework does not assign specific timelines or milestones. Rather, it promotes a long-term, principled approach to integrating AI in education, recognizing that technological development outpaces regulatory adaptation. The emphasis lies on the continuous alignment of AI applications with ethical values such as transparency, accountability, fairness, and human dignity.

### Timeline of AI Ethical Maturity in Education



An important consideration in this context is the regulatory landscape shaped by the European Union's General Data Protection Regulation (GDPR) (Regulation 2016/679). It simultaneously plays a crucial role in establishing safeguards against the misuse of data, infringement of privacy, and potential manipulative control of individuals. AI systems, like any machinery or robots must be subject to fundamental ethical limitations. These include respect for autonomy, protection of individual rights, and the prevention of harm. Therefore, the ethical integration of AI in higher education institutions must not only address technical efficiency but also critically examine the societal, psychological, and legal implications of AI-assisted learning, assessment, and decision-making. Accordingly, AI should serve as a tool to enhance the core values of education, not compromise them. (Ethics of AI (Artificial Intelligence), Dhruvitkumar Talati, 2024).

## 21. Accountability and Responsibility

A primary ethical challenge in the integration of artificial intelligence (AI) in education lies in establishing clear lines of accountability when these systems produce adverse outcomes. In essence, the fundamental dilemma emerges when AI contributes to or directly causes unintended negative consequences, such as biased grading, inappropriate learning path recommendations, or exclusionary access to educational resources. In such cases, attributing responsibility becomes highly complex. Should the blame lie with the developers who designed the algorithm, the educational institution that implemented and relied on it, or the AI system itself as an autonomous agent? This phenomenon, often referred to as the “diffusion of responsibility,” creates a vacuum of accountability that can leave students, educators, and institutions without clear channels for redress or correction.

For example, if an AI-based assessment system unfairly downgrades students from marginalized backgrounds due to biased training data, or if a recommender system limits a student’s learning potential based on past performance rather than future capacity, it can have serious long-term consequences. Yet identifying who is accountable for these outcomes remains opaque. This lack of clarity undermines trust in AI systems and may discourage responsible adoption.

To better understand the broader picture, these ethical risks are often weighed against the potential benefits of AI in education, such as personalization, scalability, and efficiency. This contrast is effectively illustrated in a Benefit vs. Risk matrix, which helps stakeholders visualize the trade-offs between innovation and ethical responsibility.

## 22. Bias and Fairness

AI systems are fundamentally reliant on data for training, and this data often reflects the historical, social, and cultural biases present in the environments from which it is sourced. As a result, artificial intelligence does not operate in a vacuum. It absorbs and potentially amplifies these existing prejudices. This issue becomes particularly critical in the field of education, where AI tools are increasingly used to support decision-making, assessment, and personalized learning. When biased data underpins these systems, the consequences can be far-reaching and harmful.

For example, AI-based assessment tools may systematically underestimate the performance of students from underrepresented backgrounds, particularly if the training data lacks diversity or embeds discriminatory assumptions. Similarly, personalized learning systems, designed to adapt educational content to individual needs may reinforce existing achievement gaps by offering lower-level content or fewer opportunities for academic advancement to certain groups of students, based on flawed or incomplete data profiles.

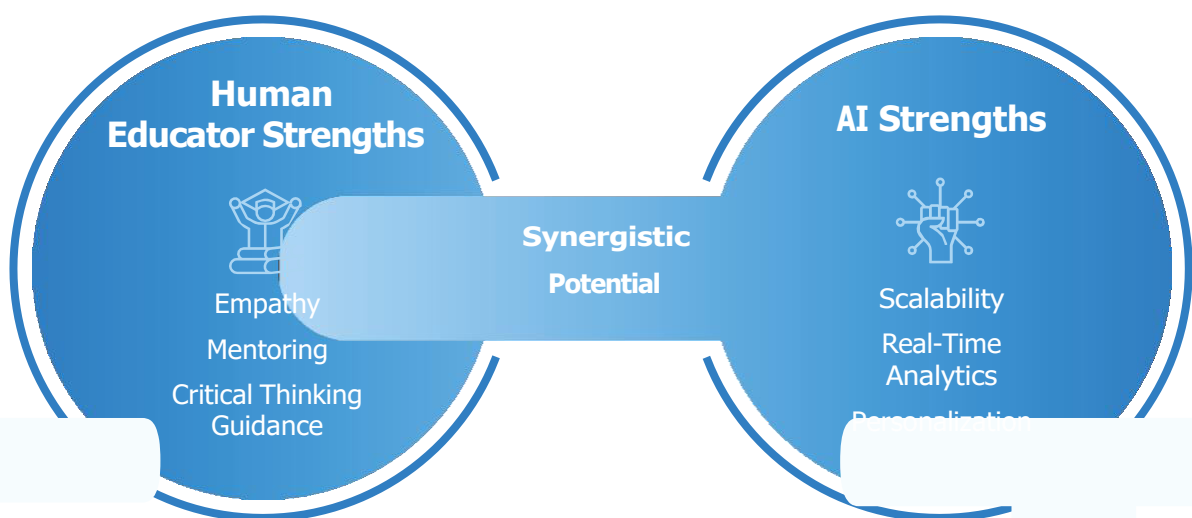
Researchers like Luckin and Selwyn (2016) have drawn attention to the ethical risks posed by these developments, especially in AI-driven assessment. They argue that such tools, if left unchecked, may reproduce social inequalities within educational systems by embedding bias into what are perceived as objective and data-driven processes. This undermines the promise of AI to democratize learning and instead contributes to a stratified system in which marginalized learners continue to face systemic barriers.

Addressing these challenges requires a proactive and ethical approach to AI design and deployment. This includes rigorous scrutiny at every stage of the AI development pipeline: from inclusive and representative data collection to the careful pre-processing of datasets to remove or mitigate bias; from transparent and explainable algorithm design to robust mechanisms for regular auditing and monitoring. It is essential to evaluate the impact of AI systems on different student groups continuously, with particular attention to those historically disadvantaged in educational settings. Only through such comprehensive efforts can we ensure that AI contributes to greater equity and fairness in education rather than reinforcing the very inequalities it seeks to overcome.

## 23. Human Autonomy and Agency

AI systems in education have the potential to subtly influence students' learning paths, career choices, and even their understanding of themselves. Recommender systems for learning resources or AI-driven career guidance tools, while seemingly beneficial, can inadvertently narrow students' horizons and undermine their independent decision-making. This echoes broader concerns about AI's impact on human autonomy, where algorithms can shape opinions and behaviors in ways that erode individual agency. Maintaining student autonomy requires transparent AI systems that empower learners with information about how recommendations are generated and provide them with genuine choices.

The integration of AI in education also raises questions about the future of the teaching profession.



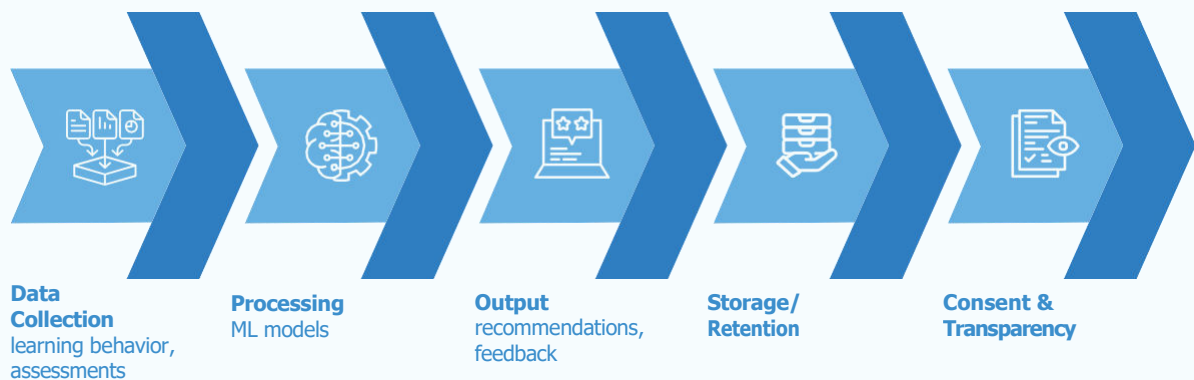
While AI can automate certain tasks, such as grading or administrative duties, concerns exist regarding the potential displacement of teachers and the erosion of the crucial human element in education. Selwyn (2017) explores this potential displacement, emphasizing the unique role of educators in fostering critical thinking, social-emotional development, and personalized guidance – aspects that current AI systems struggle to replicate fully. Ethical considerations must guide the implementation of AI in ways that augment, rather than replace, the vital role of educators to keep maintaining student agency within AI-enhanced learning environments.

## 24. Privacy and Data Protection

The integration of artificial intelligence in education brings substantial benefits, including personalized learning and predictive analytics. However, these systems depend heavily on the continuous collection, storage, and analysis of vast volumes of student data. This data may include academic records, behavioral patterns, emotional responses, and even sensitive personal information such as demographic background or mental health indicators. Such extensive data processing raises serious ethical and legal concerns about privacy and data protection. One of the most pressing issues is the potential misuse of data, whether through unauthorized access, data breaches, or repurposing data for commercial or surveillance purposes. There is also the risk of creating detailed student profiles that could lead to bias, discrimination, or long-term reputational consequences for learners.

To address these challenges, it is essential to implement strong data protection frameworks and ethical governance structures. This includes enforcing strict access controls, anonymizing data where possible, and ensuring data minimization principles are followed. AI systems in education must be designed with transparency in mind. Students, parents, and educators need to understand how data is collected, for what purposes it is used, and who has access to it.

### Data Flow & Privacy Lifecycle Schema



Equally important is the requirement for informed consent. Learners and their guardians must be fully aware of the implications of data collection and provide clear, voluntary agreement before their data is used. Educational institutions must also ensure compliance with national and international regulations such as the General Data Protection Regulation (GDPR) and promote digital literacy to empower users to make informed decisions.

Ultimately, safeguarding student privacy is not just a legal obligation but a moral imperative. Ethical data use fosters trust in educational technologies and ensures that AI serves as a tool for empowerment rather than control.

## 25. Safety and Security

The integration of artificial intelligence into educational environments brings not only new opportunities but also serious responsibilities, particularly when it comes to safety and security. Ensuring the reliability and security of AI systems in education is of critical importance, as these systems increasingly influence teaching practices, administrative decisions, and student learning outcomes.

A malfunction in an AI-powered learning platform can do more than just cause technical inconvenience. It can disrupt the entire learning process, delay educational progress, and erode trust in digital solutions. For example, an error in an adaptive learning system may result in misaligned content delivery or incorrect feedback, thereby hindering a student's academic development.

Even more concerning is the potential for malicious misuse. As AI systems process vast amounts of sensitive data, such as academic records to personal student profiles, they become attractive targets for cyberattacks. Unauthorized access or manipulation of AI algorithms can lead to biased assessments, privacy violations, or even intentional sabotage of academic results. Such scenarios not only jeopardize individual students but also undermine the credibility of educational institutions.

To address these challenges, robust safeguards must be implemented at every stage of AI system development and deployment. This includes rigorous testing before implementation, the application of strong cybersecurity protocols, real-time monitoring, and rapid-response mechanisms for identifying and mitigating risks. Ethical oversight, transparency in algorithmic processes, and ongoing collaboration between IT experts, educators, and policymakers are also key to building AI systems that are not only effective but also secure and trustworthy.

By prioritizing safety and security, educational institutions can foster a digital learning environment where innovation is supported by integrity, and technology serves as a reliable partner in the pursuit of knowledge.

## 26. Inclusivity

In the context of education, particularly within Higher Education Institutions, the principle of inclusiveness requires that the design, deployment, and governance of AI systems actively promote equity, diversity, and accessibility. Inclusiveness is not just about avoiding discrimination. It is also about creating systems that recognize, accommodate, and support the rich diversity of student experiences, backgrounds, and learning needs.

To uphold this principle, HEIs should take deliberate steps to ensure that AI does not perpetuate or exacerbate existing inequalities by preventing algorithmic bias that may negatively affect students based on race, gender, disability, language proficiency, or socio-economic background, ensuring that training data used by AI systems is representative and that outcomes are regularly audited to detect and mitigate any form of discrimination; supporting diverse learning styles and educational needs by leveraging AI tools that enable personalized learning, adaptive content delivery, and flexible assessment formats, so that educational processes are tailored to students' strengths, pace, and contexts; engaging

diverse stakeholders, including students, staff, and representatives of marginalized or underrepresented groups, in the co-creation and monitoring of AI policies, as participatory governance increases the social legitimacy of AI use, encourages trust, and helps institutions anticipate ethical risks from multiple perspectives; and expanding accessibility through the use of inclusive AI technologies such as real-time transcription, automatic translation, and content customization, which can reduce barriers for students with disabilities, language learners, or others who may face challenges in traditional academic environments.

Inclusiveness also means recognizing systemic barriers that may affect certain students' ability to benefit from AI technologies equally, such as lack of digital skills, limited access to devices, or cultural differences in engagement with technology. HEIs must therefore accompany AI integration with targeted support measures, such as training, mentoring, and digital literacy programs.

## 27. Transparency and Explainability

As artificial intelligence becomes increasingly integrated into higher education, ensuring transparency and explainability in AI systems has emerged as a critical ethical and practical priority. Many advanced AI models, particularly those based on deep learning, function as complex "black boxes," where the internal decision-making processes are not readily accessible or understandable to users. This opacity can pose significant challenges in educational settings, where trust, accountability, and fairness are foundational values.

In learning environments, AI is often used to support or even automate key processes such as student assessment, personalized learning recommendations, admissions decisions, and academic advising. When AI systems are involved in making high-stakes or sensitive decisions, the inability to trace and understand the rationale behind a given output undermines confidence in these tools. Students and educators may question the validity or fairness of recommendations, particularly if they cannot see or challenge the logic that produced them.

Furthermore, lack of explainability impedes the ability to identify and correct errors or biases embedded in algorithms. For example, if a student receives an unfavorable evaluation from an AI-based system, both the student and instructor must be able to understand how that outcome was reached. Without transparency, there is little room for appeal, learning, or improvement - both for the AI system itself and for the human users it is intended to support.

Therefore, developing AI tools with transparent mechanisms and explainable outputs is essential not only for ethical compliance, but also for enhancing user engagement, improving outcomes, and ensuring that AI technologies serve as partners rather than opaque authorities in the learning process. Increasing research efforts are being directed toward creating explainable AI (XAI) models, which aim to make algorithms more interpretable without significantly compromising their performance. Integrating such models in higher education can help build a culture of trust, ensure informed consent, and empower users with greater autonomy in their learning journeys.



### 3. Implementing AI Ethically in Higher Education Institutions: Key Recommendations

This section outlines practical and policy-oriented recommendations for the integration of Artificial Intelligence (AI) within Higher Education Institutions (HEIs) in a way that aligns with internationally recognized ethical standards, particularly those set by UNESCO in its 2023 guidance. As AI technologies increasingly influence academic processes, from teaching and assessment to administration and research, it is essential that their adoption be guided by principles that safeguard human dignity, academic freedom, equity, and institutional responsibility.

The recommendations presented here provide a structured roadmap to help HEIs adopt AI systems that are transparent, inclusive, fair, and respectful of privacy and autonomy. They focus on establishing institutional governance mechanisms, aligning AI practices with data protection laws, addressing algorithmic bias, promoting human oversight, and integrating ethical awareness into academic culture. Each recommendation is directly linked to core ethical domains ensuring that the use of AI strengthens, rather than undermines, the values of higher education.

By following these recommendations, HEIs can build trust, mitigate risks, and ensure that AI serves as a tool for innovation and social good by empowering educators, students and communities alike.

## 3.1. Establishing an Ethical Governance Structure for AI

To ensure the ethical deployment of Artificial Intelligence (AI) within Higher Education Institutions (HEIs), it is essential to establish a robust and transparent governance structure that clearly defines roles, responsibilities, and oversight mechanisms. Ethical governance serves as the foundation for accountable AI use, ensuring that human agency is preserved and that decision-making remains traceable, justified, and aligned with institutional values.

**Linked principle:** Accountability and Responsibility

### 3.1.1 Forming an AI Ethics and Governance Committee

The ethical integration of Artificial Intelligence (AI) in Higher Education Institutions (HEIs) begins with a clear and robust governance structure, one that places humans at the center of decision-making and ensures that institutions remain accountable for the design, deployment, and consequences of AI systems.

At the heart of this structure should be the formation of an AI Ethics and Governance Committee, a multidisciplinary, inclusive body responsible for overseeing the ethical dimensions of AI usage across the institution. This committee provides strategic direction, ensures adherence to international ethical standards (e.g., UNESCO's 2023 AI ethics framework), and fosters trust among students, staff, and external stakeholders.

## Composition of the AI Ethics and Governance Committee

To uphold accountability in both decision-making and risk oversight, the committee must reflect a diversity of roles and expertise. It ensures shared responsibility by including academic, technical, legal, student, and community perspectives in the governance process. This multidisciplinary approach guarantees that decisions related to AI are made transparently, equitably, and with full awareness of social impact, preventing responsibility from being centralized in one unit or delegated to automated systems.

The recommended composition could include the following members:

**Academic Staff from diverse disciplines.** Representatives from fields such as Artificial Intelligence, Ethics, Education, Law, Social Sciences, and Philosophy, who contribute interdisciplinary insights into the ethical, legal, and pedagogical implications of AI in higher education.

**Student Representatives.** Participants from various study programmes and backgrounds, ensuring that learners' voices are heard—particularly in relation to academic rights, fairness in automated decision-making, and digital inclusion.

**Legal and Data Protection Experts.** Professionals with expertise in data governance, GDPR compliance, and institutional legal frameworks, ensuring AI systems meet all relevant ethical and legal requirements.

**IT and AI Systems Administrators.** Technical staff who oversee the deployment, maintenance, and auditing of AI tools within the institution. Their presence ensures that system functionalities and limitations are ethically reviewed and technically explained.

**Representatives of Marginalized or Underrepresented Groups.** Individuals advocating for students and staff with disabilities, ethnic minorities, or other vulnerable populations. Their inclusion promotes fairness, equity, and social justice in AI decision-making processes.

**External Advisors.** Experts from civil society, the public sector, or private industry with a background in digital ethics, AI policy, or human rights. These members contribute an external, unbiased perspective and provide benchmarking against national or international standards.

## Core Responsibilities of the AI Ethics and Governance Committee

The AI Ethics and Governance Committee plays a central role in operationalizing ethical principles across the lifecycle of Artificial Intelligence (AI) deployment within Higher Education Institutions (HEIs). Its work ensures that AI systems used in teaching, research, administration, and student services align with core values.

The following key responsibilities outline the committee's scope of action and institutional value:

### Develop Ethical AI Guidelines

1.

Establishing clear, institution-wide guidelines ensures proactive responsibility for the development and application of AI systems. These standards define what the institution permits or prohibits, and hold departments accountable for compliance, creating a normative framework that anchors accountability in practice.

## **Review and Approve AI Systems**

**2.**

By assessing AI tools before deployment, the committee guarantees that no AI system enters institutional use without ethical and legal evaluation. This function ensures institutional responsibility is asserted upfront, and potential harms or violations of academic values are prevented in advance.

## **Ensure Human Oversight**

**3.**

Defining the boundaries of AI autonomy and the mandatory involvement of humans in critical decisions ensures that institutions do not abdicate responsibility to algorithms. AI becomes a tool, not a decision-maker, upholding the principle that accountability must always rest with human actors.

## **Monitor, Audit, and Respond**

**4.**

The committee takes responsibility for ongoing supervision of AI systems, maintaining ethical standards and responding to issues post-implementation. This reinforces the idea that accountability is continuous, not a one-time approval process.

## **Act as a Redress Mechanism**

**5.**

Providing a transparent and accessible process for raising concerns ensures the institution is accountable for any negative consequences resulting from AI use. The presence of a redress mechanism formalizes institutional responsibility toward students and staff, protecting their rights and interests.

## **Promote AI Ethics Literacy**

**6.**

Educating the academic community about AI risks, responsibilities, and ethical standards distributes awareness and accountability across the institution. It empowers individuals to act responsibly and reinforces a shared culture of ethical integrity.

The establishment of an AI Ethics and Governance Committee is the institutional embodiment of accountability and responsibility in the era of AI. By assigning clear roles, ensuring human oversight, and implementing continuous evaluation and redress procedures, HEIs uphold their ethical and legal obligations while fostering trust and transparency. In alignment with UNESCO's (2021) ethical vision, this approach ensures that AI serves the public good under human and institutional control, never operating beyond the bounds of democratic oversight and moral responsibility.

### **3.12 Ensuring Human Accountability in AI Decisions**

One of the most fundamental aspects of ethically integrating Artificial Intelligence (AI) into Higher Education Institutions (HEIs) is establishing clear and enforceable accountability mechanisms that ensure humans, not algorithms remain ultimately responsible for all decisions made or assisted by AI systems. While AI can process data, support decision-making, and automate administrative or academic tasks, legal, ethical, and professional responsibility must remain with identifiable human actors and governing bodies.

Institutions must clearly assign responsibility for AI-supported processes, embedding these roles within job descriptions, committee mandates, and internal regulations. Decision-making must be transparent and traceable, with systems that log how AI decisions are made, what data they rely on, and who holds ultimate responsibility for review and approval. This traceability not only ensures that decisions can be audited or explained but also allows individuals to appeal AI-driven outcomes when necessary.

Particularly in high-stakes contexts such as admissions, grading, and staff evaluations, meaningful human oversight must be mandatory. AI should never fully replace human judgment in areas where fairness, ethics, and individual circumstances matter. Institutions should also establish structured mechanisms to receive and resolve complaints, including clear channels for appeals and protections for those who report ethical concerns.

Moreover, before any AI system is implemented, it should undergo a thorough ethical risk assessment, such as a Data Protection Impact Assessment or a broader ethical review, approved by a dedicated institutional body. This reinforces institutional responsibility and ensures no system is adopted without critical reflection. Accountability is further supported through routine internal and external reporting, which monitors the performance of AI tools, tracks complaints or risks, and demonstrates responsiveness to any emerging harms or biases.

By embedding responsibility into every stage of the AI lifecycle, from the early stages of design and data selection, through development, piloting, and institutional deployment, to long-term monitoring and human oversight, Higher Education Institutions send a clear and powerful message that Artificial Intelligence must always serve, and never override, human values, professional ethics, and the institution's educational mission.

### **3.13 Aligning AI Strategies with Core Ethical Principles**

An essential aspect of building ethical governance for AI in Higher Education Institutions (HEIs) is the alignment of institutional AI strategies with globally recognized ethical standards, particularly those outlined in UNESCO's 2023 Recommendation on the Ethics of Artificial Intelligence. These principles - human rights, transparency, sustainability, and social inclusion must not remain aspirational; they must become operational foundations that guide how institutions adopt, deploy, and monitor AI technologies.

This alignment reinforces accountability, as it requires institutions to clearly articulate who is responsible for ensuring that AI systems do not violate fundamental rights such as dignity, autonomy, and non-discrimination. It also demands that institutions define internal responsibilities for evaluating whether AI-supported practices align with social goals (e.g., equitable access to education) and do not reinforce inequalities or systemic biases.

By embedding these ethical standards into strategic documents, such as AI roadmaps, procurement policies, assessment frameworks, and staff training programs HEIs create institutional obligations rather than vague commitments. For example, implementing transparency as a guiding principle involves not only documenting how AI systems function but also naming the individuals or bodies accountable for explaining AI-based decisions to affected students or staff. Human rights alignment requires institutions to put in place impact assessments and redress mechanisms when AI affects learners' access to services or opportunities.

Responsibility must also be visible in how sustainability is treated, not only environmental sustainability (e.g., energy efficiency of AI infrastructure) but also long-term social sustainability, such as how AI shapes labor, pedagogy, and institutional equity. Similarly, operationalizing social inclusion means institutions must be accountable for evaluating whether their AI tools are accessible to all students, including those from underrepresented or marginalized communities.

Ultimately, aligning AI strategies with UNESCO's core ethical principles transforms responsibility from a reactive duty into a proactive institutional commitment. It ensures that AI is not merely managed by technical teams or isolated experts but is governed by a shared ethical vision that holds institutions collectively and visibly accountable to the public, to students, and to society.

### **3.14 Developing and adopting institution-wide AI use policies**

A foundational element of ethical AI governance in HEIs is the creation of clear, enforceable, and institution-wide policies that define acceptable and prohibited uses of AI. These policies serve as a formal expression of institutional accountability, establishing who is responsible for the use of AI, under what conditions, and with what safeguards.

Such policies must go beyond general ethical declarations and instead provide practical, operational guidance. This includes clearly defined permitted use cases, such as AI-assisted tutoring, administrative automation, or predictive analytics for student support, as well as explicitly prohibited applications, for example, automated decision-making in admissions without human oversight, or opaque AI systems used in staff performance evaluation.

By developing these policies, institutions demonstrate responsibility not only in designing AI use but also in foreseeing risks and preventing misuse. These documents must outline the obligations of staff and faculty when engaging with AI tools, ensure that human review is required in high-stakes contexts, and specify mechanisms for reporting, auditing, and remedying ethical breaches.

Moreover, these policies must be transparent and inclusive in their development, involving input from legal experts, students, academic staff, and marginalized groups to ensure they reflect both legal compliance and institutional values. Once adopted, they must be monitored and regularly updated, with designated personnel or committees held accountable for policy enforcement and alignment with international frameworks, such as UNESCO's principles on AI and human rights.

By formalizing AI use in this way, HEIs embed ethical considerations into everyday academic and administrative processes, ensuring that AI systems are not used arbitrarily or

irresponsibly. Instead, they become tools governed by clear norms, institutional obligations, and public trust, where accountability is both traceable and enforceable.

## Key Components of Effective AI Use Policies

**Scope and purpose.** Clearly state where AI can be used (e.g., admissions, grading, HR, administrative automation) and define its intended benefits and risks.

**Permissible and prohibited use cases.** Specify which AI applications are acceptable, conditionally allowed (e.g., with human oversight), or prohibited altogether.

**Role assignment.** Define who is responsible for approving, maintaining, and monitoring AI tools at the institutional and department levels.

**Transparency requirements.** Mandate that staff and students be informed when AI tools are used in processes that affect them.

**Appeal and redress mechanisms.** Ensure that individuals have clear pathways to contest decisions made or supported by AI.

Institution-wide AI use policies are essential governance tools that enshrine accountability and uphold institutional integrity in the digital age. By clearly defining the scope, limitations, and ethical expectations of AI systems, HEIs ensure that the deployment of AI does not compromise human judgment, academic values, or legal responsibilities. These policies provide a structured framework through which responsibilities are assigned, risks are managed, and trust is built among students, faculty, and administrative staff. Ultimately, such policies turn abstract ethical principles into actionable safeguards, guaranteeing that AI serves institutional goals while respecting the rights and dignity of all members of the academic community.

Establishing a strong ethical governance structure is the cornerstone of responsible AI integration in higher education. By forming an inclusive AI Ethics and Governance Committee, aligning institutional strategies with international ethical standards such as UNESCO's 2023 framework, and adopting clear, enforceable AI use policies, Higher Education Institutions create a robust foundation for accountability, transparency, and human-centered oversight. These structures ensure that the deployment of AI supports educational values rather than undermines them by keeping human agency, institutional responsibility, and social trust at the core of every AI-related decision. Ultimately, ethical governance is not a one-time action, but a continuous institutional commitment to using AI in ways that serve both academic integrity and the public good.

## 3.2 Safeguarding Privacy in AI Use: Responsible Data Practices and Institutional Readiness

As Higher Education Institutions integrate Artificial Intelligence into teaching, research, and administration, they increasingly process large volumes of personal, sensitive, and behavioral data. This makes the ethical management of data privacy and protection not only a legal necessity but also a critical pillar of institutional trust and student safety. Ensuring privacy in AI use reflects the institution's commitment to safeguarding individual rights, academic freedom, and data security in a digital learning environment.

AI systems used in education, such as learning analytics platforms, automated grading tools, or student advising bots often rely on sensitive personal data like academic records, health information, browsing behavior, or even biometric data. If left unregulated, these tools may expose students and staff to data breaches, misuse of personal information, or surveillance-like environments that undermine freedom of thought and expression.

To prevent such risks, HEIs must embed data protection into every stage of AI system design and deployment, from procurement to implementation and long-term use. A proactive, privacy-by-design approach ensures that data minimization, informed consent, and secure storage are prioritized from the outset.

**Linked principle:** Privacy and data protection

### 321 AI in Education: Ensuring Privacy and Data Minimization

As Higher Education Institutions increasingly integrate Artificial Intelligence into their academic, administrative, and research environments, they face a critical responsibility: to safeguard the privacy of students, staff, and stakeholders. AI systems often rely on vast amounts of personal data to function effectively, but the ethical and legal risks of misuse, overcollection, or un-transparent processing are significant and growing.

This is why privacy-by-design must be at the heart of every AI deployment in education. Rather than treating privacy as a technical patch or an after-the-fact compliance task, this principle demands that privacy be embedded into the architecture, configuration, and operation of AI systems from the earliest stages of development. It involves anticipating potential harms, minimizing data exposure, and designing for user consent, data transparency, and secure storage as default, not optional - features.

Data minimization, a closely aligned principle, ensures that only the information strictly necessary for a defined educational purpose is collected and processed. For instance, an AI tool developed to recommend learning resources should not access a student's disciplinary history or health records. By limiting the scope of data, institutions not only reduce the risk of breaches or ethical violations but also enhance data governance and system efficiency.

In practice, these principles translate into several actionable measures:

- **Conducting Privacy Impact Assessments (PIAs)** or Data Protection Impact Assessments (DPIAs) prior to the adoption of any AI system, especially those involving automated decision-making or sensitive data (e.g., health, ethnicity, behavior).
- **Applying data anonymization and encryption techniques** wherever possible to protect identities and reduce the risks in the event of a breach.
- **Using role-based access controls** to ensure that only authorized personnel can view or manipulate personal data, and only within the scope of their responsibilities.
- **Setting strict data retention policies**, so that personal data is not stored longer than necessary or used beyond its original purpose.
- **Ensuring user consent and transparency**, giving students and staff clear information about what data is collected, how it is used, and how they can manage or withdraw their data rights.

The importance of these practices cannot be overstated. In an educational setting, privacy violations can erode trust, chill academic expression, and disproportionately affect vulnerable or marginalized populations. If learners feel constantly monitored, profiled, or dehumanized by data-driven systems, they may become disengaged or feel unsafe participating fully in their education.

Moreover, legal frameworks such as the General Data Protection Regulation (GDPR) in Europe, or equivalent laws in other regions, place specific duties on HEIs to ensure lawful, fair, and transparent data processing. Violations can result in not only financial penalties but also severe reputational damage—particularly in institutions that claim to uphold human rights and academic freedom.

By institutionalizing privacy-by-design and data minimization, HEIs demonstrate that AI adoption does not come at the cost of individual rights. Instead, they affirm that ethical innovation is possible, where technology amplifies educational quality without compromising human dignity. This approach also empowers individuals by ensuring that their data is treated respectfully and that their participation in digital systems is based on informed choice rather than passive submission.

### **322 Conducting Data Protection Impact Assessments (DPIAs) before deploying AI tools**

As Artificial Intelligence systems become more integrated into higher education institutions, their ability to collect, process, and analyze vast amounts of personal and sensitive data introduces significant privacy risks for a broad range of individuals, not only students, but also academic staff, administrative personnel, researchers, applicants, external partners, and alumni.

AI tools used for automated decision-making, behavior prediction, learning analytics, human resources, or research management often handle data such as grades, health information, personal communications, performance metrics, or behavioral patterns. The use of such data must be governed with utmost care to preserve privacy, dignity, and legal compliance.

To meet these standards, HEIs must conduct Data Protection Impact Assessments (DPIAs) before deploying any AI systems that involve the processing of personal or sensitive information. A DPIA is a formal process designed to ensure that the use of AI systems complies with legal and ethical standards related to personal data protection. Particularly relevant in higher education contexts where AI technologies intersect with sensitive data, the DPIA enables institutions to proactively identify risks and design safeguards that protect the rights and freedoms of all individuals whose data may be processed. This includes students, staff, researchers, applicants, and external partners.

The implementation of AI systems in Higher Education Institutions inevitably affects multiple stakeholder groups beyond students, each with distinct data protection risks. Conducting DPIAs ensures that potential harms are identified and mitigated before the deployment of AI technologies, promoting a culture of ethical accountability, compliance with data protection regulations, and institutional trust.

## For Students

Students generate and share vast quantities of data throughout their academic journey, including grades, attendance, LMS interactions, behavioral patterns, and sometimes health or psychological information. AI systems used for academic analytics, behavioral prediction, or personalized learning paths must be assessed to ensure:

- ➡ Academic records are not repurposed for unintended uses (e.g., behavioral surveillance or marketing).
- ➡ Students are not subject to profiling or stereotyping based on socioeconomic background, nationality, gender, or past performance.
- ➡ Privacy violations do not compromise student well-being, mental health, or trust in the institution.

DPIAs help prevent discriminatory or opaque algorithmic decisions, especially in high-stakes areas like grading, disciplinary actions, or access to support services.

## For Academic and Administrative Staff

Staff data, including contracts, performance evaluations, workload indicators, communication logs, and professional development records can be used in AI-supported human resource management systems. DPIAs ensure that:

- ➡ Recruitment algorithms do not exhibit hidden biases or violate equal opportunity principles.
- ➡ Performance monitoring tools respect privacy boundaries and do not foster surveillance cultures.
- ➡ Decision-making regarding promotions, workloads, or discipline remains transparent and includes human oversight.

This safeguards employee dignity, reinforces fairness, and aligns with labor law and ethical employment practices.

## For Researchers

AI applications in research (e.g., data mining, participant recruitment, lab management) often process sensitive data, especially in disciplines like health sciences, psychology, and social research. DPIAs provide a critical framework to:

- Protect the privacy of research participants, particularly in studies involving minors, vulnerable populations, or stigmatized topics.
- Ensure that AI tools used for data analysis or prediction comply with research ethics and institutional review board (IRB) protocols.
- Avoid unauthorized data sharing with third-party platforms or commercial AI vendors.

This promotes responsible research conduct and preserves academic freedom while meeting legal and ethical standards.

## For Applicants (Prospective Students and Staff)

AI systems used in admissions or hiring, especially those powered by predictive analytics or automated assessments carry the risk of reproducing structural inequalities or reinforcing unconscious bias. DPIAs help institutions:

- Ensure transparency in how application data is processed and scored.
- Prevent algorithmic discrimination based on age, gender, nationality, or disability.
- Maintain procedural fairness by requiring human review of AI-generated recommendations.

By applying DPIAs in these contexts, HEIs demonstrate that technological innovation does not come at the expense of justice, equity, and privacy.

By systematically applying Data Protection Impact Assessments (DPIAs) across all domains where AI interacts with personal or sensitive data, Higher Education Institutions not only fulfill regulatory obligations but also embed a culture of ethical vigilance and human-centered accountability. Whether safeguarding students from algorithmic profiling, protecting staff from surveillance-driven management, preserving the integrity of research, or ensuring fairness in admissions and hiring—DPIAs function as a cornerstone of responsible AI governance. This proactive approach reinforces institutional trust, promotes compliance with data protection principles, and affirms the university's role as a steward of both technological advancement and fundamental rights.

### 323 Training on Data Rights and GDPR Compliance

As Artificial Intelligence (AI) systems become increasingly integrated into academic, administrative, and research environments, Higher Education Institutions (HEIs) must recognize that data literacy is not merely a technical or legal necessity as it is a fundamental ethical obligation. AI technologies used in learning analytics, admissions algorithms, HR automation, and research tools often collect, process, and analyze vast amounts of personal and sensitive data. In such contexts, the ability of institutional stakeholders, including students, academic staff, researchers, and administrators to understand and exercise their data rights is essential for preserving human dignity, autonomy, and institutional trust.

Providing targeted, ongoing education on the GDPR and relevant national legislation equips individuals to critically assess when and how their personal data is being used, stored, and shared. This includes understanding their rights to access, rectify, and erase personal data; the right to object to automated processing and profiling; and the right to transparency in how AI-powered systems arrive at decisions that affect them.

When staff and students are empowered to question data practices, request explanations, and seek redress for misuse, they become active participants in shaping an ethical data culture. This shift from passive data subjects to informed rights-holders strengthens institutional accountability and reinforces the principle that responsibility for ethical AI use lies not only with developers or administrators, but is shared across the academic community.

Moreover, fostering data rights awareness helps prevent common ethical pitfalls, such as over-surveillance, unauthorized profiling, and non-consensual data reuse. It creates a proactive defense against reputational and legal risks while also supporting a more inclusive and democratic approach to digital transformation in higher education.

When students, faculty, and staff are equipped with a clear understanding of their data rights and protections, HEIs:

- **Promote informed consent and individual autonomy** in interactions with AI systems. Users are more likely to trust and responsibly engage with digital tools when they know their choices matter and their data is handled respectfully.
- **Reduce the risk of unethical or illegal data practices** by building literacy around core data protection principles such as data minimization, purpose limitation, and the right to be forgotten.
- **Encourage a culture of accountability** across all institutional levels, where data is not seen as a limitless resource but as a responsibility. Everyone from developers and decision-makers to students and administrators becomes an active steward of ethical data use.

#### Effective Training on Data Rights

Training on data rights and privacy within Higher Education Institutions must extend well beyond regulatory compliance or technical checklists. It should be conceived as a transformative educational process that cultivates a deep understanding of data ethics, individual rights, and institutional responsibilities in the context of emerging AI technologies.

As AI systems become embedded in core academic, administrative, and research functions from predictive learning analytics to AI-assisted grading and automated HR systems, every stakeholder within the institution must be equipped to understand not only how these systems function, but also how they affect personal autonomy, fairness, and dignity.

Effective training empowers students, faculty, researchers, and administrative staff by:

**Raising awareness of their personal data rights**, including the right to be informed, access data, contest decisions, and demand accountability.

**Fostering critical thinking**, about the ethical implications of data use in AI-powered systems, especially in areas involving evaluation, selection.

**Encouraging proactive engagement**, with institutional processes, such as seeking redress, filing complaints, or participating in policy development.

**Creating a shared culture of transparency and trust**, where all actors understand the value of consent, fairness, and respect for human agency.

Rather than viewing data protection as a niche legal requirement, HEIs should embed privacy training into broader institutional development strategies linking it with digital literacy, academic integrity, and responsible innovation. When delivered effectively, such training builds an informed academic community that can confidently navigate and shape ethical digital transformation.

## **These are essential components of impactful and comprehensive training programs:**

### **1. GDPR Fundamentals**

Training should begin with a clear overview of core GDPR principles, which underpin ethical and legal data use in AI-driven activities within Higher Education Institutions. Lawfulness, fairness, and transparency ensure that data is collected and used in ways that are legally justified, ethically sound, and clearly communicated. Purpose limitation and data minimization require that only necessary data is collected for defined objectives. Storage limitation ensures data is not retained longer than needed, while integrity and confidentiality emphasize strong safeguards against unauthorized access or breaches. These principles are key to evaluating the fairness and responsibility of AI tools in academic environments.

### **2. Individual Rights under GDPR**

Participants should understand how to exercise their GDPR data rights, which are essential in AI-driven academic settings. These include the right to access, correct, delete, or restrict the use of their personal data, as well as the rights to object to processing and request data portability. Training should connect these rights to university contexts like academic records, learning analytics, and staff evaluations so that all community members feel empowered to manage their data and challenge unfair or unlawful uses.

### 3. AI-Specific Contexts

Training should connect GDPR principles to real-world AI applications in higher education. Participants need to understand how algorithmic decision-making is used in grading, admissions, or staff evaluation, and why human oversight is essential. They should also learn the risks of automated profiling, especially when used to predict academic performance or behavior. In addition, learning analytics tools that process LMS data, attendance, or engagement must be assessed for fairness, transparency, and privacy. Linking these practices to GDPR safeguards equips staff and students to evaluate AI tools critically and responsibly.

### 4. Practical Scenarios and Simulations

To complement theoretical knowledge, training should include interactive elements such as role-playing exercises where students practice requesting access to their data or challenging AI-generated decisions. Case studies on issues like unauthorized data reuse or profiling can prompt critical thinking, while group discussions encourage reflection on fairness, consent, and transparency. These activities enhance not only understanding but also ethical awareness in real-life situations involving AI.

### 5. Clear Institutional Procedures

Clear institutional procedures are essential to ensure that all community members can act on their data rights. Training should explain how to contact the Data Protection Officer (DPO), file internal appeals against AI-influenced decisions, and raise concerns safely, including whistleblower protections. Participants should also know where to find key documents like the university's AI Use Policy and Privacy Policy. This helps turn awareness into action and supports a culture of transparency and accountability.

Training staff and students on data rights is a cornerstone of responsible and accountable AI use in higher education. By equipping all members of the academic community with the knowledge to understand, question, and exercise their rights under the GDPR and national laws, institutions foster a culture of transparency, trust, and ethical integrity. This empowers individuals to actively participate in safeguarding their privacy and ensures that AI technologies are implemented with full respect for human dignity and institutional responsibility.

In an age where AI systems increasingly rely on vast amounts of personal and sensitive data, Higher Education Institutions bear a critical responsibility to protect the privacy and rights of all members of their academic community. By implementing privacy-by-design principles and data minimization practices, institutions embed protection into the very structure of AI technologies. Through the systematic use of Data Protection Impact Assessments (DPIAs), they anticipate and mitigate risks before harm can occur. And by training staff and students on their data rights under the GDPR and national laws, HEIs empower individuals to actively participate in ethical data governance. Taken together, these actions form a comprehensive approach to privacy and data protection, one that ensures AI use remains accountable, transparent, and aligned with the values of trust, autonomy, and institutional integrity.

## 33. Promoting Fairness and Preventing Bias in AI-Supported Decision-Making

To implement Artificial Intelligence ethically in Higher Education Institutions, it is essential to ensure that AI systems operate fairly, without reinforcing discrimination or structural inequalities. Promoting fairness and preventing bias means that all individuals, regardless of gender, ethnicity, disability, socioeconomic status, or other characteristics must be treated equitably in any AI-supported decision-making process. AI technologies used in admissions, grading, student support, or hiring processes can unintentionally perpetuate historical biases embedded in data or algorithm design. Therefore, HEIs must take proactive steps to detect and correct such risks.

**Linked principle:** Bias and fairness

### 33.1 Auditing AI for Bias in Key Academic Systems

Regular audits of AI systems are a critical safeguard for ensuring fairness and mitigating bias in higher education contexts. AI technologies are not inherently neutral. They are shaped by the data they are trained on, the assumptions built into their algorithms, and the human decisions behind their design and deployment. As a result, they may unintentionally reinforce systemic inequalities or produce discriminatory outcomes.

**Example:**

an AI system used to grade essays might systematically disadvantage non-native speakers due to language patterns not reflected in its training data. A recruitment algorithm may prioritize certain profiles based on historical hiring patterns, unintentionally excluding qualified candidates from underrepresented groups. Predictive tools used for student support may flag certain students as “at-risk” based on socioeconomic indicators that reflect broader societal inequalities rather than individual needs.

### The Importance of Regular Auditing for Bias Mitigation:

- **Detect bias early:** Continuous evaluation helps identify when and how bias enters the system whether from skewed training data, flawed algorithmic logic, or unequal system performance across different groups.
- **Improve system reliability:** Auditing builds trust in AI systems by ensuring their outputs are consistent, explainable, and equitable over time.
- **Protect vulnerable groups:** Audits can reveal whether AI tools disproportionately harm or neglect specific communities (e.g., students with disabilities, ethnic minorities, non-traditional applicants) and prompt corrective action.
- **Ensure accountability:** Regular review processes clarify who is responsible for monitoring AI tools, responding to audit findings, and making necessary adjustments.

## Conducting Audits to Promote Fairness and Prevent Bias

To ensure AI systems in Higher Education Institutions (HEIs) operate fairly and equitably, audits must be conducted as structured, transparent, and actionable processes. These audits should assess both the technical performance of AI tools and their broader social impact, especially in sensitive domains such as admissions, grading, student advising, and staff recruitment.

An effective audit begins with quantitative analysis, such as examining false positive and false negative rates across different demographic groups (e.g., gender, ethnicity, disability status). These metrics help detect whether certain populations are systematically disadvantaged by the algorithm's outputs. For example, if a predictive model for academic risk flags students with disabilities more often than others without clear justification, this indicates a serious bias requiring attention.

In parallel, qualitative evaluations are essential for capturing contextual and experiential insights. Feedback from students, faculty, and staff, especially those directly affected by AI decisions can highlight unintended harms or concerns that technical metrics may overlook. Focus groups, structured interviews, and anonymous surveys can reveal whether users feel fairly treated, adequately informed, and able to contest automated outcomes.

To enhance impartiality and public trust, HEIs should involve external reviewers or independent ethics committees in the auditing process. These may include academic experts in digital ethics, civil society representatives, or data protection authorities. Their role is to offer independent scrutiny, ensure that audits meet ethical standards, and prevent internal bias or conflict of interest from skewing results.

Most importantly, audits must not end at identification. Audit results must be thoroughly documented, publicly reported when appropriate, and used to drive meaningful change. Follow-up procedures may involve retraining machine learning models with more diverse or representative data, adjusting algorithmic thresholds that unfairly exclude certain groups, or even decommissioning tools that cannot be ethically justified. Institutions should establish internal policies outlining who is responsible for these steps and how often re-auditing should occur.

Rigorous, transparent, and inclusive auditing practices are a cornerstone of ethical AI governance. By embedding them into institutional processes, HEIs affirm their commitment to fairness, equity, and academic integrity in the age of automation.

### 332 Ensuring Diverse Data to Prevent Social Bias

One of the most effective ways to prevent bias in Artificial Intelligence systems is to ensure that the data used to train and operate these systems is both diverse and representative of the entire academic community. In Higher Education Institutions, this means moving beyond datasets that reflect only dominant or historically privileged student populations and ensuring inclusion of individuals from different genders, ethnic backgrounds, socioeconomic groups, abilities, age ranges, and learning paths.

AI systems are only as fair as the data they are trained on. If the training data is skewed, such as overrepresenting students from traditional academic trajectories or underrepresenting minority groups, the AI may replicate and amplify existing inequalities. For example, an AI tool designed to predict student success might perform well for majority populations but inaccurately flag students from marginalized backgrounds as “at risk,” leading to unnecessary interventions or stigmatization.

To counteract this, HEIs must adopt data governance frameworks that mandate inclusivity and equity in data collection and usage. This includes:

Ensuring that datasets used in grading systems, admissions analytics, or learning management platforms include examples from a broad cross-section of students.

Collecting disaggregated data that captures intersectional identities (e.g., a student may be both a first-generation college attendee and have a disability).  
cross-section of students.

Implementing safeguards to avoid “overfitting” to the dominant group while still protecting the privacy of small or vulnerable groups.

Moreover, institutions should engage with affected communities when developing datasets to understand context-specific challenges.

**Example:** students from linguistic minorities might use different terminology or writing patterns that, if misunderstood by AI-powered writing tools, could result in inaccurate feedback or lower automated grades. Including their voices in system design helps ensure AI responds equitably.

Transparency in how datasets are built and maintained is also essential. Institutions should publish information about data sources, demographic coverage, and limitations to enable critical scrutiny and trust. In addition, regular bias testing and dataset updates must be built into the system lifecycle to ensure that tools remain fair as demographics and contexts evolve.

Ultimately, by using diverse and representative data, HEIs signal their commitment to a more just and inclusive academic environment—one where AI technologies do not perpetuate discrimination but actively support equity, access, and equal opportunity.

### 333. Appealing Mechanisms for AI Decisions

To uphold fairness and prevent harm caused by unjust or inaccurate algorithmic decisions, Higher Education Institutions must implement clear, accessible, and trustworthy appeal mechanisms. These mechanisms ensure that individuals, whether students, faculty, or staff retain their right to challenge and seek redress when impacted by AI-generated outcomes.

As AI tools are increasingly used in areas such as admissions, grading, performance reviews, and student support, the potential for error, bias, or unintended consequences grows. For example, an AI-powered grading system might unfairly penalize students whose writing style deviates from dominant linguistic norms, or an automated performance monitoring tool might misinterpret staff work patterns due to incomplete data.

An effective appeal mechanism includes the following elements:

- **Transparency:** Individuals must be informed when an AI system is involved in a decision that affects them and be told how to access appeal procedures.
- **Human Review:** All appeals must be reviewed by qualified human decision-makers, not by another automated system, ensuring that final decisions respect context, empathy, and academic judgment.
- **Procedural Clarity:** The process for contesting an AI-generated decision should be well-documented and easy to navigate. This includes clear timelines, responsibilities, and available forms of resolution.
- **Non-retaliation:** Individuals must feel safe to raise concerns without fear of negative consequences or institutional backlash.
- **Feedback Loops:** Institutions should treat appeals not only as individual cases but also as system-level indicators. Recurrent issues may reveal structural bias, poor training data, or misuse of AI tools, prompting re-evaluation or system redesign.

These mechanisms help protect the rights and dignity of all members of the academic community. More broadly, they strengthen trust in the institution's digital transformation by demonstrating that AI is a tool for enhancement, not a replacement for human accountability or ethical standards. In line with the principles of bias prevention and fairness, establishing appeal pathways ensures that those affected by AI remain at the center of institutional decision-making.

Promoting fairness and preventing bias in the use of AI is not a one-time technical fix, but an ongoing institutional commitment to equity, inclusion, and academic integrity. By regularly auditing AI systems, Higher Education Institutions can identify and correct hidden forms of discrimination or systemic imbalances. Ensuring the use of diverse and representative datasets helps prevent the reproduction of social inequalities within algorithmic processes. Most importantly, providing clear appeal mechanisms empowers individuals to challenge unfair outcomes and reinforces the principle that AI must remain accountable to the people it serves. Together, these measures create a culture where AI enhances educational experiences without compromising human dignity, fairness, or trust.

## 34. Aligning AI Use with Institutional Ethics and Human Agency

Supporting human autonomy and agency means ensuring that Artificial Intelligence systems in Higher Education Institutions do not override, diminish, or replace human judgment, but rather enhance it. In educational environments, autonomy is closely linked to academic freedom, learner-centered teaching, and ethical responsibility in decision-making. AI should be a tool to empower, not control, both students and staff.

In academic settings, students must retain the freedom to think critically, make choices, and shape their learning journeys. Faculty must maintain pedagogical authority and academic integrity. When AI tools, such as adaptive learning platforms, grading algorithms, or predictive analytics start making recommendations or decisions automatically, there is a real risk of disempowering the individuals they are meant to support. For instance, a student valued as high risk by an AI system might be pushed into remedial support without consultation or personal input. A faculty member might feel forced to rely on algorithmically generated grades, even if they disagree with the outcome. Administrative staff might be subject to automated performance assessments with no opportunity for contextual explanation.

Without safeguards, these systems can unintentionally reduce people to data points, strip away individual judgment, and erode trust in educational processes. Therefore, promoting autonomy and agency requires deliberate design choices, transparent practices, and institutional policies that reinforce the role of human decision-making and uphold the dignity and voice of every member of the academic community.

**Linked principle:** Human autonomy and agency

### 34.1 Ensuring Human Oversight of all High-stakes AI Applications

In Higher Education Institutions, high-stakes decisions, such as those related to academic grading, staff evaluations, admissions, or disciplinary actions, carry significant consequences for individuals' futures. When such decisions are influenced or driven by Artificial Intelligence, it is essential to maintain strong human oversight to preserve fairness, trust, and personal agency.

Human oversight means that AI does not operate autonomously in contexts where ethical judgment, contextual sensitivity, and academic values are critical. Instead, AI should function as a decision-support tool that assists, but never replaces as the final judgment of qualified educators, evaluators, or administrators.

This oversight should include:

### 1. Mandatory Human Review of AI-Generated Decisions Before Finalization

In high-stakes environments like grading, staff evaluations, or student support allocation, even small inaccuracies or biases in AI-generated decisions can have disproportionate consequences on individuals' academic progress, career trajectories, or mental well-being. Therefore, mandatory human review is not merely a procedural formality. It is an ethical safeguard that ensures AI remains subordinate to human judgment.

This human-in-the-loop approach means that no critical decision, especially one that affects someone's rights, status, or opportunities should be made solely by an algorithm. Instead, a qualified educator, administrator, or reviewer must interpret, validate, and, if necessary, adjust the AI's output in light of broader institutional values, contextual understanding, and professional discretion.

#### Example:

If automated grading software scores an essay poorly due to unconventional language use, a faculty member should review the work manually to ensure that creativity or linguistic diversity isn't unfairly penalized.

If an AI system flags a student as "at risk," the advisor should review the student's academic and personal situation holistically before any intervention is taken.

If an AI system flags a student as "at risk," the advisor should review the student's academic and personal situation holistically before any intervention is taken.

#### Required Institutional Commitment:

- ✓ Train staff to critically evaluate AI outputs rather than accepting them at face value
- ✓ Define clear roles and responsibilities for decision-makers who act as reviewers or moderators.
- ✓ Ensure transparency about when and how AI is used, so affected individuals know when human judgment will be involved.

Keep documentation that shows both the AI's recommendation and the rationale behind the human decision, creating accountability and traceability.

### 2. Clear documentation and audit trails that show how decisions were made, what role AI played, and who held final responsibility

Clear documentation and audit trails are essential components of ethical and accountable AI implementation in Higher Education Institutions. They serve as both preventive and corrective tools by making visible how decisions involving AI are made, who was involved, and how responsibility is distributed between automated systems and human actors. In practice, this means that every high-stakes decision, whether related to grading, admissions, student support, or staff evaluation that involves AI must be traceable.

The documentation should clearly indicate:

- ✓ What the AI system recommended or identified (e.g., a risk prediction score, a suggested grade, or a resource allocation).
- ✓ Who reviewed the AI output, what criteria they used, and whether the recommendation was accepted, modified, or rejected.
- ✓ Why the final decision was made, including any human judgment applied, contextual factors considered, or institutional policies referenced.

This process not only upholds transparency, but also strengthens accountability by enabling institutions to thoroughly investigate errors or appeals, demonstrate compliance with data protection laws and ethical guidelines, monitor decision-making patterns for signs of systemic bias or misuse of AI tools, and foster trust among students and staff. When individuals know that automated systems are not operating unchecked, and that human oversight remains a central part of high-stakes decisions, they are more likely to feel respected, protected, and engaged within the academic community.

Moreover, consistent documentation and auditability make it possible to learn from experience. Over time, institutions can assess whether certain tools consistently require manual correction, whether staff are properly trained to interpret AI outputs, and how these tools align with institutional values such as fairness, inclusion, and academic integrity.

### 3. Defined escalation procedures that allow questionable AI outcomes to be flagged and reviewed by humans with authority to override or amend decisions

Defined escalation procedures are a critical safeguard within any ethical AI governance framework in higher education institutions (HEIs). These procedures ensure that when an AI system produces a decision or recommendation that appears questionable, unfair, or potentially harmful, there is a structured, transparent, and responsive mechanism for human intervention.

Such procedures clarify who is responsible for reviewing flagged AI outputs, under what circumstances escalation should occur (e.g., when students contest a grade, staff dispute a performance rating, or data appears misinterpreted), and how the review process will be conducted. The escalation pathway must include individuals with sufficient institutional authority and contextual understanding to override or amend the AI's output where appropriate. By embedding these procedures into institutional workflows, HEIs reinforce the principle that AI is assistive, not authoritative. This prevents blind reliance on automation and ensures that human judgment remains central in decisions that impact individual rights, academic progression, or professional development. It also enables institutions to continuously learn from flagged cases using them to refine AI tools, improve fairness, and uphold ethical standards over time.

### **342 Preserving Academic Freedom in the Age of AI**

Avoiding over-reliance on AI systems in Higher Education Institutions is essential to preserving academic freedom, professional autonomy, and the integrity of institutional roles. While AI tools can support efficiency, personalization, and data-informed decisions, their unchecked use can gradually erode the unique contributions of educators, researchers, and staff.

In teaching, for example, excessive dependence on AI-generated lesson plans, grading algorithms, or content recommendations may reduce the creative and pedagogical agency of faculty. Educators risk becoming facilitators of machine-generated instruction rather than active shapers of learning environments. Similarly, in research and administration, staff may feel sidelined when AI systems drive decisions about funding, performance, or student support without sufficient human input.

Over-reliance can also undermine academic freedom by narrowing intellectual diversity or promoting standardized thinking. For instance, if predictive tools guide students toward certain courses or careers based on past data, this could restrict exploration and reinforce status quo outcomes. Academic institutions thrive on dissent, experimentation, and dialogue, qualities that cannot be outsourced to algorithms.

To support human autonomy, HEIs must establish clear boundaries for AI use, promote human-AI collaboration rather than substitution, and regularly assess the impact of AI on core academic values. This means involving staff and students in AI-related decisions, encouraging critical engagement with automated outputs, and protecting space for independent judgment, creativity, and ethical responsibility.

### **343 Empowering Students and Staff through AI Literacy and Participation**

Empowering students and staff to meaningfully engage with Artificial Intelligence begins with equipping them to understand it. In Higher Education Institutions, AI literacy must be treated as a core competency, not just for data scientists or IT staff, but for the entire academic community. Without it, users are left to passively accept AI-generated outputs, unable to question or contribute to how these systems shape learning, research, or administrative life.

AI literacy programs should go beyond technical explanations to include critical thinking about ethical implications, data rights, bias, and the limits of machine decision-making. When students understand how a predictive model flags them as “at risk,” or when staff can trace how an algorithm determines workload performance, they are better positioned to respond, challenge, and participate in shaping those systems.

AI literacy refers to the ability to understand how AI systems work, what they are capable of, where their limitations lie, and how their use impacts rights, ethics, and outcomes. For HEIs, promoting AI literacy is not only a technical skill, but a form of civic and academic empowerment.

By enhancing AI literacy:

- ➔ **Students** can better navigate learning analytics platforms, automated feedback tools, or AI-assisted assessments. They can identify errors, understand how decisions are made, and advocate for their rights when necessary.
- ➔ **Faculty** can critically assess AI recommendations in grading, plagiarism detection, or curriculum personalization, making informed pedagogical decisions that retain their professional judgment.
- ➔ **Staff** can engage with HR or administrative automation tools in a way that preserves human oversight and recognizes when outputs should be challenged.

Well-designed AI literacy programs should provide a holistic understanding of the ethical frameworks that govern AI use, such as fairness, accountability, and transparency. They should also cover practical aspects of data rights, including how GDPR principles apply in educational settings and what meaningful consent looks like. Participants should engage with real-world examples of algorithmic bias and misuse to better understand the social and institutional risks of poorly governed AI. Finally, programs should include hands-on workshops that familiarize students and staff with the actual AI tools used within the institution, empowering them to use these systems critically and effectively.

### 344. **Monitoring the impact of AI on labor and employment conditions**

Monitoring the impact of AI on labor and employment conditions within Higher Education Institutions is essential to ensure that technological innovation does not erode workers' rights, degrade working conditions, or increase precarity. As AI tools are introduced into administrative workflows, academic management, and human resource practices, their influence on job roles, expectations, and performance metrics must be carefully assessed.

This includes tracking whether automation is leading to increased workload for remaining staff, creating unrealistic expectations based on machine efficiency, or subtly shifting power dynamics in ways that disempower educators and administrative professionals. For example, AI systems used to monitor productivity or allocate tasks may inadvertently foster surveillance cultures or reduce trust, especially if decisions are opaque or one-sided.

HEIs must regularly consult with staff representatives, unions, and HR departments to assess the effects of AI integration on job satisfaction, autonomy, and role clarity. They should also establish feedback channels that allow staff to report negative impacts or ethical concerns linked to AI use in employment-related contexts. Where necessary, institutions should revise job descriptions, update workplace policies, or place limits on AI use to preserve fair labor practices and ensure that technology augments rather than replaces human expertise and dignity.

Monitoring AI's impact on labor helps institutions uphold academic values, promote ethical employment, and foster a workplace where both human and digital capabilities are respected and aligned.

To ensure ethical AI use, Higher Education Institutions must keep human judgment at the center of all high-stakes decisions. Human oversight, academic freedom, and staff roles must not be diminished by automation. Empowering students and staff through AI literacy and participatory design builds trust and engagement. Monitoring AI's impact on employment helps protect fair working conditions. Together, these measures uphold a human-centered approach where technology enhances, rather than replaces, human agency.

## **35. Ensuring Inclusive and Accessible AI Use**

Ethically implementing Artificial Intelligence in Higher Education Institutions demands a firm commitment to inclusivity and accessibility at every stage of AI system design, deployment, and use. AI should not reinforce structural disadvantages or create new forms of exclusion. Instead, it must actively contribute to fair, equitable, and inclusive educational experiences for all, particularly for historically marginalized or underserved groups.

In practice, advancing inclusivity and accessibility through AI means developing tools that follow universal design principles, ensuring compatibility with assistive technologies such as screen readers or voice commands, while also accommodating neurodiversity learners and a variety of learning styles. It involves promoting linguistic and cultural inclusivity by creating AI systems that support multiple languages, dialects, and regional expressions and are especially important in multilingual and international educational environments. Equally crucial is the prevention of algorithmic discrimination, which requires regular audits of AI systems to identify and mitigate biased training data or unequal performance across demographic groups, particularly in sensitive areas like grading, admissions, or student support. Inclusivity also depends on the active involvement of underrepresented groups in the design and testing phases of AI development, ensuring that their unique needs and challenges are addressed from the start rather than as an afterthought. Finally, institutions must commit to monitoring digital equity over time by analyzing who benefits from AI tools and who might be excluded, and then using these insights to revise institutional policies and promote fairer access to educational opportunities.

**Linked principle:** Inclusivity

### **35.1 Ensuring AI systems and platforms comply with accessibility standards**

Ensuring that AI systems and platforms comply with accessibility standards, such as the Web Content Accessibility Guidelines (WCAG), is a foundational step in advancing inclusivity and accessibility within Higher Education Institutions (HEIs). These standards provide a globally recognized framework for designing digital content and technologies that can be used by everyone, including individuals with visual, auditory, cognitive, or motor impairments.

In the context of AI, accessibility compliance means more than just checking technical boxes. It requires a proactive approach to ensure that students and staff with disabilities can fully engage with AI-powered tools whether they're using learning management systems with integrated chatbots, AI-driven assessment platforms, or predictive analytics dashboards for student support.

An AI tool that provides personalized learning pathways must be operable with screen readers, support keyboard-only navigation, and offer captions or transcripts for multimedia content. AI-based writing assistants should be usable by students with dyslexia or ADHD, offering custom settings for pacing, visual contrast, and feedback delivery. Speech recognition tools should be adaptable for users with different speech patterns or accents, including those with physical or linguistic challenges.

To implement this, HEIs should integrate accessibility audits into their AI procurement and development processes, include users with disabilities in testing phases, and maintain documentation that demonstrates compliance with accessibility guidelines. Compliance must also be treated as an evolving target: AI systems should be continuously updated to meet newer versions of WCAG and other relevant standards.

Designing AI with accessibility in mind affirms the institution's ethical and legal obligations while sending a clear message that diversity of ability is a valued part of academic life. As UNESCO (2023) emphasizes, inclusive digital transformation in education must ensure that no learner is left behind, especially as reliance on AI technologies increases.

### 3.5.2 Using AI to Support Inclusive Education

Leveraging Artificial Intelligence (AI) to support students with disabilities, linguistic minorities, and underrepresented groups means using technological innovation to close existing equity gaps, not widen them. In Higher Education Institutions, AI tools should be designed and implemented to identify, understand, and accommodate the diverse needs of students who have historically faced barriers to full participation in academic life.

#### **For students with disabilities**

For students with disabilities, AI-powered tools offer transformative potential by reducing access barriers and enabling more independent and personalized learning. Real-time captioning can support students who are deaf or hard of hearing, while speech-to-text software assists those with mobility impairments or dyslexia. Predictive tools can flag when a student might need academic or emotional support before challenges escalate. Personalized learning interfaces designed in line with accessibility standards like WCAG can adapt content adjusting font size, contrast, navigation, or enabling voice commands - to meet various needs and respect different learning preferences. Embedding accessibility into AI systems from the start shows a proactive institutional commitment to inclusion and equal opportunity, helping make education more equitable for all.

#### **For linguistic minorities**

Expanding multilingual inclusion through AI is essential in today's diverse academic environments. Students from different language backgrounds often face challenges not due to ability, but because of linguistic barriers in instruction and assessment. Natural Language Processing (NLP) tools can help overcome these challenges by providing real-time translations, simplifying complex academic texts, and enabling students to interact in their native languages. These technologies support participation, collaboration, and comprehension in multilingual classrooms.

To ensure effectiveness and equity, HEIs should implement these tools with attention

to accuracy, cultural sensitivity, and opportunities for human clarification. In doing so, they foster truly inclusive learning environments where language differences are a strength, not a barrier.

### **For underrepresented and marginalized groups**

Artificial Intelligence can support equity by helping institutions identify students who may benefit from additional academic, financial, or social support. AI-powered early warning systems can flag at-risk students based on engagement or performance trends, enabling timely, personalized interventions like mentoring, tutoring, or counseling. However, such systems must be governed by strong ethical safeguards. Interventions should empower, not stigmatize by avoiding algorithmic profiling that reinforces stereotypes. Models must be explainable, fair across diverse populations, and developed with input from the communities they aim to serve. Students should be informed about how AI is used in these processes and retain the right to question or appeal AI-driven classifications. By combining responsible AI use with human judgment and inclusivity principles, HEIs can reduce disparities and foster more equitable learning environments.

Effectively leveraging AI to support students with disabilities, linguistic minorities, and underrepresented groups requires more than deploying advanced technologies. It demands a deliberate and ethical commitment to equity. By designing AI systems that are accessible, multilingual, and sensitive to the needs of marginalized learners, Higher Education Institutions can remove long-standing barriers and promote meaningful inclusion. When implemented transparently and in consultation with the communities they serve, these tools can shift institutional practices from reactive accommodation to proactive empowerment ensuring that all students have the opportunity to thrive.

### **3.5.3. Reducing AI Inequalities Through Digital Literacy**

As Artificial Intelligence becomes more embedded in Higher Education Institutions, ensuring equitable access to its benefits requires intentional strategies to address digital divides. Not all students and staff begin their academic journey with the same level of familiarity, access, or confidence in using digital and AI-powered tools. Without targeted support, existing inequalities, whether related to socioeconomic status, disability, linguistic background, or geographical location can be reinforced or even worsened.

Targeted digital literacy training means designing and delivering educational programs that are customized to meet the needs of diverse learners, particularly those who may be disproportionately excluded from digital advancements.

This includes:



### Basic digital skills training

For individuals with limited experience in navigating online learning environments or digital communication platforms.

### AI-specific literacy

To help students and staff understand how AI tools work, what their limitations are, how data is used, and how to critically interpret algorithmic outputs.



### Confidence-building and empowerment

Through peer-led learning, mentoring programs, and inclusive design approaches that actively involve marginalized groups in shaping the training content and tools.

### Ethical and rights-based education

To inform participants about their data rights, privacy protections (e.g., under the GDPR), and mechanisms for appeal or feedback when interacting with AI systems.



### Context-sensitive delivery

Recognizes the diverse realities of learners, for instance, offering training in multiple languages, through accessible formats and via flexible modalities.

UNESCO's 2023 guidance on AI in education highlights the urgent need for AI to support, not replace human capabilities, and stresses that inclusive digital transformation depends on equal opportunity for participation. Aligned with this vision, targeted digital literacy training plays a preventive and empowering role: it reduces the risk of algorithmic exclusion, fosters more inclusive participation in digital academic life, and strengthens the capacity of all institutional members to engage critically and responsibly with AI systems.

By investing in tailored digital literacy strategies, HEIs signal a commitment to inclusion not only in access to knowledge but also in the tools that increasingly shape educational experiences, decision-making, and opportunities for success.

To ethically implement AI in Higher Education, institutions must prioritize inclusivity and accessibility from the start. This means designing AI tools that meet accessibility standards, support diverse languages and learners, and empower underrepresented groups without reinforcing bias. Targeted digital literacy training further ensures that all students and staff can engage meaningfully with AI. As highlighted by UNESCO (2023), inclusive digital transformation requires proactive, human-centered strategies that leave no one behind.

## **3.6. Embedding AI Ethics into the Curriculum**

Embedding AI ethics into the curriculum means equipping students with the ability to critically engage with the ethical, societal, and human implications of Artificial Intelligence. As AI continues to shape decisions in education, business, healthcare, governance, and everyday life, Higher Education Institutions have a responsibility to prepare learners to navigate these complexities thoughtfully and responsibly. This involves integrating ethical reflection into existing programs as well as developing dedicated courses that explore topics such as algorithmic bias, privacy, surveillance, transparency, accountability, and human rights in the age of AI. Interdisciplinary approaches are especially valuable, allowing students to examine how AI affects different fields, from law and journalism to design and public health.

**Linked principle:** Specific domains

### **3.6.1. Integrating AI Ethics and Social Responsibility Across Disciplines**

To foster a human based and ethically grounded approach to AI, Higher Education Institutions must embed AI ethics, digital responsibility, and the social implications of AI across all academic disciplines, not just within computer science or engineering faculties. This cross-disciplinary integration ensures that every student, regardless of their field of study, is equipped to understand, evaluate, and shape the impact of AI on society, democracy, and individual rights.

AI systems now influence decision-making in nearly every professional domain from algorithmic trading in finance to diagnosis support in healthcare, from automated grading in education to AI-generated content in journalism and the arts. As these technologies become more embedded in daily life, the risks of bias, inequality, misinformation, surveillance, and disempowerment grow, especially when users lack the ethical tools and knowledge to critically engage with AI systems. Therefore, ethical AI literacy must be democratized across the curriculum, not siloed in technical courses. Accordingly:

- ➔ **Humanities and social sciences students** should explore how AI affects societal structures, justice systems, privacy rights, and cultural narratives.
- ➔ **Business and economics students** should engage with algorithmic decision-making in markets, ethical dilemmas in automation, and the impact of AI on labor and corporate responsibility.
- ➔ **Health sciences students** should critically examine AI in diagnosis, patient monitoring, and data privacy.
- ➔ **Art and media students** should reflect on authorship, originality, and the ethics of AI-generated creativity.

Courses and modules across disciplines should incorporate a diverse range of pedagogical strategies to foster critical understanding of the ethical, social, and cultural dimensions of Artificial Intelligence. This integrative approach not only strengthens students' digital literacy, but also cultivates ethical sensitivity, civic responsibility, and interdisciplinary thinking.

## CASE STUDIES

Exploring real-world instances of algorithmic harm, such as bias in facial recognition technologies, discriminatory outcomes in algorithmic lending or recruitment systems, or surveillance-based educational analytics can help students connect abstract ethical principles to tangible societal consequences. These scenarios invite critical reflection on how AI systems can unintentionally perpetuate inequalities or infringe on human rights when left unchecked.

## CRITICAL ANALYSIS OF EXISTING AI ETHICS FRAMEWORKS,

including the UNESCO Recommendation on the Ethics of Artificial Intelligence (2021), the EU's Ethics Guidelines for Trustworthy AI, or national policies on digital responsibility, encourages students to examine the philosophical underpinnings, practical implications, and limitations of different ethical governance models.

## INTERACTIVE LEARNING METHODS,

such as role-playing, structured debates, or participatory design exercises, simulate real-life ethical dilemmas involving competing interests. These activities enhance empathy, collaborative decision-making, and a nuanced understanding of stakeholder perspectives.

## DISCUSSIONS OF THE HISTORICAL, CULTURAL, AND POLITICAL CONTEXTS

in which AI technologies are developed and deployed reveal how technological systems are not neutral, but shaped by and shaping societal power structures, norms, and values. Embedding these perspectives supports critical technology studies and equips students to question dominant narratives about technological progress.

Together, these strategies support a curriculum that goes beyond technical proficiency to foster ethically grounded, socially aware graduates, capable of contributing responsibly to the development and application of AI in any field.

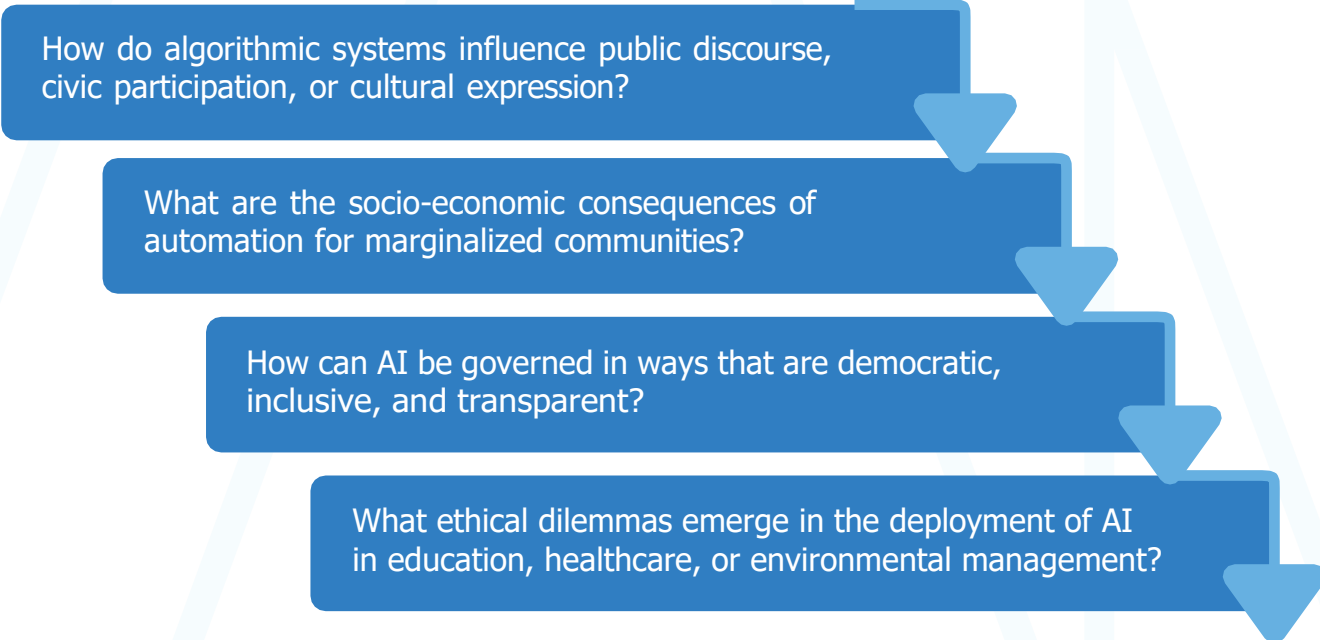
Example:

### 3.62 Promoting Interdisciplinary Approaches to AI in Society

Embedding AI ethics into the curriculum requires more than standalone modules in computer science. It calls for a broad, interdisciplinary educational approach that recognizes the pervasive impact of Artificial Intelligence across all areas of academic inquiry and professional practice. Higher Education Institutions should actively foster the development of interdisciplinary courses and research initiatives that explore the ethical, social, legal, economic, and cultural dimensions of AI.

Such initiatives bring together diverse academic disciplines ranging from philosophy, sociology, law, political science, and education to health sciences, business, media, and the arts - each offering unique perspectives on how AI technologies are designed, implemented, and experienced in society. Interdisciplinary courses encourage students to critically examine how AI reshapes human agency, labor, privacy, equity, governance, creativity, and public trust.

Research collaborations across faculties can explore questions such as:



How do algorithmic systems influence public discourse, civic participation, or cultural expression?

What are the socio-economic consequences of automation for marginalized communities?

How can AI be governed in ways that are democratic, inclusive, and transparent?

What ethical dilemmas emerge in the deployment of AI in education, healthcare, or environmental management?

By promoting these cross-disciplinary learning and research opportunities, institutions empower students and academics to address complex, real-world challenges related to AI from multiple vantage points. This fosters deeper ethical engagement, critical thinking, and innovation grounded in human-centered values.

### 3.63 Fostering Ethical and Socially Responsible Student Innovation

To prepare future leaders in the age of AI, it is essential not only to teach ethical theories and frameworks but to actively create environments where students are empowered to lead innovation grounded in ethical reasoning and social impact. Ethics should not be treated as a separate or abstract discipline, but it must become an integral part of the innovation mindset, influencing how students define problems, design solutions, and evaluate outcomes.

Student-led innovation involves giving students agency: the freedom and support to identify relevant challenges, propose novel ideas, and take ownership of real-world projects. But this freedom must come with a deep sense of responsibility. Students need to be equipped to consider not only what technology can do, but what it should do. This means embedding ethical reflection throughout the process, from the selection of a problem space to the implementation and deployment of a solution. Cultivating this mindset involves training students to anticipate the broader societal consequences of their innovations.

**Example:**

An AI-based hiring tool may improve efficiency, but if it relies on biased data or lacks transparency, it can reinforce systemic inequalities. A student-led chatbot may improve user engagement, but it must also respect user privacy and avoid spreading misinformation. By engaging students in such ethical reflection, it is helped them see innovation not just as a technical achievement, but as a social act with real human consequences.

Moreover, leadership in AI demands the ability to navigate complexity by balancing innovation with justice, performance with inclusivity, and profit with public interest. Students should be encouraged to lead projects that embrace these tensions rather than avoid them, and to work collaboratively across disciplines and perspectives. Educational institutions must therefore support this development through project-based learning, mentorship from diverse experts, and opportunities to engage with real stakeholders.

**What this means in practice:**

1. Empowering students as ethical creators

Students should be supported in initiating their own projects, startups, or research around AI and digital technologies. These initiatives should be evaluated not just by technical functionality, but by how well they address ethical dimensions such as data privacy, algorithmic fairness, transparency, environmental impact, and social equity.

2. Embedding real-world social challenges

Encourage students to identify and solve real-world problems through their innovations, especially those affecting marginalized communities or underrepresented groups.

3. Ethics as a design principle, not an afterthought

Guide students to think critically during the innovation process by encouraging them to identify all stakeholders, anticipate unintended consequences, ensure the solution is inclusive, and verify that data is ethically sourced and representative. This approach helps students create technologies that are not only effective but also fair and socially responsible.

4. Mentorship and multidisciplinary collaboration

Institutions should create spaces - innovation hubs, ethical AI labs, design studios where students can receive mentoring from experts in AI, philosophy, law, and social sciences. Interdisciplinary collaboration enriches the ethical depth of student projects and mirrors real-world decision-making environments.

### 5. Ethical evaluation as part of project assessment

Project evaluation rubrics should include ethical reasoning, stakeholder analysis, and sustainability considerations alongside technical performance and creativity.

### 6. Recognition and visibility of ethical innovation

Celebrate and showcase student projects that exemplify responsible innovation through awards, showcases, publications, and funding opportunities. This signals that ethical leadership and social impact are valued and rewarded.

In essence, student-led innovation that reflects ethical and social considerations prepares learners not just to build technology, but to shape the future with intention, responsibility, and integrity. It nurtures leaders who are not only capable of solving problems, but also of defining what problems are worth solving and doing so in ways that uplift, rather than undermine, societal well-being.

Embedding AI ethics into the curriculum is essential for preparing students to navigate and shape the impact of AI in society. This requires integrating ethical thinking across all disciplines, fostering interdisciplinary learning, and supporting student-led, socially responsible innovation. By equipping learners with the tools to critically engage with AI technologies, Higher Education Institutions help develop future leaders who are not only technically skilled but also ethically aware and committed to the common good.

## 3.7. Updating Institutional Policies to Reflect AI Ethics

As AI systems increasingly influence academic decisions, research processes, and administrative operations, Higher Education Institutions must update their internal policies to ensure responsible and ethical use of these technologies. A critical ethical priority is transparency and explainability ensuring that AI tools used within the institution operate in ways that are clear, understandable, and open to scrutiny. Institutional policies should require that all AI - based systems, whether used in student assessment, admission processes, plagiarism detection, or academic advising are implemented with full transparency.

Accordingly, all members of the academic community, including students, faculty, and administrative staff should be adequately informed about the use of AI systems, understand their underlying logic and intended function, and be able to access clear explanations of how automated decisions are made, particularly in cases where such decisions have a direct impact on academic or personal outcomes.

**Linked principle:** Transparency and explainability

### 3.7.1. Updating Policies for AI-Specific Contexts

As artificial intelligence becomes increasingly integrated into teaching, learning, research, and institutional management, Higher Education Institutions must ensure that their governance frameworks evolve accordingly. This includes the systematic review and revision of existing institutional policies, particularly those related to academic integrity, assessment, research ethics, data protection, intellectual property, and student rights to reflect new ethical and operational realities introduced by AI.

Many of the policies currently in place were designed before the proliferation of generative AI tools, algorithmic grading systems, and AI-assisted research methods. As a result, they may lack the specificity or scope needed to provide clear guidance in situations where AI systems affect decision-making, automate tasks traditionally performed by humans, or raise concerns about transparency, fairness, or accountability.

Incorporating the principles of transparency and explainability into policy reform ensures that AI systems are not used as opaque or unquestioned mechanisms. Instead, policies should:

- ✓ Require that AI tools used in academic and administrative settings are clearly documented, including information about their purpose, design logic, data sources, and limitations.
- ✓ Ensure that users - students, faculty, researchers, and administrators are informed about when and how AI is used, particularly in decisions that affect their academic trajectory, grading, research evaluations, or disciplinary outcomes.
- ✓ Mandate explainability in automated decision-making, providing mechanisms for users to access understandable explanations, contest decisions, and receive human support when necessary.

#### Example:

Research ethics policies must address the ethical implications of AI-assisted data collection, analysis, and generation. This includes requiring disclosure of AI use in methodology sections, ensuring that data bias and system limitations are acknowledged, and that the research remains interpretable and replicable by human scholars. Institutional review boards (IRBs) and ethics committees should be equipped with the knowledge and procedures to evaluate projects involving AI components.

#### Similarly:

Revised academic integrity policies should not only define how AI-generated content is treated but also clarify how AI-detection systems work and how students are informed about their use. Assessment policies should require that when automated grading or feedback tools are deployed, students are given a meaningful explanation of how their work was evaluated and a pathway to request human review.

### Operationalizing AI ethics in institutional governance

Translating ethical principles into practice requires more than abstract commitments demands the creation of concrete institutional structures, processes, and accountability mechanisms that ensure AI is implemented in a transparent, fair, and responsible manner. For Higher Education Institutions, this means embedding AI ethics into governance by clearly defining roles, responsibilities, and workflows for the oversight of AI technologies used in academic and administrative settings.

## Clarifying institutional responsibilities for ethical oversight

HEIs must establish formal procedures and designated bodies responsible for overseeing the development, acquisition, deployment, and evaluation of AI systems. These responsibilities should be distributed across relevant institutional units, including:

- IT departments, responsible for technical evaluations, system integration, and ensuring compliance with data protection standards.
- Teaching and learning centers, which assess the pedagogical implications of AI tools used in classrooms and learning management systems.
- Ethics committees or AI governance boards, which provide cross-disciplinary ethical review of AI applications, particularly those that may affect student rights, academic outcomes, or staff performance.

The goal is to create a coordinated and accountable framework where ethical oversight is embedded at every stage of the AI system lifecycle.

### Establishing clear documentation and impact assessment requirements

HEIs should require that all AI tools and systems undergo formal ethical documentation and impact assessments prior to implementation. These may include:

#### **Fairness Audits**

Fairness audits involve a structured examination of whether an AI system produces outcomes that are biased or discriminatory toward particular individuals or groups. These audits are especially important in academic settings, where AI may be applied in areas such as admissions algorithms, automated grading systems, academic risk prediction tools, and plagiarism detection systems. A comprehensive fairness audit assesses whether the training data is representative of the student population and whether it reflects or perpetuates historical inequalities. It also examines whether the system's outputs lead to systematic disadvantages for certain groups based on characteristics such as gender, race, ethnicity, disability, language proficiency, or socioeconomic background. Additionally, it analyzes the decision-making logic of the system to determine whether specific parameters or structures contribute to biased outcomes. The audit process should be conducted by cross-functional teams that include technical experts, ethicists, and relevant stakeholders, and should result in concrete recommendations ranging from system redesign and the implementation of bias mitigation strategies to the possible rejection of tools that cannot meet fairness standards.

#### **Explainability reports**

As AI systems increasingly inform and influence decisions in academic contexts ranging from student assessment and admissions to research analytics and administrative processes, it becomes essential to ensure that these systems are not only technically robust but also intelligible, reviewable, and open to scrutiny. This is where explainability reports play a vital role. Designed to make AI systems comprehensible to a broad audience, including non-technical users, explainability reports are a foundational tool for supporting informed participation, institutional accountability, and ethical governance.

## Functions of explainability reports:

- Translate the technical complexity of AI systems into accessible, structured narratives that can be understood by students, faculty, administrators, policymakers, and other stakeholders.
- Demystify the internal logic of AI-driven decisions, helping users understand why a system arrived at a particular outcome.
- Provide a foundation for institutional oversight, enabling ethics boards, governance bodies, or independent auditors to evaluate whether the system aligns with institutional values, legal obligations, and fairness criteria.
- Uphold the principle that those affected by automated decisions have a right to understand and question how those decisions are made.

Explainability reports are a critical mechanism for achieving transparency. They counter the tendency for AI systems, especially those using complex or opaque algorithms to function as “black boxes” whose internal processes are inaccessible or unintelligible to end-users. Without such clarity, users cannot meaningfully engage with or challenge automated decisions, which undermines trust, fairness, and autonomy.

Equally, these reports are essential for accountability. They document how and why systems operate in a particular way, creating a trail of justification that can be reviewed by internal and external actors. This ensures that the deployment of AI in education is not left solely to vendors or technical teams, but is subject to institutional responsibility and ethical oversight.

When shared proactively, whether through internal portals, policy documents, or onboarding materials, explainability reports empower students, faculty, and staff. They promote transparency in decision-making, reduce confusion or mistrust, and foster a shared understanding of how technology intersects with educational values and processes.

To fully leverage the value of explainability reports, HEIs should:

- ➔ Integrate them into AI procurement and deployment protocols, making them a standard requirement for any third-party or internally developed AI tool.
- ➔ Ensure accessibility, not just in terms of availability but also in terms of language clarity and format.
- ➔ Use them as a foundation for dialogue, encouraging stakeholders to ask questions, provide feedback, and participate in continuous ethical review.
- ➔ Link them to policy frameworks, for example, requiring them as part of ethical approval processes, system audits, or student rights documentation.

## **Risk and Benefit Assessments**

Ethical Risk and Benefit Assessments are a crucial element of responsible AI governance in higher education. They serve to evaluate the broader consequences of deploying AI systems, moving beyond technical or operational concerns to consider the ethical, social, psychological, and institutional dimensions of AI use. These assessments ensure that decisions to implement AI technologies are not solely based on efficiency or innovation potential, but are grounded in ethical reflection, transparency, and accountability.

Rather than being reactive, these assessments are proactive tools designed to identify potential harms and benefits before, during, and after AI deployment, and to provide institutions with a framework for making ethically sound decisions about technology use in academic environments.

Crucially, risk and benefit assessments support accountability by documenting anticipated outcomes, assigning responsibility for managing risk, and identifying thresholds at which a system should be revised, delayed, or discontinued. They also foster transparency, by encouraging open communication about trade-offs and uncertainties associated with AI adoption.

Such documentation ensures that decisions about AI deployment are evidence-based, transparent, and ethically justified, and that relevant stakeholders, such as faculty, students, and administrators can review and understand them.

### **3.7.2 Ensuring Transparency in AI Use**

As AI tools become more deeply embedded in higher education, supporting everything from academic content creation to administrative decision-making, Higher Education Institutions must adopt clear policies that mandate transparency in all uses of AI within their academic ecosystem. This includes both the use of AI by the institution (e.g., in automated grading, plagiarism detection, admissions filtering) and the use of AI by students and staff (e.g., generative AI tools in writing, research, or assignments).

Transparency in this context means that users and stakeholders are fully informed about when and how AI is being used, what decisions it affects, and what rights individuals have in relation to those systems. It also means that AI systems must be accompanied by clear and understandable explanations of their logic, data sources, and decision-making processes.

#### **Academic Work and Student Submissions**

As AI tools become more accessible and widely used by students in their academic work, Higher Education Institutions face the growing need to ensure that their learning environments remain transparent, ethical, and aligned with core educational values. This requires the explicit disclosure of AI use in all forms of student submissions, supported by clear institutional guidelines and a shared understanding of academic integrity in the digital age.

Requiring students to declare their use of AI tools, such as ChatGPT, Grammarly, GitHub Copilot, or AI-powered data analysis platforms is not intended to discourage innovation or technological fluency. Rather, it serves several critical ethical and pedagogical purposes:

- ➔ Preserving academic integrity by making the learning process visible and accountable;
- ➔ Upholding fairness by ensuring that all students are evaluated on comparable terms;
- ➔ Empowering educators to assess not only the quality of the final product but also the student's individual contribution and understanding;
- ➔ Encouraging responsible AI use, rather than hidden or inconsistent reliance on external tools.

Transparency also creates a foundation for dialogue between students and educators about what constitutes ethical assistance versus academic dishonesty in the context of emerging technologies.

To support consistent and fair evaluation, institutions should develop policies that ask students to clearly indicate:

- ➔ What tools were used: Naming specific AI platforms or applications (e.g., ChatGPT, Grammarly, QuillBot, or specialized AI data processors).
- ➔ For what purposes: Explaining how the tool supported the work (e.g., idea brainstorming, paraphrasing, coding suggestions, citation generation, data visualization).
- ➔ To what extent AI contributed: Estimating how much of the submitted work was shaped, edited, or produced with the help of AI. For example, was it used only to check grammar, or did it generate entire paragraphs?

This level of disclosure helps contextualize the student's output and allows instructors to provide meaningful, personalized feedback while upholding academic standards.

### **Balancing innovation and responsibility**

Many AI tools offer powerful learning support, facilitating brainstorming, enhancing clarity, or reducing barriers for students with language challenges. When used transparently and appropriately, AI can be a valuable educational ally. However, undisclosed or excessive reliance on AI may undermine the learning process, compromise academic assessment, and raise concerns about authorship and cognitive engagement.

By requiring disclosure, institutions reinforce the idea that AI should support, not replace, human learning. This distinction enables educators to guide students on the appropriate use of AI within disciplinary expectations, address over-reliance or misuse through educational interventions rather than punitive measures, and foster digital literacy by helping students reflect on how AI shapes their thinking and writing.

To implement this approach effectively, institutions must:

- ➔ Clearly define what constitutes acceptable and unacceptable AI use in different contexts (e.g., essays, exams, group work, lab reports);
- ➔ Provide templates or standardized statements for students to declare AI use in submissions;
- ➔ Educate students and faculty about the ethical, legal, and intellectual implications of AI-generated content;
- ➔ Include AI use declarations in academic integrity policies, treating non-disclosure as a form of misrepresentation when relevant.

Institutions should also recognize that norms around AI use will vary across disciplines and tasks. For instance, AI-assisted coding may be more accepted in computer science than in philosophy essays. Policies should be adaptable and responsive to evolving pedagogical needs and disciplinary values.

Encouraging transparent AI use in academic work is a way to build a culture of ethical, reflective, and responsible innovation. By making AI involvement visible and discussable, institutions promote integrity, support fair evaluation, and prepare students to navigate the ethical complexities of AI in their future academic and professional lives.

### 3.7.3 Involving Stakeholders in the Co-Creation of AI Policies

For AI ethics to be meaningfully embedded in HEIs, policy development must not be a top-down, technocratic process. Instead, it should be inclusive, participatory, and transparent, reflecting the lived experiences, concerns, and values of those most affected by AI deployment, namely, students, faculty, and administrative staff. Involving diverse stakeholders in the co-creation of AI policies strengthens both the legitimacy and the effectiveness of institutional governance, while upholding the ethical principles of transparency and explainability. Given the high stakes involved, it is ethically and strategically essential to include affected stakeholders - students, faculty, and staff in the development of policies governing AI use.

Involving stakeholders directly in the design and review of AI policies offers numerous benefits:

**Shared ownership and collective responsibility:** When students, faculty, and staff are engaged as co-creators, they are more likely to view AI policies as legitimate, fair, and aligned with the institution's academic mission. This sense of shared responsibility fosters a proactive culture of ethical engagement rather than reactive compliance.

**More responsive and context-sensitive governance:** Participatory policy development ensures that AI regulation is informed by actual institutional practices and needs. It allows policies to be tailored to specific departments, cultures, and learning models, and ensures that disciplinary diversity is not lost in standardized, one-size-fits-all solutions.

**Richer understanding of ethical risks and trade-offs:** Different stakeholder groups bring different ethical perspectives to the table. Students may focus on privacy and autonomy, while faculty may emphasize academic freedom or grading fairness. Including these views leads to more robust, well-rounded policies that anticipate real-world challenges and social dynamics.

**Greater transparency and explainability through process:** Involving stakeholders from the outset helps ensure that not only the AI systems themselves, but also the institutional policies that govern them, are transparent, understandable, and subject to critical dialogue. This is explainability in its broader, democratic sense by not only explaining how a system works, but also why it is used, who decided to use it, and how it fits into the values of the academic community.

Explainability also means ensuring that the policies governing AI use are not only technically sound, but also clear, contextually meaningful, and grounded in the social and educational realities of the academic community. Accordingly, stakeholder involvement helps:

- ➡ Translate abstract technical and legal language into policies that make sense in real academic contexts;
- ➡ Surface discipline-specific concerns (e.g., what “originality” means in design vs. philosophy);
- ➡ Enable meaningful scrutiny of assumptions, data practices, and implementation strategies;
- ➡ Build institutional capacity for digital literacy through ongoing dialogue.

When students and staff are engaged in shaping the policies that affect them, they are more likely to understand and support AI-related decisions, and to critically reflect on how AI intersects with educational values and responsibilities.

To operationalize this recommendation, HEIs should begin by creating multi-stakeholder working groups or advisory councils dedicated to AI ethics and governance. These bodies can provide a structured space for dialogue, reflection, and policy development involving representatives from across the academic community.

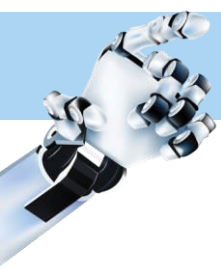
In addition, institutions should organize consultation forums, workshops, where students, educators, and support staff have the opportunity to express concerns, share expectations, and engage in shaping institutional approaches to AI. Ensuring diversity and representativeness in these processes is essential, particularly by including voices from

marginalized or underrepresented communities whose experiences with technology and education may differ significantly from dominant narratives.

To support meaningful engagement, participants should be given access to clear, accessible background information, enabling all individuals, regardless of technical expertise, to contribute thoughtfully to the discussion. Finally, HEIs should establish feedback mechanisms and revision loops that not only collect input but also communicate how that input influenced final policies, reinforcing a sense of transparency, inclusion, and ongoing stakeholder influence.

Stakeholder involvement is a core element of ethical AI governance. By engaging students, faculty, and staff in meaningful dialogue about the design, deployment, and regulation of AI systems, institutions can create policies that are more legitimate, contextually appropriate, and socially just. This inclusive approach helps ensure that AI is used to enhance, rather than disrupt, the values of higher education: trust, equity, autonomy, and shared knowledge creation.

As AI becomes more embedded in higher education, institutions must ensure their policies reflect core ethical principles—especially transparency and explainability. By updating existing regulations, requiring disclosure in academic and administrative uses of AI, and involving stakeholders in shaping these policies, HEIs can foster trust, accountability, and responsible innovation. Ethical AI governance is not only about managing technologies, but about protecting academic values and ensuring that all members of the community can engage with AI in informed and equitable ways.



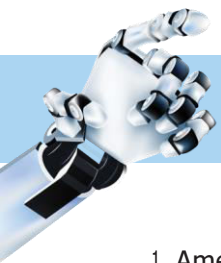
## Conclusion

Looking ahead, as AI systems become increasingly sophisticated, it is imperative to consider the long-term implications for the future of human-AI interaction in education. While the immediate focus in education is on more narrow AI applications, a forward-thinking ethical framework must consider the evolving relationship between humans and AI in the learning process and ensure that AI serves to enhance, rather than diminish, human flourishing and educational equity. When considering the long-term implications for the future of humanity in learning, it is aimed that AI systems in education become increasingly sophisticated and integrated into the fabric of learning, it is essential to define the long-term ethical implications for the future of human-AI interaction within educational contexts. While current applications often focus on narrow AI tasks, reflecting on broader discussions surrounding artificial general intelligence and superintelligence, serves as a crucial reminder to proactively consider the potential societal consequences of growing autonomous AI systems in shaping the minds of future generations.

The integration of AI into education holds immense promise, but realizing its full potential requires a strong ethical compass. Addressing the complex issues of accountability, bias, autonomy, labor, privacy, safety, and transparency is not merely a matter of compliance but a fundamental imperative for ensuring the responsible and beneficial use of AI in shaping the future of learning. By proactively engaging with these ethical norms, fostering interdisciplinary dialogue, and developing robust regulatory frameworks, it is possible can navigate the algorithmic landscape of education in a way that prioritizes the well-being, equity, and holistic development of all learners.

**Please scan QR code and provide your feedback**





## References

1. American Association of University Professors. (2025). Artificial Intelligence and the Academic Professions. <https://www.aaup.org/report/artificial-intelligence-and-academic-professions>
2. Bostrom, N. (2014). *Superintelligence: Paths, dangers, strategies*. Oxford University Press.
3. Chaudhry, M. A., Cukurova, M., & Luckin, R. (2022). A Transparency Index Framework for AI in Education. [https://arxiv.org/abs/2206.03220?utm\\_source=chatgpt.com](https://arxiv.org/abs/2206.03220?utm_source=chatgpt.com)
4. EDUCAUSE. (2025). AI Ethical Guidelines – EDUCAUSE Library. [https://library.educause.edu/resources/2025/6/ai-ethical-guidelines?utm\\_source=chatgpt.com](https://library.educause.edu/resources/2025/6/ai-ethical-guidelines?utm_source=chatgpt.com)
5. European Commission. (2019). *Ethics Guidelines for Trustworthy AI*. Independent High-Level Expert Group on Artificial Intelligence. Brussels. <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>
6. European Commission. (2020). *White Paper on Artificial Intelligence – A European Approach to Excellence and Trust*. [https://commission.europa.eu/system/files/2020-02/commission-white-paper-artificial-intelligence-feb2020\\_en.pdf](https://commission.europa.eu/system/files/2020-02/commission-white-paper-artificial-intelligence-feb2020_en.pdf)
7. European Parliament & Council of the European Union. (2016). Regulation (EU) 2016/679 – General Data Protection Regulation (GDPR). <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32016R0679>
8. GEANT. (2024, October 18). AI, data privacy, and ethics in higher education: Navigating the future. <https://connect.geant.org/2024/10/18/ai-data-privacy-and-ethics-in-higher-education-navigating-the-future>
9. GDPR Advisor. (2023). *GDPR Compliance in the Education Sector: Protecting Student Data in Learning Environments*. [https://www.gdpr-advisor.com/gdpr-compliance-in-the-education-sector-protecting-student-data-in-learning-environments/?utm\\_source=chatgpt.com](https://www.gdpr-advisor.com/gdpr-compliance-in-the-education-sector-protecting-student-data-in-learning-environments/?utm_source=chatgpt.com)
10. General Data Protection Regulation (GDPR), European Union. <https://eur-lex.europa.eu/eli/reg/2016/679/oj>
11. Georgieva, M., & Stuart, J. (2025). Ethics Is the Edge: The Future of AI in Higher Education. *EDUCAUSE Review*. <https://er.educause.edu/articles/2025/6/ethics-is-the-edge-the-future-of-ai-in-higher-educatio>
12. Justin B. Bullock (2023) *The Oxford Handbook of AI Governance, Harms of AI*. <https://doi.org/10.1093/oxfordhb/9780197579329.013.65>
13. Karran, A. J., Charland, P., Martineau, J.-T., Ortiz de Guinea López de Arana, A., Lesage, A. M., Senecal, S., & Leger, P.-M. (2024). Multi Stakeholder Perspective on Responsible Artificial Intelligence and Acceptability in Education. *NPJ Sci Learn*. [https://arxiv.org/abs/2402.15027?utm\\_source=chatgpt.com](https://arxiv.org/abs/2402.15027?utm_source=chatgpt.com)
14. Luckin, R., & Selwyn, N. (2016). Digital learning and the 'AI turn': what difference do artificial intelligence and data science make? *British Journal of Educational Technology*, 48(6), 1152-1163.
15. Luckin, R. (2017). *Towards artificial intelligence-based assessment systems*. Nature Human Behavior.
16. OECD (2021). *Framework for the Classification of AI Systems*. Organization for Economic Co-operation and Development (OECD). <https://www.oecd.org/publications/oecd-framework-for-the-classification-of-ai-systems-cb6d9eca-en.htm>
17. Porayska-Pomsta, K., Holmes, W., & Nemorin, S. (2024). The ethics of AI in education. *arXiv*. <https://arxiv.org/abs/2406.11842>
18. Selwyn, N. (2017). *Education and technology: Key issues and debates*. Bloomsbury Publishing
19. Selwyn, N. (2019). Should robots replace teachers? AI and the future of education. *Learning, Media, and Technology*.
20. Sharples, M. (2023). Towards social generative AI for education: theory, practices and ethics. *arXiv*. Retrieved from <https://arxiv.org/abs/2306.10063>
21. UNESCO. (2021). *Recommendation on the Ethics of Artificial Intelligence*. <https://iep.utm.edu/ethics-of-artificial-intelligence>
22. Zhang, X., Ahmed, T., Brookes, J., & Yamada, K. (2025). Fairness and Effectiveness in AI-Driven Educational Assessments: Challenges and Mitigation Strategies. *Journal of Innovation and Development*, 11(2), 55–84. [https://drpress.org/ojs/index.php/jid/article/download/30638/30018/45085?utm\\_source=chatgpt.com](https://drpress.org/ojs/index.php/jid/article/download/30638/30018/45085?utm_source=chatgpt.com)

# AI-ERIT: AI Integration Framework for Higher Education: Ensuring HEIs' Readiness for Inescapable Transformation

Project Code: 2024-1-LT01-KA220-HED-000251565

Discover our resources  
on the project website

[www.ai-erit.eu](http://www.ai-erit.eu)

Created by:



Fachhochschule  
des Mittelstands



ASPIRA<sup>®</sup>  
UNIVERSITY  
OF APPLIED SCIENCES

UNIVERSIDADE  
DA MAIA

