FEB 2026

COLUMBIA CLIMATE SCHOOL
CENTER FOR SUSTAINABLE DEVELOPMENT

FII INSTITUTE
Future Investment Initiative Institute
Impact on Humanity

AI AND ETHICS OF SMART EDUCATION

Prepared by

Prof. Yanis Ben Amor
Center for Sustainable Development,
Columbia University

Prof. Cristina Godoy
University of São Paulo

Executive Summary

Artificial intelligence is rapidly reshaping education systems worldwide, accelerating the development of "smart education" models that integrate digital platforms, learning analytics, and generative AI into teaching, assessment, and administration. These technologies offer significant opportunities to expand personalized learning, strengthen accessibility through translation and adaptive tools, support teachers by reducing administrative burdens, and improve system-level planning through real-time data insights. However, the same technologies introduce ethical risks that are uniquely concentrated in education, where children represent a highly vulnerable population and AI-enabled decisions can shape long-term life trajectories.

This report examines the ethics of smart education through six country case studies: the European Union (France), Qatar, Brazil, China, Japan, and the United States. The comparative analysis demonstrates that ethical outcomes depend less on technical capability than on the governance models through which AI is deployed. Three dominant pathways emerge. The European Union, exemplified by France, reflects a rights-based and precautionary approach grounded in the AI Act and GDPR, which classify many educational AI systems as high-risk and impose strict requirements for transparency, human oversight, data minimization, and protection of minors. China, Qatar, and Japan illustrate state-driven strategies that frame smart education as national infrastructure for competitiveness and societal resilience, with varying degrees of central coordination and emphasis on literacy, security, and sovereignty. The United States and Brazil demonstrate fragmented governance environments where adoption is often market-led and decentralized, resulting in uneven safeguards and greater vulnerability to surveillance drift and vendor-driven experimentation.

Across all cases, a consistent finding is that education concentrates multiple high-risk AI applications simultaneously. Longitudinal learner profiling through learning analytics, automated classification and placement systems, AI-supported grading and discipline tools, and the rapid diffusion of generative AI tutoring and content systems create pathways for structural harm if deployed without enforceable safeguards. The most prominent shared risk is the "datafication of childhood," whereby continuous monitoring generates persistent student profiles with unclear retention limits, secondary uses, and accountability structures. The report also highlights the growing significance of epistemic integrity and child safety risks introduced by generative AI, including hallucinations, misinformation, academic misuse, and deepfake-enabled bullying. These risks are particularly acute in contexts where detection and enforcement capacity is limited.

The case studies further show that equity challenges extend beyond algorithmic bias and increasingly depend on infrastructure. Qatar's Arabic-first AI strategy demonstrates how language capability can function as a fairness prerequisite rather than a localization feature. China's experience underscores that regional disparities in infrastructure and capacity can produce uneven adoption outcomes. France and Japan emphasize the importance of human-centered governance, including literacy and developmental safeguards, while Brazil and the United States illustrate the consequences of weak ex-ante oversight, where biometric surveillance tools and vendor systems may enter schools without adequate transparency, contestability, or public accountability.

Based on these findings, the report proposes a set of policy recommendations designed to support ethical smart education. These include: (1) classifying key educational AI applications as high-risk; (2) institutionalizing contestability and redress mechanisms; (3) enforcing strict data minimization and retention limits for minors; (4) regulating procurement and vendor contracts as primary governance instruments; (5) restricting biometric and emotion recognition technologies in schools; (6) embedding AI literacy and ethics training across students and educators; (7) establishing protocols to address generative AI threats to integrity and child safety; (8) protecting teacher agency and ensuring meaningful human oversight; (9) investing in trusted education data infrastructures and sovereign capacity where appropriate; and (10) integrating digital wellbeing safeguards into smart education strategies.

The report concludes that smart education is not primarily a technical challenge, but a governance challenge. If deployed without robust safeguards, AI risks entrenching inequality, weakening trust in learning institutions, and normalizing surveillance in childhood. If governed responsibly, it can strengthen human development, educational inclusion, and national resilience. The pathway forward therefore requires moving beyond voluntary principles toward enforceable standards that protect children's rights, preserve human-centered learning environments, and ensure that innovation serves the public interest.

Panel 1: Key Report Findings

KEY FINDING	EXECUTIVE TAKEAWAYS
Education is a High-Risk AI Domain	▫ Education concentrates multiple high-stakes AI uses (grading, placement, profiling, tutoring). Errors and bias can directly shape long-term life trajectories and social mobility. Because learners are minors, education requires safeguards stronger than most sectors.
Datafication of Childhood is the Central Structural Risk	▫ Smart education relies on continuous data extraction and longitudinal learner profiling. Persistent "digital dossiers" risk extending beyond school contexts and time horizons. Retention limits and deletion rights remain weakly operationalized in most systems.
Contestability Gaps Create "Black Box" Schooling	▫ Many AI-assisted educational decisions lack transparent rationale and traceability. Contestability requires enforceable procedures (appeals, documentation, timelines, accountability). Fragmented governance increases the risk of unreviewable outcomes affecting students' futures.
Generative AI Expands Ethics Into Safety and Integrity	▫ GenAI introduces misinformation and hallucination risks that undermine learning reliability. Deepfakes are emerging as a major child safety threat, including sexualized harassment. Education governance must address AI as a safety and integrity issue, not only privacy.
Teacher Agency is the Primary Safeguard—Yet Under Pressure	▫ AI can support instruction, but also risks deskilling educators into compliance moderators. "Human-in-the-loop" rhetoric often masks structural incentives toward automation. Ethical deployment requires teachers to retain authority to override AI outputs.
Governance Architecture Determines Outcomes More Than Technology	▫ Three pathways shape ethics: rights-based regulation, state-driven integration, and fragmented adoption. The same tools yield different outcomes depending on procurement rules and oversight capacity. Weak governance environments tend toward reactive enforcement after harm occurs.
Sovereignty vs. Privacy is an Emerging Strategic Dilemma	▫ Building sovereign educational AI requires representative data while protecting minors' rights. Excess dependence on foreign platforms risks structural lock-in and loss of control. Trusted public data spaces plus privacy-preserving methods offer the most viable compromise.

Introduction

The advent of Artificial Intelligence (AI) is transforming education systems worldwide. "Smart education" refers to the deployment of AI-enabled technologies—including generative AI, adaptive learning systems, predictive analytics, and automated decision-support tools—across teaching, assessment, and education administration. While these tools may strengthen personalized instruction, improve accessibility, and reduce administrative burdens, they also raise significant ethical concerns, particularly when deployed at scale in learning environments involving minors. The integration of AI in classrooms has therefore created a growing policy challenge: ensuring that education systems harness innovation while safeguarding fairness, privacy, transparency, and human-centered decision-making.

Public concern about these risks is already widespread. According to FII Institute's 2025 PRIORITY Compass, education ranks **#6 among global priorities**, and **76%** of respondents express concern that the use of AI in education may **increase the digital divide**. Education is also ranked as a top priority across multiple regions—including **MENA (#3)**, **Africa (#4)**, and **Asia (#5)**—underscoring both the urgency and the global relevance of ethical governance in this domain.

This report explores the evolving smart education landscape in Brazil, China, France (as an illustrative European Union case), Japan, Qatar, and the United States. While the boundaries of AI-enabled learning are still being rapidly pushed, the governance response varies widely across national contexts. The report identifies three distinct governance pathways shaping smart education policy: **rights-based precautionary regulation** (EU/France), **state-driven strategic integration and sovereignty-oriented models** (China, Japan, Qatar), and **market-led fragmented adoption environments** (Brazil, United States). By analyzing these case studies through a set of core ethical pillars, the report highlights cross-cutting risks and opportunities and proposes actionable policy recommendations to support innovation while protecting learners' rights, educational integrity, and public trust.

Core Ethical Pillars

Panel 2: Core Ethical Pillars

ETHICAL PILLAR	EXECUTIVE TAKEAWAYS
Education is a High-Risk AI Domain	▫ Education concentrates multiple high-stakes AI uses (grading, placement, profiling, tutoring). Errors and bias can directly shape long-term life trajectories and social mobility. Because learners are minors, education requires safeguards stronger than most sectors.
Datafication of Childhood is the Central Structural Risk	▫ Smart education relies on continuous data extraction and longitudinal learner profiling. Persistent "digital dossiers" risk extending beyond school contexts and time horizons. Retention limits and deletion rights remain weakly operationalized in most systems.
Contestability Gaps Create "Black Box" Schooling	▫ Many AI-assisted educational decisions lack transparent rationale and traceability. Contestability requires enforceable procedures (appeals, documentation, timelines, accountability). Fragmented governance increases the risk of unreviewable outcomes affecting students' futures.
Generative AI Expands Ethics Into Safety and Integrity	▫ GenAI introduces misinformation and hallucination risks that undermine learning reliability. Deepfakes are emerging as a major child safety threat, including sexualized harassment. Education governance must address AI as a safety and integrity issue, not only privacy.
Teacher Agency is the Primary Safeguard—Yet Under Pressure	▫ AI can support instruction, but also risks deskilling educators into compliance moderators. "Human-in-the-loop" rhetoric often masks structural incentives toward automation. Ethical deployment requires teachers to retain authority to override AI outputs.
Governance Architecture Determines Outcomes More Than Technology	▫ Three pathways shape ethics: rights-based regulation, state-driven integration, and fragmented adoption. The same tools yield different outcomes depending on procurement rules and oversight capacity. Weak governance environments tend toward reactive enforcement after harm occurs.
Sovereignty vs. Privacy is an Emerging Strategic Dilemma	▫ Building sovereign educational AI requires representative data while protecting minors' rights. Excess dependence on foreign platforms risks structural lock-in and loss of control. Trusted public data spaces plus privacy-preserving methods offer the most viable compromise.

1. Fairness & Allocative Justice: Mitigating bias in predictive analytics used for student tracking, streaming, and admissions.

The distinction between treating students *fairly* and treating them *equally* is foundational to contemporary debates on educational justice. Equality assumes that all learners begin their educational trajectories from comparable starting points and therefore require identical resources. In reality, students' opportunities are profoundly shaped by pre-existing social, economic, and institutional inequities—including those arising from systemic racism and structural discrimination. As a result, equal treatment often reproduces rather than redresses disparities. Fairness, by contrast, requires prioritizing the needs of historically marginalized and under-represented students and their families to counterbalance accumulated disadvantages.

Allocative justice in education entails the equitable distribution of resources, opportunities, experiences, and outcomes so that all learners—regardless of race, gender, ethnicity, language background, disability status, sexual orientation, family structure, or socioeconomic position—can attain meaningful readiness for both career and life. Crucially, equity does not mean offering every student *the same* resource, but rather ensuring that each student has timely access to the specific supports, pedagogical rigor, and culturally responsive content they require to succeed. Fairness involves differentiated, context-sensitive provision that advances genuine equality of opportunity and supports human flourishing across diverse learner communities.

Artificial intelligence holds significant potential to advance fairness and allocative justice by enabling personalized learning pathways, enhancing accessibility through real-time content translation, supporting instructional delivery via intelligent tutoring systems, automating assessments (customized to the needs of individual students, and keeping humans in the loop), and generating actionable insights through learning analytics. Yet concerns persist about the capacity of algorithmic systems to replicate and entrench existing inequities when trained on biased data or deployed without appropriate safeguards. Research demonstrates that such systems can inadvertently encode disparities linked to protected characteristics such as race, nationality, gender, home language, parental education, military affiliation, and socioeconomic status.

Addressing these risks requires a comprehensive and ethically grounded approach to the design and governance of predictive analytics used in student tracking, course placement, and admissions. Key elements include rigorous data-quality standards, continuous human oversight, clear mechanisms for transparency and explainability, and regular auditing to detect disparate impacts. The overarching objective is to ensure that algorithmic tools promote allocative justice rather than perpetuate historical patterns of exclusion, thereby contributing to a more equitable and socially responsive education system.

2. Contestability & Explainability: The right of students to challenge algorithmic grading ("Black Box" opacity).

In the context of artificial intelligence, *the concern of explicability and goal of explainable AI* (*XAI*) concerns the capacity of a system to illuminate the rationale underlying its decisions, thereby promoting transparency and interpretability. *Contestability*, by contrast, focuses on the mechanisms through which users can challenge, appeal, and influence those decisions—enabling not only understanding, but also action, correction, and redress. As AI systems increasingly inform high-stakes decisions, including in the education sector, contestability has emerged as a central safeguard for users' rights, dignity, and autonomy. It is conceptualized as a dynamic, participatory process that allows stakeholders to interrogate AI-generated outcomes, prompt reconsideration, and obtain revised determinations. In doing so, contestability embeds accountability and human oversight within AI governance.

Explainability is widely regarded as a necessary precondition for meaningful contestability. Without a clear understanding of how and why an AI system has reached a given conclusion, users cannot effectively question or contest its outputs. Scholarly literature distinguishes between *descriptive explanations*—which merely describe model behavior—and *justificatory explanations*, which offer normative reasoning that directly supports inquiry and challenge. Effective contestability therefore requires explanations that are contextualized, actionable, and calibrated to the informational needs of different user groups. When integrated into interactive system design, explainability and contestability become mutually reinforcing: transparent, user-centered AI systems empower stakeholders to exercise informed oversight and foster trust in automated processes.

These principles are particularly important given the longstanding problem of *black-box opacity*, a defining characteristic of many advanced AI systems. Black-box opacity refers to the difficulty—often impossibility—of understanding the internal logic, decision pathways, and parameter weightings of complex models such as deep neural networks. Because these systems learn highly abstract and non-linear representations from data, their outputs cannot easily be traced back to interpretable reasoning steps, even by experts. This opacity presents significant challenges for external validation, ethical auditing, and regulatory verification, underscoring the critical importance of embedding explainability and contestability into AI systems intended for educational and other high-stakes applications. In addition, it is critical to ensure greater inclusivity of training data and decision rules to mitigate bias.

3. Data Minimization: moving beyond "consent" to strict limitations on data retention

The notion of consent in digital environments has come under increasing scrutiny, particularly when users interact with large corporate platforms that wield disproportionate bargaining power. In such contexts, individuals are often compelled to accept extensive data-processing terms as a precondition for accessing essential services. This asymmetry undermines the voluntariness of consent and calls into question its legitimacy as a meaningful safeguard for user autonomy. Consequently, scholars, regulators and digital rights advocates argue for moving beyond a model that relies primarily on user consent toward one that structurally embeds privacy protection within AI systems themselves. Central to this shift is the principle of data minimization, which places the burden on organizations—rather than users—to limit the collection, use, and retention of personal data to what is strictly necessary for a clearly defined task.

Data minimization is a foundational data-protection principle that mandates the collection, processing, and storage of only the minimal amount of personal information required to fulfill a specific and legitimate purpose. It emphasizes necessity, relevance, and proportionality. For instance, while an e-commerce platform may justifiably require an address for order delivery, collecting additional attributes such as marital status would exceed what is necessary and introduce avoidable privacy risks. By constraining the scope of data collection in this manner, organizations strengthen public trust, reduce the likelihood of misuse or breach, and align their practices with regulatory frameworks such as the EU General Data Protection Regulation (GDPR).

This principle is operationalized through three interrelated components: **purpose limitation, collection minimization, and storage limitation**.

- Purpose limitation requires that personal data be collected only for specified, explicit, and legitimate purposes, and prohibits further processing that is incompatible with those purposes.
- Collection minimization mandates that personal data be adequate, relevant, and limited to what is strictly necessary for the stated aims.
- Storage limitation requires that personal data be retained only for the duration necessary to fulfill the specified purpose. Where continued use is justified (e.g., research or system improvement), data should be irreversibly anonymized only when re-identification risk is demonstrably low under realistic linkage attacks; otherwise, it should be treated as pseudonymous personal data and governed accordingly. Because metadata can enable re-identification through unique combinations and external linkage, ethical implementation also requires metadata minimization, strict controls on granularity (e.g., coarse time/location), and routine testing for re-identification and attribute inference risks.

Together, these principles provide a coherent framework for reducing the scale and sensitivity of data collected by AI systems. By limiting data flows at every stage of the lifecycle, they significantly mitigate risks of privacy infringement, discrimination, and security breaches, while reinforcing organizational accountability for responsible data stewardship. Ultimately, shifting from a consent-dependent model to one grounded in structural protections such as data minimization represents a critical step toward safeguarding individuals' rights and freedoms in an era of pervasive and increasingly opaque AI technologies.

4. Human-Centric Decision Making: ensuring that AI remains a support tool, not the final arbiter of student potential.

Human-centered AI (HCAI) refers to the design, development, and deployment of artificial intelligence systems that explicitly foreground human needs, values, and capabilities. Rather than seeking to replace human judgment or automate human roles, HCAI emphasizes the creation of technologies that augment human abilities, promote well-being, and reinforce social and ethical commitments. This approach recognizes that AI systems operate within complex cultural, social, and institutional contexts, and therefore must be designed to be accessible, usable, and beneficial to diverse populations. HCAI is closely aligned with the field of human–AI interaction, which studies how humans and intelligent systems communicate, collaborate, and share decision-making authority.

Implementing a human-centered approach requires interdisciplinary collaboration throughout the development lifecycle, which goes from the development of general-purpose AI tools based on LLMs (embodying AI's "horizontal capabilities") to their adaptation to specific domains and tasks and sometimes incorporation of specialized data sets for this purpose (AI "vertical capabilities") in what some refer to as "small language models." In the proposed approach designers and engineers work alongside psychologists, ethicists, educators, and domain experts to ensure that AI systems are transparent, interpretable, and accountable. This collaborative process situates HCAI within the broader movement for ethical AI, highlighting the importance of fairness, respect for human rights, and responsiveness to cultural and individual differences.

In the context of education, a human-centered orientation entails prioritizing the needs and experiences of students, teachers, and communities. AI technologies are deployed not to supplant human pedagogical expertise, but to enhance personalized learning, support learner diversity, and reduce administrative burdens in ways that preserve educators' professional agency. Human-centered AI for education also places strong emphasis on privacy protections, ethical data use, and the promotion of creativity, critical thinking, and equitable learning opportunities. Designing AI tools with real users in mind—ensuring transparency, providing meaningful user control, and positioning AI as an assistive partner rather than a dominant force—helps safeguard human dignity and autonomy while leveraging technology to strengthen educational outcomes.

These pillars are not exhaustive; rather, they provide a normative baseline through which national strategies can be assessed and compared.

A human-centered orientation entails prioritizing the needs and experiences of students, teachers, and communities.

Systemic and Developmental Risks

As students begin engaging with large language models (LLMs) and smart education platforms to assist their learning, they are being exposed to multiple risks that could affect their livelihood and human development. These risks can take the form of the datafication of childhood, cognitive outsourcing, surveillance from external stakeholders, and labor displacement for teachers/educators. These include:

- **The Datafication of Childhood:** The creation of permanent, interoperable digital dossiers that follow students into adulthood.
- **Cognitive Outsourcing:** The risk of atrophy in critical thinking and writing skills due to over-reliance on generative tools.
- **Surveillance:** Ethical implications of gaze-tracking, emotion recognition, and attention-monitoring software in classrooms.
- **Labor Displacement:** The shift from "teacher augmentation" to de-professionalization and scripted instruction.

Datafication of Childhood

All aspects of a student's childhood, through increased datafication and smart education (especially if a student is utilizing a single platform throughout the entirety of their education), could be documented and follow them throughout their lives. For example, behavioral issues, the presence of sensitive medical data (including medical disorders), juvenile criminal records, their personal choices, and an array of other profiling topics about the student can be exploited and potentially sold. This sensitive data in a child's developmental years could potentially be accessed by employers, credit agencies, or other third parties that could take advantage of such data.

Cognitive Outsourcing

Reliance on AI at a younger age can result in students being unable to conduct tasks such as writing and computing. Students would be unable to verify correct information and generate original thought without an algorithm assisting them. The risk here is that students will be unable to develop complex reasoning skills, and unable to interrogate and verify with an LLM further—based on blind trust of AI. Smart education platforms create this risk, if they do not encourage independent thought and reasoning as a basis for a student education.

Surveillance

Smart education creates a platform that can be exploited for ubiquitous surveillance. Gaze-tracking cameras, emotion recognition software, and attention-monitoring wearables make the classroom a place where students are highly vulnerable. Surveillance can also result in a culture of penalizing students whose learning styles or responses do not adhere to how a smart education model has been trained. The classroom can put students at risk of high levels of stress and behavior compliance, depending on how the smart education platform is trained. If students know they are being watched, they will self-censor and comply, rather than engage in genuine inquiry and learning.

Labor Displacement

Looking at the future of work, smart education platforms have the potential to displace or degrade the work of teachers. This societally essential work entails substantive expertise, pedagogical skills, emotional labor, and psychological support, among other skillsets. While smart education in many forms could be a co-pilot or assistant to teachers to help with grading and other tasks, their role could be shifted from being a primary educator to at best a "human in the loop" that is moderating AI. Such an outlook would result in lower skills requirements for overseeing smart education, foreclose teachers' autonomy and creativity in customizing content such as lesson plans, and erode sensitivities and emotional care that are otherwise brought in by educators. Educators can also play the role of understanding a child's needs at home. A smart education platform that displaces or trivializes the work of educators would put students who are in vulnerable positions at increased risk. Where teachers and educators are involved in co-creating AI tools, and/or their associations and unions are involved in this process or setting up the terms for engagement and participation in implementation in collective or other forms of bargaining or consultation (as some are beginning to in some national and local contexts), then smart education might more fully realize its potential for reskilling work and enhancing pedagogical practice, and mitigate the potential to downgrade and deskill or displace invaluable human labor and the professional training and expertise it embodies.

Methodology and Case Study Selection

This report applies a comparative qualitative approach to examine the ethical governance of smart education across diverse national contexts. It draws on publicly available national strategies, regulatory texts, policy guidance documents, and secondary research. Findings are synthesized through a shared ethical framework focused on fairness, transparency, privacy, accountability, and human-centered decision-making in education.

The selection of case studies was designed to ensure both analytical diversity and regional relevance.

Case Study Selection Criteria and Scope

Case studies were selected using five criteria:

- **Governance model diversity:** capturing distinct approaches to AI governance in education, including rights-based precautionary regulation, state-driven strategic integration, and market-led fragmented adoption.
- **Regional representation:** ensuring coverage across major geopolitical regions shaping global AI and education policy debates (Europe, North America, Asia, MENA, and Latin America).
- **Policy relevance to high-risk use cases:** prioritizing contexts where AI intersects with learner profiling, generative AI tutoring, biometric surveillance, and automated classification or placement.
- **Global influence and transferability:** focusing on countries with strong policy signaling power and relevance for cross-country learning.
- **Availability of public documentation:** ensuring sufficient accessible material to support transparent desk-based analysis.

While Africa is not represented as a dedicated case study in this edition, this reflects scope and documentation constraints rather than a lack of relevance. Education is a top regional priority in Africa according to FII's 2025 PRIORITY Compass, and future iterations of this work would benefit from dedicated Africa-focused case studies examining smart education governance under infrastructure, capacity, and resource constraints.

Panel 3: Three Governance Models in Smart Education

GOVERNANCE MODEL	EXECUTIVE TAKEAWAYS
Education is a High-Risk AI Domain	▫ Anchors smart education in enforceable rights frameworks (e.g., high-risk classification, data protection, human oversight). ▫ Strength: clearer accountability, stricter limits on misuse of minors' data, stronger baseline trust. → **Watch-out:** privacy constraints can complicate access to representative training data and slow iteration unless public data infrastructure exists.
Datafication of Childhood is the Central Structural Risk	▫ Treats smart education as national infrastructure for competitiveness, capacity-building, and AI readiness (often linked to sovereignty goals). ▫ Strength: coordinated scaling, investment in platforms/compute, and potential to embed literacy and national standards quickly. → **Watch-out:** principles may outpace operational safeguards—especially education-specific contestability, retention limits, and transparent accountability.
Contestability Gaps Create "Black Box" Schooling	▫ Adoption is driven by decentralized procurement and institutional autonomy, with uneven safeguards across jurisdictions and providers. ▫ Strength: rapid experimentation and innovation diffusion; localized policy pilots and sandboxes can emerge quickly. → **Watch-out:** "deployment first, governance later" increases exposure to surveillance drift, vendor opacity, and reactive enforcement after harm occurs.

CASE STUDY	RATIONALE FOR INCLUSION
European Union (France)	▫ Included as a representative example of a rights-based and precautionary regulatory model, where education is increasingly governed through enforceable safeguards such as the GDPR and the EU AI Act. France also illustrates the strategic tension between innovation ambitions and strong child-centered privacy protections.
Qatar	▫ Included as a leading example of state-driven digital transformation in education, where AI is positioned as part of a national human capital and sovereignty agenda. Qatar is also distinctive for framing Arabic language capability as a structural fairness issue rather than a secondary localization feature.
Brazil	▫ Included as a case of rapid decentralized deployment under a legislative vacuum, where AI adoption in education often outpaces governance safeguards. Brazil illustrates the risks of procurement-led experimentation and the role of courts and prosecutors in shaping reactive enforcement.
China	▫ Included as a prominent model of state-coordinated scaling and sovereignty-oriented governance, with strong national platforms and structured guidance on generative AI use in schools. China also provides insight into how smart education strategies intersect with regional inequality and national digital infrastructure priorities.
Japan	▫ Included as a model of human-centric AI governance aligned with the Society 5.0 vision, emphasizing literacy, wellbeing, and responsible integration of AI into social systems. Japan offers a valuable lens on balancing innovation with integrity concerns in higher education and creative sectors.
United States	▫ Included as a globally influential case of market-led and federally fragmented adoption, where education governance is highly decentralized and shaped by EdTech markets. The U.S. illustrates both rapid innovation diffusion and heightened ethical risk exposure, particularly regarding student privacy, surveillance tools, and deepfake-enabled harms.

Country Case Studies

Panel 4: "Cross-Country Findings"

ETHICAL PILLAR	EXECUTIVE TAKEAWAYS
Education is a High-Risk AI Domain	▫ Education concentrates multiple high-stakes AI uses (grading, placement, profiling, tutoring). Errors and bias can directly shape long-term life trajectories and social mobility. Because learners are minors, education requires safeguards stronger than most sectors.
Datafication of Childhood is the Central Structural Risk	▫ Smart education relies on continuous data extraction and longitudinal learner profiling. Persistent "digital dossiers" risk extending beyond school contexts and time horizons. Retention limits and deletion rights remain weakly operationalized in most systems.
Contestability Gaps Create "Black Box" Schooling	▫ Many AI-assisted educational decisions lack transparent rationale and traceability. Contestability requires enforceable procedures (appeals, documentation, timelines, accountability). Fragmented governance increases the risk of unreviewable outcomes affecting students' futures.
Generative AI Expands Ethics Into Safety and Integrity	▫ GenAI introduces misinformation and hallucination risks that undermine learning reliability. Deepfakes are emerging as a major child safety threat, including sexualized harassment. Education governance must address AI as a safety and integrity issue, not only privacy.
Teacher Agency is the Primary Safeguard—Yet Under Pressure	▫ AI can support instruction, but also risks deskilling educators into compliance moderators. "Human-in-the-loop" rhetoric often masks structural incentives toward automation. Ethical deployment requires teachers to retain authority to override AI outputs.
Governance Architecture Determines Outcomes More Than Technology	▫ Three pathways shape ethics: rights-based regulation, state-driven integration, and fragmented adoption. The same tools yield different outcomes depending on procurement rules and oversight capacity. Weak governance environments tend toward reactive enforcement after harm occurs.
Sovereignty vs. Privacy is an Emerging Strategic Dilemma	▫ Building sovereign educational AI requires representative data while protecting minors' rights. Excess dependence on foreign platforms risks structural lock-in and loss of control. Trusted public data spaces plus privacy-preserving methods offer the most viable compromise.

I. Rights-Based Precautionary Regulation

France

France pursues a strategy for artificial intelligence that balances ambition with a responsible mindset, in alignment with the European risk-based approach.

In the "AI for Humanity" strategy presented in 2018 in the Villani report, AI governance should not hinder innovation, but ensure that AI is deployed strategically from an ethical design perspective. Villani emphasizes that, to mitigate risks while fostering innovation, Europe must develop public data infrastructures to avoid structural dependence on major private platforms. The precautionary principle should be grounded in a clear-sighted, measured, and responsible vision that goes beyond purely technological considerations.

He argues for a collective governance of AI, explicitly questioning who designs it, for whom, and for what purposes. AI should be conceived as a tool serving the public interest, particularly in key sectors such as health, the environment, transport, defense, and education, rather than as a mere driver of economic or financial profitability. He also draws attention to the ecological and social costs of AI, notably the environmental impact of digital infrastructures (data centers, energy consumption) and the risk that "digital ecology" may be overlooked in the absence of a big picture systemic perspective.

THE REGULATORY FRAMEWORK: AI ACT AND GDPR

The European AI Act classifies certain educational uses of AI systems as "high-risk," as they may directly influence learners' futures (guidance, assessment, equity). This classification entails reinforced requirements in terms of transparency, human oversight, documentation, bias management, and auditing.

In parallel, the GDPR imposes enhanced protections for minors' personal data, limiting both data collection and secondary data use. In educational contexts, this means that any use of data must be clearly justified, proportionate, and secure.

THE KEY CHALLENGE: RECONCILING PRIVACY PROTECTION AND THE SOVEREIGNTY OF EDUCATIONAL MODELS

The central challenge lies in reconciling the protection of privacy with the sovereignty of educational models. Developing AI systems tailored to European pedagogical contexts requires representative, contextualized, and ethically sourced datasets, while access to students' real data remains tightly regulated. The dilemma is twofold: protecting minors from any form of surveillance or misuse, while at the same time avoiding leaving the field open to non-European private actors who already possess massive datasets and can impose their tools.

The solutions proposed by Villani and other experts include the creation of public educational data spaces that are anonymized, secure, and governed collectively; the establishment of European standards for ethics, transparency, and auditing; and the use of technical approaches such as synthetic data, federated learning, and advanced anonymization techniques.

Within this framework, the *France 2030* strategy aims to position the country among the leading European countries in AI by supporting innovation in key areas such as embedded AI, trustworthy AI, frugal AI, and generative AI. In education, however, the introduction of AI primarily raises questions about its impact on cognitive development processes, the development of students' intellectual, relational, and critical capacities, and their education for citizenship.

BALANCING QUALITY EDUCATION AND INNOVATION

AI should therefore be introduced gradually, in line with students' age and level of development, with particular caution before upper secondary education in order to safeguard the development of critical thinking. Its use by teachers must be transparent and responsible, taking into account both its pedagogical relevance and its ecological impact.

The 2024 "Enfants et écrans" (Children and Screens) report aligns with this perspective: it warns against the harmful effects of excessive digital exposure while acknowledging that, within a structured framework and under educational supervision, digital technologies can serve as effective pedagogical tools. Nevertheless, it prioritizes approaches

that foster human interaction and foundational learning outside of screens, especially for younger children, and emphasizes the development of psychosocial skills (critical thinking, emotional regulation, and the management of digital behaviors) as a prerequisite for a reflective relationship with technology.

In a context in which mental health has been designated as a major national cause in 2025, the French state is attempting to subtly reconcile sometimes contradictory orientations: on the one hand, restricting the use of screens and mobile phones in schools (for example through the *"portable en pause"* [phone-free breaks] policy in lower secondary education), and on the other hand, promoting training digital technologies and AI in educational policies. This tension reflects an effort to balance the preservation of educational quality, the protection of young people, and economic competitiveness.

Several legislative proposals under discussion in late 2025 aim to limit, or even prohibit, the use of social media in order to reduce screen time linked to declining educational performance. More broadly, France faces a persistent tension between ambitions for innovation and ethical concerns, particularly regarding health, copyright, data protection, and the safeguarding of minors. While initiatives are being rolled out to integrate generative AI tools into education or to develop sovereign and secure models, other public policies simultaneously seek to prevent the harmful effects of excessive digital consumption (57% of young adults aged 18–24 feel they spend excessive time on platforms–streaming, gaming, social media- of which social media is a major concern in the educational field.

II. State-Driven Strategic Integration

Qatar

GOVERNANCE INSTRUMENTS AND ORIENTATION

Qatar frames "smart education" as part of a state-led digital transformation agenda oriented toward national capacity-building. The National AI Strategy (2019) links AI deployment to human capital development through an "AI+X" paradigm (including AI+Education) and presents AI education as a cross-level curriculum priority. The governance problem in education is straightforward: scaling data-driven personalization and analytics without normalizing permanent learner surveillance and unequal outcomes. For AI governance, the public record points to a layered stack rather than a single education regulator. The AI Committee is described as a cross-government coordination mechanism. Qatar's Ministry of Communication and Information Technology (MCIT) has Principles and Guidelines (2024) which set non-binding ethical expectations for AI development and deployment, and they explicitly include students among intended stakeholders. They also identify risks associated with generative AI, including misinformation/inaccuracy and deepfakes. NCSA's 2024 guidance is presented as cybersecurity and information-security-oriented direction for organizations adopting AI, with emphasis on risk management and protective controls. "Qai" (launched in December 2025) is positioned publicly as a Qatar Investment Authority subsidiary focused on developing and investing in AI infrastructure and "secure and trusted" AI systems across sectors. This matters for education because

it signals that "sovereignty" is pursued not only through procurement choices, but also through domestic compute capacity—though public documentation does not, at this stage, specify an education-only mandate.

Against this governance and infrastructure backdrop, education emerges as a sector where the ethical stakes of AI deployment become especially concentrated. Even when policy texts remain general, educational settings consistently aggregate high-risk AI uses: the construction of longitudinal learner profiles through adaptive learning systems and learning analytics; the deployment of generative AI for tutoring, feedback, or assessment support, where risks of misinformation and bias are non-trivial; automated or semi-automated classification practices such as placement, risk scoring, or performance prediction; and forms of administrative automation that redistribute responsibility for decisions and errors across human and technical actors. These exposure points help explain why questions of language, equity, and governance design become central in the Qatari case discussed below.

LANGUAGE SOVEREIGNTY AS EQUITY INFRASTRUCTURE

A notable feature of Qatar's framework is treating Arabic language capability as a fairness and inclusion issue, rather than a cosmetic localization feature. The Strategy identifies Arabic language processing as a strategic priority (Pillar 5), explicitly acknowledging that English-centric AI ecosystems can be a poor fit for Arabic linguistic and cultural contexts. The Fanar initiative operationalizes this orientation through Arabic-first model development, including dialect support. In education terms, the implication is clear: if the language layer is structurally underserved, "personalization" becomes selective access, and exclusion gets baked into the infrastructure.

OPEN QUESTIONS AND POTENTIAL ROADMAPS

The governance framework is coherent at the level of national strategy and general-purpose guidance, but education tends to expose the places where principles need auditable sector standards. First, smart education commonly depends on continuous data collection and longitudinal learner profiles. In the publicly accessible AI guideline layer, education-specific guardrails on learner-data retention periods, deletion rights for minors, and strict purpose limitation for learning analytics are not clearly operationalized as enforceable sector rules. The practical

The governance framework is coherent at the level of national strategy and general-purpose guidance, but education tends to expose the places where principles need auditable sector standards.

risk is the datafication of childhood: persistent dossiers built from performance and behavioral traces with unclear boundaries on scope and duration. Second, contestability is stated at the level of principle, but procedures are the hard part. MCIT's guidelines articulate a general right for affected users to question and challenge decisions that affect them. What remains unclear in the public record is education-specific operationalization—student and parent-facing routes for appeal, timelines, documentation standards for automated or AI-assisted decisions, and named accountability points for high-stakes outcomes such as placement or performance prediction.

KEY OBSERVATIONS

Qatar's case is useful for comparison because it treats "sovereign AI" as an institutional and infrastructural project, not only a list of ethical principles. Qai's recent launch underscores the infrastructure dimension of sovereignty. The language axis (Arabic-first capacity) shows how equity questions can be pushed down to the infrastructure layer rather than handled as after-the-fact translation. A continuing area of development is the translation of national strategy and cross-sector guidance into education-specific standards with visible accountability and child-centered protections, particularly around data minimization, retention and erasure clarity, and workable contestability mechanisms.

Japan

MAINSTREAMING SOCIETY 5.0 AS THE NORTH STAR

The Government of Japan publishes the Science, Technology, and Innovation Basic Plan every five years to outline its strategic goals. The Basic Plan published in 2016 first introduced the concept of Society 5.0, which was further developed in the Sixth Basic Plan published in 2021. In the Sixth Basic Plan, Society 5.0 is defined as "a society that is sustainable and resilient against threats and unpredictable and uncertain situations, that ensures the safety and security of the people, and that enables individuals to realize diverse forms of well-being."

The widespread adoption of generative AI since 2022 has significantly accelerated its societal impact. In response, the Seventh Science, Technology, and Innovation Basic Plan, scheduled for publication in March 2026, will include a strategic framework focused on AI Readiness. The foundations of AI Readiness have been discussed in the Social Principles of Human-Centric AI, which aim to realize an "AI-ready society" as part of the Society 5.0 vision. These principles propose seven core elements: (1) Human-Centricity, (2) Education and Literacy, (3) Privacy Protection, (4) Ensuring Security, (5) Fair Competition, (6) Fairness, Accountability, and Transparency, and (7) Innovation.

The principle of Education and Literacy seeks to foster a broad societal understanding of AI, grounded in ethical considerations, in order to prevent intentional misuse. In both formal education (K–12 and higher education) and workforce reskilling, implementation must be inclusive to avoid creating societal divides in access to learning.

EDUCATION AND AI

In higher education, universities are providing guiding principles for students, faculty, researchers, and staff regarding the use of AI. Universities are being challenged to determine how best to cultivate critical thinking, logical reasoning, and creativity in the age of AI. At the same time, it is essential for students to acquire the AI literacy required by society. In addition, strong emphasis must be placed on cultivating integrity and ethical awareness.

In the research community, scholarly publishing is currently facing a fundamental challenge related to the misuse of AI, particularly the risk of AI hallucinations. As a result, universities are increasingly required to provide education in AI ethics for researchers, faculty, and students to ensure the ethical use of AI in academic research.

In K–12 education, the Ministry of Education, Culture, Sports, Science and Technology (MEXT) has initiated the GIGA School Program, which aims to provide one connected digital device (i.e., a tablet) per student. The GIGA School Program is advancing the adoption of digital textbooks, including the potential integration of AI-ready features. In addition, MEXT has issued guidelines for K–12 educators and staff regarding the use of generative AI, with the aim of strengthening information literacy.

KEY ISSUES

Japan is experiencing a sustained population decline, resulting in structural labor shortages across a wide range of economic sectors, a trend that is projected to intensify in the future. These shortages are particularly acute in non-urban regions. AI and physical AI are often viewed as labor-substituting solutions to address these structural shortages. However, such labor substitution increases the need for workforce reskilling. Meeting this emerging educational demand represents a critical role that educational institutions are expected to fulfill.

Similarly, the Agency for Cultural Affairs oversees Japan's soft power, as represented by film, anime, games, music, manga, and other creative arts. The Agency has issued a document titled "General Understanding on AI and Copyright in Japan," which explains how copyright law may apply to AI. The document examines potential risks of copyright infringement at two key stages: the AI training phase, in which copyrighted works may be used as training data, and the content generation phase, in which AI systems produce new outputs. In addition, it discusses whether and under what conditions AI-generated materials (i.e., content) may be eligible for copyright protection.

On the flip side, universities, creative companies, and artists and designers are exploring how AI can be embraced as a new creative partner. Through practices such as AI manga, AI art and AI-generated music, AI is giving rise to an emerging creative genre—one that has the potential to expand the boundaries of human expression and fundamentally reshape the future of the creative arts.

Japan is experiencing a sustained population decline, resulting in structural labor shortages across a wide range of economic sectors, a trend that is projected to intensify in the future.

China: State-Driven Sovereignty Model

China's Ministry of Education (MoE) has taken strong efforts to boost AI-enabled education by taking various measures. The MoE initially issued the Action Plan for AI Innovation in Higher Education Institutions (Action Plan) in April 2018, which specifies the utilization of AI technologies to support the reform of teaching methods. Soon after, the MoE established the AI Technology Innovation Expert Group in July 2018, which is responsible for conducting AI-related research, consultation and guidance on promoting the Action Plan. In August 2018, the MoE issued the Notice on Launching Pilot Programs for Using AI to Support Teaching Faculty Development. Ningxia and Beijing Foreign Studies University (BFS) were selected into the pilot programs and would receive tailored guidance on using AI in education from the MoE, as well as funding support. Attention should be paid that Ningxia is located in the west of China and home to many ethnic minorities, while BFS is located in the capital, therefore, it is evident that revealed the MoE has intended to balance regional development in using AI in education. In 2022, the MoE launched the National Digital Strategy Action for Education and the "Smart Education of China" platform. Since then, the MoE has propelled the use of AI in primary and middle schools, and vocational institutions, broadened the pilot programs, and released samples of AI deployment in education. Nevertheless, China considers 2025 as the inaugural year of smart education in China, ambitious to integrate AI into the entire process of education and teaching in future.

REGULATION

There are two instructive documents on the use of GenAI in education. The first is the Guide to Using GenAI in Primary and Secondary Schools (2025 Edition) and the second is the Guide for Teachers to Use GenAI (Version I). Both documents provide samples on how to use GenAI in education and emphasize the protection of personal privacy, data security, and compliance with relevant existing law. The Guide to Using GenAI in Primary and Secondary Schools (2025 Edition) restricts students' use of GenAI, e.g., students in primary schools are prohibited from using GenAI independently, students in junior middle schools are only permitted to a certain extent to explore the logic of generated content, and students in senior middle schools can use GenAI to conduct inquiry-based learning by following technological principles. In contrast, the Guide for Teachers to Use GenAI (Version I) is more to equip teachers with AI literacy and knowledge.

The guardrails on the use of AI in education are embedded in China's existing law and regulation. Pertinent legal documents include but are not limited to the Personal Information Protection Law, Data Security Law, Interim Measures for the Administration of Generative Artificial Intelligence Services, Measures for the Security Management of the Application of Facial Recognition Technology, and the updated Cybersecurity Law.

KEY ISSUES

Three issues have appeared in deploying AI in education in China. First, GenAI has not been fully adapted to the pedagogical models and most AI tools can only assist teaching rather than providing personalized teaching, which means the displacement of teachers has not become a conspicuous issue in China, while an urgent task is to integrate AI with the subject knowledge structure and teaching logic. Second, although the MoE has tried to balance the development of AI-enabled education among different regions, due to the reality of regional disparities in technological and economic development, the deployment of AI in education is more advanced in the eastern regions than the west, especially in terms of vocational education. Third, guiding college students to use AI ethically is a thorny issue. Not all the universities in China provide AI literacy courses and some college students are not fully aware of the risks of using AI. Due to the wide spread of AI models on the market, college students can easily resort to AI tools and may rely on AI to write papers without checking the accuracy of generated content.

III. Market-Led / Fragmented Adoption

Brazil: Judicial Activism vs. Legislative Vacuum

Unlike the U.S. federalist fragmentation, Brazil has a centralized legislative tradition, yet the deployment of AI in education is aggressively decentralized among its 27 states. Without a specific AI law—as the "Brazilian AI Act" is still pending—public administrators in states like São Paulo and Paraná have rushed to procure AI solutions from private vendors to "modernize" efficiency. This has created a landscape of "deployment first, governance later," where complex surveillance tools are installed in public schools before ethical guidelines are established. At the same time, current Brazilian policy discussions increasingly frame artificial intelligence not as a substitute for teachers but as a catalyst for redefining professional roles and competencies within the education system. Recent guidelines under development by the Ministry of Education (MEC), including the national framework for responsible AI in education and the "AI in Basic Education" reference a process launched in Brasília in 2025, emphasise that educators must remain at the center of AI deployment as pedagogical mediators, ethical supervisors, and critical interpreters of algorithmic outputs. Rather than replacing teaching functions, AI is expected to reshape teachers' work by reducing administrative burdens, supporting lesson planning and assessment, and expanding personalised learning strategies, while simultaneously increasing the demand for continuous professional development in digital literacy, data governance, and ethical oversight. This emerging "augmentation paradigm" reflects a broader policy consensus that the future of smart education systems in Brazil depends on strengthening teacher agency, not automating it, and on ensuring meaningful human-in-the-loop governance across all AI-enabled educational environments. Although the MEC is currently drafting guidelines, in practice, state governors have the autonomy to sign contracts with EdTechs. The result is a clash between rapid technological adoption by the Executive branch and corrective actions by the Judiciary and Public Prosecutors. To address this, the Ministry of Education (MEC) is now organizing a public call for a regulatory sandbox on AI to test solutions in the education sector to avoid lawsuits related to the use of AI applications in public schools.

REGULATION

Brazil currently operates in a regulatory vacuum specific to AI in education, and governance relies on a patchwork of general laws rather than sector-specific mandates. The Federal Constitution of 1988 guarantees privacy as a fundamental right but lacks specific digital provisions. The LGPD (General Data Protection Law), modeled after the GDPR in Europe, requires specific parental consent for minors and "best interest" standards, making it currently the only effective legal tool to challenge abuses. Legislative efforts are underway with Bill 2338/23; if passed, this bill would align Brazil with the EU AI Act, classifying emotion recognition in schools as "excessive risk"—effectively banning it—and other educational AI as "high risk," mandating rigorous algorithmic impact assessments (AIA). Currently, there is an enforcement gap: in the absence of the AI Act, regulation is reactive, driven by Public Prosecutors (Ministério Público) or lawyers filing lawsuits after harm has occurred rather than ex-ante compliance.

The State of Paraná deployed facial recognition across its public network under the guise of administrative efficiency to automate attendance-taking. A conflict arose because, while the stated purpose was bureaucratic, the contract with the private vendor included provisions for emotion analysis and engagement monitoring without public transparency or legitimate purpose. This constituted an ethical failure, as the state effectively turned classrooms into laboratories for emotional surveillance, cross-referencing students' attention with the National Curriculum (BNCC) and exceeding the legal mandate. The impact was severe due to the lack of a Data Protection Impact Assessment (DPIA), which led to the mass collection of minors' biometrics without granular consent, resulting in false negatives for absent students and no recourse for families. This case typifies the Brazilian resolution pattern: the Executive branch implemented a "black box" system, and the Judiciary and the Public Prosecutor intervened to halt it based on general privacy laws (LGPD), highlighting the urgent need for the specialized governance proposed in Bill 2338/23.

MULTILATERAL CONTEXT

Finally, this domestic tension now unfolds alongside a broader multilateral movement: in 2025, BRICS countries signed a Joint Declaration on Artificial Intelligence in Education, formally establishing a technical and professional cooperation alliance focused on ethical use, capacity-building, and the exchange of best practices. While the declaration signals a political commitment to South–South cooperation and shared concerns regarding AI in education, its practical contribution to governance remains uncertain. The initiative lacks binding regulatory mechanisms, clear accountability structures, and enforceable safeguards, raising questions about its capacity to meaningfully influence domestic regulatory gaps such as those currently observed in Brazil. Rather than resolving the tension between rapid deployment and weak ex-ante oversight, the BRICS framework may, at best, function as a soft coordination platform—one that can complement, but not substitute, robust national legislation, independent supervisory authorities, and context-sensitive ethical standards grounded in Brazilian constitutional and data protection principles.

This constituted an ethical failure, as the state effectively turned classrooms into laboratories for emotional surveillance

United States

Policies and governance frameworks regarding use of AI in education in the United States are shaped and constrained by federalism, the absence of comprehensive AI or privacy regulations, and deference to markets and technology companies and in this instance "ed tech." Federalism means education is primarily a state and local responsibility, with only eight percent of total education spending in 2022–2023 coming from federal sources such as the Department of Education (DOE). The first timid steps toward developing an overall AI governance framework were taken in October 2023, with Executive Order (E.O.) 14110 on "Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence." EO 14110 directed federal agencies to create AI safety regulations, and set goals like mandatory reporting for powerful AI systems, establishment of an AI Safety Institute, and developing a risk management framework. In October 2024 under the outgoing government the DOE issued a Toolkit for Safe, Ethical, and Equitable AI Integration to give guidance to both primary and secondary education (K–12) and post-secondary institutions for responsible adoption of AI tools. Among the principles outlined were: humans in the loop for high-stakes decisions with automated tools such as in grading and disciplinary actions; caution against discriminatory potential and clarification of relevance of civil rights laws to AI's use in education; careful vetting by schools of AI tools that could harm data security and with transparency for parents; need for guidance on deep fakes; and promoting AI literacy.

DEVOLUTION OF POWER

In this context, setting AI policy for K–12 schools as in other aspects of education has fallen to the state and local level, while public (state) university systems and private universities generally set their own policies for higher ed. As of October 2025, 31 of the 50 states had issued some sort of AI policy framework for primary and secondary education. While the state policies vary widely, common themes are: preparing the workforce of the future; human-centered and responsible use to augment rather than replace humans; AI literacy for professional development; equity, inclusion, and accessibility regardless of student backgrounds or abilities; data privacy, security, and safety; consideration of how AI can be used to improve pedagogy and learning outcomes; and broadening access to computer science education. Moreover, whether exercising the considerable autonomy they sometimes have or in response to state guidance, some school districts have forged ahead with AI policies One increasingly common approach among states that are most strongly pushing for innovation balanced with guardrails—led by governments from each of the two major parties—is state oversight boards and/or regulatory sandboxes, in which there can be flexible but supervised experimentation.

Massachusetts is an example of states trying to balance these concerns. The state's Technology Collaborative and STEM Advisory Council are partnering with the NGO Project Lead the Way (PLTW) to pilot an AI STEM curriculum. In parallel, the state Department of Elementary and Secondary Education has issued guidance on use of AI for school districts, which stresses five guiding principles of data privacy and security, transparency and accountability, bias awareness and mitigation, human oversight and educator judgment, and academic integrity. Experimentation with AI is happening in many ways here, as teachers "generate rubrics, lesson plans, and instructional materials," and students "draft essays," "brainstorm ideas," or "[translate] text for multilingual learners" and district officials use AI to schedule, adapt assessments, and allocate resources.

In public and private universities and university systems, the adoption of policies regarding AI use in the classroom and course management has become increasingly common across the U.S.—a 2023 survey indicated 58 percent of students indicated their program or university had an AI policy, while a separate 2025 survey showed a figure of 80 percent. Yet training is still weak and spotty in how to use such tools in ways that promote rather than detract from critical thinking, analytical reasoning, research skills, student career preparation, and other key learning goals and outcomes. There is much debate and controversy and considerable experimentation and trial and error. Students increasingly use AI tools but are ambivalent about their use and about their role in shaping instruction, and their use is extremely difficult if not impossible to detect in the way that traditional plagiarism was through online tools, generating both false negatives and false positives. Concerns about displacement or degrading of human labor of faculty and teaching assistants are often expressed by faculty and graduate school unions as well as university professor associations.

KEY ISSUES

The growing if contested use of facial recognition technology in schools as well as the potential for deepfakes of students are key ethical challenges. A sizeable and growing minority of schools use facial recognition technology (FRT), typically under contracts with private vendors in schools for purposes of monitoring attendance and security, partly in response to the increase in school shootings in recent years. However, FRT is known to be inaccurate (particularly for persons of color) and invasive of privacy, and introduces the potential for commercial use and misuse of students' biometric data. There is a tension with federal privacy safeguards for students contained in the Family Educational Rights and Privacy Act (FERPA, for student records) and Protection of Pupil Rights Amendment (PPRA, regarding marketing and surveys). Some districts have withdrawn plans to install cameras given resistance from parents. New York State became the first state to outlaw FRT in schools in 2023. While no other states have banned them specifically in schools, several states and some cities have banned or placed restrictions on FRTs or their use by law enforcement more generally. With regard to student privacy and the collection, storage, and usage of student images or other data by ed tech vendors and other third parties operating in schools, a patchwork of practices and regulations is also present. The most robust state law, California's Student Online Personal Information Protection Act (SOPIPA) "prevents online service providers from using student data for commercial purposes, while allowing specific beneficial uses such as personalized learning;" it has inspired passage of similar laws modeled on it in an additional 23 states (as of May 2025). An additional 16 states also have laws that regulate both vendors and schools though they are considered weaker than SOPIPA and SOPIPA-modelled laws, while the final 11 states have laws that regulate either vendors or schools but not both.

A particular problem is deepfakes, which have become easier to produce and more difficult to detect with gen-AI. All too frequently they are sexually explicit. The National Education Association (NEA) goes on to underline goes on to underline that "[f]or the victims (mostly girls), the emotional and psychological impact can be severe and long-lasting." Compared to traditional bullying, deepfakes can be disseminated quickly and go viral, and remain online indefinitely, re-victimizing their targets repeatedly. Obstacles to effective control—in addition to clear policy definitions and provisions to prevent, detect, and respond—are lack of financial and technical resources to identify them, gaps in training of teachers and staff, and "legal ambiguity when incidents originate off-campus or outside of instructional hours." This inability to guarantee verification allows learning environments to put students in danger to violence and potential harm.

The National Education Association underlines that "[f]or the victims (mostly girls), the emotional and psychological impact can be severe and long-lasting." Compared to traditional bullying, deepfakes can be disseminated quickly and go viral, and remain online indefinitely, re-victimizing their targets repeatedly.

Relationships to Ethical Pillars

Fairness and Allocative Justice appear most explicitly where countries treat equity as an infrastructure issue rather than a downstream outcome. Qatar provides a distinctive example through its Arabic-first AI strategy, positioning language capability as a prerequisite for fair personalization and inclusion. China similarly frames equity through a regional development lens, attempting to balance pilot programs across advanced eastern provinces and less-developed western regions, though disparities persist in practice. In contrast, the United States and Brazil illustrate how fragmented governance can produce uneven protections and thereby amplify inequity, particularly where surveillance-capable tools (e.g., facial recognition) may be deployed without consistent safeguards. The EU/France model addresses allocative justice primarily through risk-based regulation and strict protections for minors, embedding fairness obligations into compliance requirements for high-risk educational systems.

Contestability and explainability are strongest where governance systems provide enforceable requirements for transparency, documentation, and redress. The France/French approach institutionalizes these principles through its risk-based regulatory model, including expectations of human oversight, auditability, and user rights related to automated decision-making. By contrast, Brazil and the United States illustrate a more fragile contestability environment, where challenges to AI-assisted decisions often occur through litigation, state-level regulation, or ad hoc district policy rather than standardized national safeguards. Qatar explicitly articulates contestability as a principle, but the case study highlights an implementation gap: education-specific procedures for appeal, timelines, and accountability ownership remain insufficiently operationalized.

Data minimization emerges as a major point of divergence across governance models. France and the EU provide the clearest example of structural constraints through GDPR and the AI Act's high-risk logic, which together emphasize proportionality, purpose limitation, and enhanced safeguards for minors. In contrast, the United States demonstrates a fragmented privacy landscape in which student data protection depends heavily on state-level laws and vendor practices, increasing the risk of what has been described as the "datafication of childhood." Brazil offers an acute illustration of these risks: state-level deployments of facial recognition and behavioral monitoring proceeded without robust impact assessment, resulting in large-scale biometric data collection of minors without meaningful consent or recourse. Qatar similarly highlights the risk that personalization systems may generate persistent longitudinal learner profiles, but the case study notes that retention and deletion safeguards are not yet clearly operationalized into enforceable education-sector rules.

Human-centric decision-making is widely endorsed across contexts, yet remains vulnerable to implementation pressures. Brazil provides the most explicit articulation of a human-centered "augmentation paradigm," positioning teachers as pedagogical mediators and ethical supervisors rather than replaceable labor. The United States similarly emphasizes "humans in the loop" in federal guidance, but implementation varies widely across states, districts, and universities, and concerns about labor displacement and deskilling remain prominent. Japan's Society 5.0 vision reinforces human-centricity as a societal organizing principle, emphasizing AI literacy, integrity, and inclusion as prerequisites for responsible integration. China also advances teacher training and structured guidance on generative AI use, yet the case study notes persistent skepticism among educators when tools fail to align with pedagogical logic or reduce workload meaningfully. Across all cases, the key risk is that administrative efficiency incentives may push AI toward substitution rather than augmentation, weakening professional agency and undermining the relational and developmental functions of education.

Cross-Country Analysis

Overview: Three Governance Pathways Shaping Smart Education Ethics

Across the six case studies (European Union/France, Qatar, Brazil, China, Japan, and the United States), a consistent finding is that ethical outcomes in smart education depend less on technical sophistication than on the governance model through which AI is deployed. Three governance pathways emerge. The European Union (illustrated through France) reflects a rights-based and precautionary regulatory approach, embedding safeguards through GDPR and the AI Act. China, Qatar, and Japan represent state-driven integration models in which AI-enabled education is framed as strategic infrastructure for national readiness and competitiveness. The United States and Brazil illustrate fragmented adoption environments, where procurement and implementation often outpace governance and where safeguards vary widely across jurisdictions.

Within these models, smart education creates distinct opportunity and risk profiles for four stakeholder groups: educators, learners, policymakers, and private sector providers.

Figure 2: Public Perceptions of AI Readiness and Education Risk (FII PRIORITY Compass 2025)

COUNTRY	EDUCATION SYSTEMS CANNOT KEEP PACE WITH TECH (AGREE)	AI WILL INCREASE DIGITAL DIVIDE (AGREE)	SUPPORT GLOBAL AI RULES (AGREE)	COUNTRY PREPARED FOR DIGITAL AGE (AGREE)
Brazil	79%	69%	61%	73%
China	58%	85%	71%	52%
France	72%	71%	52%	72%
Japan	64%	71%	48%	51%
United States	69%	71%	63%	77%

These numbers are collated from respondents who selected "strongly agree" and "mostly agree" to each question.

1. Impact on Teachers and Educators: Opportunities and Ethical Challenges

Opportunities

Across all contexts, AI-enabled education is frequently positioned as an opportunity to enhance instructional delivery and reduce administrative workload. Tools such as automated lesson planning, intelligent tutoring systems, adaptive learning platforms, and AI-supported assessment design can allow educators to devote more time to individualized student engagement and core pedagogical work. In Brazil, emerging policy discourse explicitly frames AI not as a substitute for teachers but as a mechanism to strengthen teacher agency and professional capacity, reflecting an "augmentation paradigm" in which educators remain central to governance and learning outcomes. Japan similarly emphasizes education and literacy as ethical infrastructure, recognizing that teachers must be equipped not only to use AI tools but also to guide students in responsible and reflective use.

Ethical Challenges

Despite rhetorical convergence around human-in-the-loop principles, the case studies highlight a shared risk: efficiency incentives may gradually shift teachers from autonomous professionals into compliance moderators of algorithmic systems. In the United States, institutional autonomy gives faculty wide latitude to determine AI policies, but training and implementation are inconsistent, creating uneven capacity to supervise AI responsibly. France raises additional concerns about cognitive development and the need for cautious integration, particularly for younger learners, emphasizing that educational technology must not erode critical thinking or foundational human learning processes.

Best Practices and Governance Implications

The strongest safeguard across cases is the explicit positioning of educators as decision-makers rather than passive users. Effective systems require teacher training in AI literacy, institutional policies that guarantee educators the right to override AI recommendations, and clear rules preventing fully automated grading or disciplinary outcomes. Without these protections, smart education risks deskilling the teaching profession and weakening the relational and developmental functions of schooling.

2. Impact on Students and Learners: Opportunities and Ethical Challenges

Opportunities

AI can expand access to learning by enabling personalized pathways, improving accessibility for students with disabilities, and supporting multilingual learners through real-time translation and adaptive interfaces. Qatar provides a distinctive model of fairness-oriented infrastructure by treating Arabic language capability as a strategic inclusion priority rather than a cosmetic localization feature. Japan's human-centric framework similarly emphasizes that inclusive literacy is necessary to prevent societal divides in access to AI-enabled learning. China's national platform approach reflects an ambition to scale smart education across the system, including primary, secondary, and vocational institutions.

Ethical Challenges

Across all cases, the most prominent student-centered ethical risk is the "datafication of childhood." Learning analytics systems and adaptive platforms can generate persistent longitudinal learner profiles, tracking academic performance and behavioral signals in ways that may extend beyond educational necessity and outlive the student's schooling trajectory. Qatar explicitly recognizes this risk, noting that smart education can normalize permanent learner surveillance if retention limits and deletion rights are not clearly operationalized. Brazil illustrates how this risk becomes acute when surveillance technologies are deployed without transparency, such as facial recognition systems and emotion analysis embedded in school monitoring tools. The United States demonstrates similar vulnerability, where student data governance is fragmented and vendor contracts may introduce surveillance practices unevenly across districts.

Students are also exposed to growing epistemic and safety risks linked to generative AI. Japan highlights hallucinations and misuse in higher education and research environments, emphasizing threats to integrity and knowledge credibility. China similarly identifies misuse risks among university students, where national guidance and literacy training are uneven. In the United States, deepfakes represent an escalating child safety crisis, increasingly associated with bullying and sexualized harm, often spreading rapidly and remaining online indefinitely.

France and the United States also highlight a distinct student-centered concern: digital wellbeing. Both contexts reflect growing policy efforts to restrict screen exposure and mobile phone use in schools, acknowledging that excessive digital engagement may undermine attention, mental health, and cognitive development. This creates a policy tension: education systems seek to deploy AI for innovation while simultaneously restricting technology consumption to protect learners.

Best Practices and Governance Implications

Protecting learners requires enforceable constraints on student data retention, restrictions on surveillance technologies, and the development of child-centered contestability mechanisms. It also requires proactive safety infrastructure for generative AI harms, including deepfake detection and response protocols, digital citizenship training, and clear rules for acceptable GenAI use in schoolwork. Without these safeguards, smart education may undermine trust, safety, and long-term learner autonomy.

3. Impact on Policymakers and Public Sector Systems: Opportunities and Ethical Challenges

Opportunities

For governments, smart education offers the potential to strengthen system performance and efficiency through learning analytics, resource allocation optimization, and targeted intervention models. AI can support national competitiveness agendas by building workforce readiness and improving educational quality at scale. This is most explicit in China, Qatar, and Japan, where smart education is embedded in national AI readiness strategies. France similarly frames AI as a public interest tool, emphasizing strategic deployment in sectors including education while seeking to avoid structural dependence on foreign platforms through public data infrastructures.

Ethical Challenges

The case studies demonstrate that the primary governance challenge for policymakers is ensuring that ethical principles translate into enforceable sector standards. Qatar illustrates this tension clearly: while national principles and cross-sector guidelines are coherent, education-specific rules on retention, deletion rights for minors, and contestability procedures remain less visible in public documentation. Brazil and the United States show the risks of fragmented governance, where adoption is decentralized and often driven by procurement decisions rather than consistent national frameworks. In Brazil, the absence of a dedicated AI law has led to "deployment first, governance later," with courts and prosecutors acting as de facto regulators after harm occurs. In the United States, the lack of a comprehensive national privacy framework produces a patchwork landscape in which protections vary widely across states, districts, and institutions.

France highlights a second governance challenge: reconciling privacy protections with sovereign model development. Strong constraints on access to student data protect minors, but may complicate the development of representative European educational datasets, potentially increasing dependence on non-European private actors. China's experience highlights a related tension: state-led scaling is ambitious, but regional disparities in infrastructure and implementation capacity remain persistent.

Best Practices and Governance Implications

Across cases, the strongest policy interventions are those that treat education as a high-risk AI domain requiring dedicated governance instruments. This includes mandatory impact assessments, standardized procurement requirements, clear restrictions on biometric surveillance, enforceable contestability procedures, and the development of public educational data infrastructures governed by strict child protection rules. Policymakers must also treat digital wellbeing and generative AI harms as core governance domains rather than secondary concerns.

4. Impact on Businesses and AI Solution Providers: Opportunities and Ethical Challenges

Opportunities

Smart education represents a major growth opportunity for EdTech companies, AI developers, and infrastructure providers. Across all contexts, the demand for tutoring systems, translation tools, learning analytics platforms, and generative AI educational applications is accelerating. Qatar's investment in AI infrastructure and China's national platform scaling highlight the potential for large-scale public-private ecosystems. Japan's engagement with AI creativity and copyright governance illustrates the possibility for AI to expand educational content development and creative learning markets. In the United States, a market-driven ecosystem has enabled rapid experimentation and widespread vendor adoption across districts and universities.

Ethical Challenges

The central ethical challenge for businesses is that education is a uniquely sensitive environment in which minors' data, high-stakes decisions, and trust-based institutions converge. The Brazil case illustrates the reputational and legal risks associated with opaque vendor contracts, particularly where surveillance functions (such as facial recognition and emotion monitoring) exceed publicly stated purposes. In the United States, compliance with fragmented state laws and federal safeguards creates regulatory complexity, while deepfake harms and academic integrity controversies increase pressure for safety-by-design standards. In Europe, the high-risk classification model creates higher compliance requirements and may limit rapid scaling unless vendors can demonstrate auditability, fairness, and robust data governance.

Across contexts, the private sector also faces increasing pressure around sovereignty and localization. Qatar's Arabic-first model development illustrates how language capability is becoming a core ethical requirement. France's emphasis on public data infrastructures and reduced dependence on foreign platforms signals a broader trend toward trusted domestic capacity-building. China's regulatory and sovereignty environment similarly encourages alignment with national governance priorities.

Best Practices and Governance Implications

The case studies suggest that responsible innovation in smart education will require vendors to adopt higher standards of transparency, documentation, and auditability, including clear disclosure of model capabilities and strict limitations on secondary data use. Businesses must anticipate that trustworthiness will increasingly function as a competitive advantage, particularly in high-risk educational settings where governments are likely to demand demonstrable safeguards. Providers that design systems to support teacher agency, minimize data collection, and prevent misuse (including deepfakes and hallucinations) will be better positioned to scale sustainably across diverse regulatory environments.

Cross-Cutting Takeaways

Across stakeholder groups, the evidence suggests three overarching conclusions. First, smart education amplifies both opportunity and risk because education concentrates multiple high-risk AI uses simultaneously, including profiling, classification, assessment support, and generative content generation. Second, governance architecture is the dominant determinant of ethical outcomes: rights-based models provide stronger structural protections, state-driven models enable strategic scaling but require stronger contestability mechanisms, and fragmented systems generate uneven safeguards and heightened surveillance drift risk. Third, generative AI has expanded the ethical landscape beyond privacy and bias toward urgent child safety and integrity challenges, requiring proactive policy tools rather than reactive enforcement.

Across stakeholder groups, the evidence suggests three overarching conclusions.

- → Smart education amplifies both opportunity and risk.
- → Governance architecture is the dominant determinant of ethical outcomes.
- → Generative AI has expanded the ethical landscape beyond privacy

Policy Recommendations

Panel 5: "Policy Blueprint: 5 Priorities"

PRIORITY AREA	EXECUTIVE TAKEAWAYS
1. Treat Education as a High-Risk AI Domain	▫ Classify grading, placement, admissions support, disciplinary recommendations, and learner profiling as high-risk uses. Require pre-deployment assessments to evaluate bias, reliability, privacy, and security risks before scaling. Apply enhanced safeguards (documentation, transparency, human oversight, independent review) for consequential systems.
2. Prevent the Datafication of Childhood	▫ Enforce strict data minimization and purpose limitation for learning analytics and personalization systems. Establish retention caps, deletion/erasure rights for minors, and clear boundaries on secondary data use. Treat biometric and behavioral data as highly sensitive, with heightened safeguards and narrow necessity thresholds.
3. Make Contestability Real (Not Theoretical)	▫ Mandate notice when AI is used in consequential decisions and provide meaningful explanations to users. Create formal appeal and redress pathways with defined timelines, documentation standards, and named accountability owners. Require audit logs and traceability so decisions can be reviewed, corrected, and learned from institutionally.
4. Protect Teacher Agency and Human Oversight	▫ Prohibit fully automated grading, placement, or disciplinary outcomes without educator review and authority. Guarantee educator override rights and clarify accountability so humans remain responsible decision-makers. Invest in teacher training so AI augments professional judgment rather than deskilling the workforce.
5. Govern Generative AI as Safety and Integrity Infrastructure	▫ Establish clear integrity rules for acceptable GenAI use and verification expectations in student work. Implement deepfake prevention and response protocols (reporting, rapid escalation, victim support, takedown pathways). Build AI literacy and digital citizenship programs to mitigate misinformation and misuse across age groups.

Overview

The case studies demonstrate that ethical smart education depends on translating high-level principles into enforceable safeguards that protect children, preserve trust in learning environments, and ensure that AI strengthens rather than weakens human-centered education systems. While national contexts differ, the report identifies a set of practical governance priorities that can be adapted across regulatory models. These recommendations are organized by stakeholder group to support clarity, accountability, and implementation planning.

Public Perception (PRIORITY Compass 2025)

Public concern about the equity implications of AI-enabled education is already widespread. According to FII Institute's 2025 PRIORITY Compass, 76% of respondents globally agree that the use of AI in education may increase the digital divide, reinforcing the urgency of policy safeguards focused on access, inclusion, and equitable learning outcomes. This concern is reflected across multiple country contexts examined in this report, including China (85%), France (71%), Japan (71%), the United States (71%), and Brazil (69%). These findings suggest that public trust in smart education will increasingly depend on governments' ability to implement enforceable protections around fairness, data governance, and the prevention of exclusionary AI-driven learning systems.

1. Recommendations for Teachers and Educators

Strengthen Teacher Agency as a Governance Safeguard

Education authorities should explicitly define teachers as the primary decision-makers in AI-enabled education environments, ensuring that AI functions as a support tool rather than a substitute for professional judgment. Teachers should retain the right to override AI recommendations in grading, assessment, classroom interventions, and learning pathways without institutional penalty.

Institutionalize AI Literacy and Ethical Training for Educators

Governments and education institutions should establish structured teacher training programs on AI literacy, responsible pedagogical use of generative AI, and ethical supervision of algorithmic systems. Training should cover bias awareness, detection of hallucinations and misinformation, data governance basics, and appropriate classroom integration aligned with developmental needs.

Prevent Deskilling Through Human-in-the-Loop Standards

Policies should prohibit fully automated grading, disciplinary actions, or placement decisions without educator review. Human-in-the-loop requirements should be formalized not as a voluntary principle, but as a mandatory operational standard for high-stakes educational decisions.

Provide Educators with Practical Tools and Support Systems

Teacher-centered governance requires that educators have access to clear institutional guidelines, model usage rules, and practical classroom resources. Ministries of Education should support schools in developing standardized AI use policies, rather than leaving implementation to individual teachers without training or institutional backing.

2. Recommendations for Students and Learners

Protect Learners from the "Datafication of Childhood"

Education systems should adopt strict limitations on student data collection, retention, and reuse, recognizing that smart education models often rely on continuous data extraction and longitudinal learner profiling. Policies should mandate clear retention periods, deletion rights for minors, and strict purpose limitation for learning analytics systems.

Establish Clear Rights to Transparency and Contestability

Students and parents should be informed when AI is used in assessment, placement, disciplinary recommendations, or other consequential decisions. Education authorities should implement formal appeal and redress mechanisms that allow learners and families to challenge AI-assisted outcomes, request human review, and obtain timely correction.

Restrict Biometric and Emotion Recognition Technologies in Schools

Given the sensitivity and permanence of biometric identifiers, governments should impose strict limits—or outright prohibitions—on facial recognition and emotion inference technologies in educational settings. These systems should be treated as high-risk by default, with strong justification requirements and oversight mechanisms if permitted.

Address Generative AI Harms Through Safety and Integrity Protocols

Schools and education authorities should adopt explicit protections against generative AI-enabled harms, including deepfakes, cyberbullying, and misinformation. This includes clear reporting channels, response protocols, victim support mechanisms, and structured digital citizenship education to strengthen critical thinking and responsible technology use.

Embed Developmental and Digital Wellbeing Safeguards

AI integration should be age-appropriate and aligned with cognitive and psychosocial development needs. Policymakers should integrate screen-time governance and wellbeing safeguards into smart education strategies, particularly for younger learners, to ensure technology enhances rather than undermines attention, mental health, and foundational learning.

3. Recommendations for Policymakers and Public Sector Systems

Define Education as a High-Risk AI Domain

Governments should formally classify key educational AI systems—such as grading tools, placement algorithms, admissions support systems, disciplinary recommendation tools, and learner profiling platforms—as high-risk applications requiring enhanced oversight. This classification should trigger mandatory safeguards including documentation, transparency, auditing, and human oversight.

Require Pre-Deployment Impact Assessments

Before AI tools are deployed at scale in schools or universities, governments should require Algorithmic Impact Assessments (AIAs) or Education AI Impact Assessments. These assessments should evaluate bias risks, data governance practices, cybersecurity resilience, and potential harms to student rights and wellbeing.

Strengthen Procurement Governance as a Primary Policy Lever

Public procurement is often the decisive governance mechanism in smart education. Ministries of Education should develop standardized procurement frameworks requiring full vendor transparency, audit access, restrictions on secondary data use, and contractual accountability for harms, errors, or breaches. Procurement rules should also require disclosure of hidden surveillance features, such as emotion monitoring or behavioral inference.

Operationalize Contestability Through Enforceable Standards

Governments should establish sector-wide rules requiring documentation of AI-assisted decisions, clear audit logs, and formal appeals procedures with defined timelines. Contestability should be treated as a practical governance mechanism, not merely a conceptual ethical principle.

Develop Trusted Public Education Data Infrastructures

To reduce dependence on external platforms and mitigate sovereignty risks, governments should explore the creation of public educational data spaces governed by strict child protection safeguards. Where feasible, privacy-preserving approaches such as federated learning, synthetic data generation, and secure anonymization methods should be prioritized to support innovation without exposing minors' raw data.

Integrate Generative AI Governance into National Education Policy

Governments should recognize that generative AI introduces new systemic risks that extend beyond traditional data privacy concerns. Education policies should explicitly address deepfakes, hallucinations, academic integrity challenges, and misinformation, including regulatory requirements for content safeguards and institutional response systems.

4. Recommendations for Businesses and AI Solution Providers (Private Sector)

Adopt Trust-by-Design Standards for Educational AI

AI solution providers should treat education as a high-risk environment requiring stronger safeguards than general consumer AI. Vendors should build systems that prioritize data minimization, transparency, auditability, and robust child safety protections, including guardrails against harmful outputs and misuse.

Increase Transparency and Disclosure of Model Capabilities

Providers should be required to disclose system functions clearly, including whether tools collect biometric data, generate behavioral inferences, or use sensitive student information. Transparency must include documentation of training data limitations, known biases, and model performance across different student populations.

Align Business Models with Child-Centered Data Protection

Vendors should not rely on secondary commercialization of student data. Business models should be aligned with strict purpose limitation, minimal retention, and compliance with child data protection rules, recognizing that trust and legitimacy will increasingly determine market access.

Support Teacher Agency and Responsible Pedagogical Integration

AI providers should design products that strengthen educator control, enabling teachers to guide learning pathways, override AI recommendations, and monitor system outputs. Systems should support pedagogical logic rather than impose automation that reduces teacher autonomy.

Prepare for Emerging Sovereignty and Localization Expectations

Across multiple case studies, sovereignty concerns are becoming central. Vendors should anticipate increased demand for localized language capability, domestic hosting requirements, and compliance with national standards for trusted AI certification. Qatar's Arabic-first model development highlights that language capability is increasingly an equity and inclusion requirement, not a cosmetic feature.

Summary: Priority Actions for Ethical Smart Education

Across stakeholder groups, the report identifies five cross-cutting priorities that are essential to responsible deployment:

1. Treat education as a high-risk AI domain requiring enhanced safeguards and pre-deployment review.
2. Prevent the datafication of childhood through strict data minimization, retention limits, and deletion rights for minors.
3. Institutionalize contestability and accountability so learners and families can challenge AI-assisted decisions.
4. Protect teachers as governance anchors by ensuring AI augments rather than displaces professional judgment.
5. Address generative AI safety risks proactively, including deepfakes, misinformation, and academic integrity threats.

If implemented effectively, these recommendations can support a smart education ecosystem that accelerates learning outcomes while protecting human dignity, child wellbeing, and public trust.

Table 1: Smart Education Ethics: Policy Blueprint

POLICY PILLAR	WHAT GOVERNMENTS SHOULD DO	PRIMARY POLICY INSTRUMENTS	IMPLEMENTATION LEVEL
High-Risk Classification	Formally classify AI systems used for grading, placement, admissions support, and learner profiling as high-risk	AI risk taxonomy; mandatory impact assessments; certification or approval requirements	National regulator / Ministry of Education
Contestability & Redress	Ensure students and parents can challenge AI-assisted decisions and receive timely review	Right-to-explanation provisions; appeal procedures; documentation and record-keeping requirements	Ministry of Education / School systems
Data Minimization & Retention Limits	Limit student data collection and storage to what is strictly necessary, with clear deletion rights for minors	Retention caps; purpose limitation rules; deletion and erasure requirements	Privacy authority / Ministry of Education
Procurement & Vendor Governance	Standardize EdTech procurement rules to prevent hidden surveillance features and misuse of student data	Contract templates; disclosure requirements; audit clauses; restrictions on secondary data use	Ministry of Education / Local authorities
Biometric and Emotion Recognition Controls	Restrict or prohibit biometric surveillance and emotion inference technologies in school settings	Biometric bans or strict necessity tests; consent rules; mandatory impact assessments	Legislature / Ministry of Education

Table 1: Smart Education Ethics: Policy Blueprint

POLICY PILLAR	WHAT GOVERNMENTS SHOULD DO	PRIMARY POLICY INSTRUMENTS	IMPLEMENTATION LEVEL
AI Literacy & Institutional Capacity	Embed AI literacy and ethical training across students, teachers, and education institutions	Curriculum standards; teacher training programs; institutional guidance frameworks	Ministry of Education / Teacher training institutes
Generative AI Integrity & Child Safety	Address deepfakes, misinformation, and academic misuse through school safety and integrity protocols	Deepfake response protocols; academic integrity standards; reporting and escalation pathways	Ministry of Education / Child protection agencies
Teacher Agency and Human Oversight	Protect the educator's role as the primary decision-maker and prevent deskilling through automation	Human-in-the-loop requirements; override rights; workload and responsibility standards	Ministry of Education / School leadership
Sovereignty and Trusted Infrastructure	Develop trusted national education data infrastructures and reduce dependence on external platforms	Public education data spaces; federated learning; secure national compute strategies	National digital authority / Ministry of Education
Digital Wellbeing Safeguards	Ensure smart education does not undermine cognitive development, attention, and mental health	Age-appropriate restrictions; screen-time governance; school device policies	Ministry of Education / School systems

Conclusion

Artificial intelligence is rapidly reshaping education systems worldwide, accelerating the emergence of "smart education" models that integrate adaptive learning platforms, predictive analytics, and generative AI into the core functions of teaching, assessment, and administration. The case studies in this report demonstrate that while these technologies can strengthen personalization, accessibility, and system-level efficiency, they also concentrate ethical risks in ways that are uniquely acute in educational settings. Because AI systems in schools operate on minors' data and influence high-stakes decisions that shape long-term life trajectories, education must be treated as a high-risk domain requiring safeguards that exceed those applied in many other sectors.

Across the six country contexts, the report identifies a shared governance challenge: the ethical outcomes of smart education depend less on the sophistication of technology than on the regulatory and institutional architecture that governs its deployment. Rights-based models such as the European Union emphasize precaution, enforceable safeguards, and strong data protections. State-driven strategies such as those in Qatar, Japan, and China frame smart education as national infrastructure for AI readiness and competitiveness, but often face challenges in translating principles into operational accountability mechanisms. Fragmented environments such as the United States and Brazil illustrate the risks of rapid procurement-led deployment without consistent ex-ante oversight, creating conditions for surveillance drift, unequal safeguards, and reactive governance driven by litigation or crisis response.

The comparative findings further show that smart education produces distinct opportunities and risks across stakeholder groups. For teachers and educators, AI offers the promise of reduced administrative burdens and enhanced instructional support, but also introduces the risk of deskilling and professional displacement if human oversight is treated as symbolic rather than enforceable. For students and learners, AI can expand inclusion through personalization and language accessibility, yet it simultaneously raises the risk of the "datafication of childhood," in which continuous monitoring creates persistent learner profiles with unclear boundaries on retention, reuse, and accountability. For policymakers, AI-enabled education offers new tools for system optimization and workforce development, but demands governance mechanisms that ensure transparency, contestability, and child-centered protections. For the private sector, education represents a major innovation frontier, yet one in which trustworthiness, auditability, and responsible data governance will increasingly determine legitimacy and market access.

A central conclusion of this report is that ethical smart education cannot be achieved through voluntary principles alone. The growing diffusion of generative AI has introduced new integrity and safety risks—hallucinations, misinformation, academic misuse, and deepfake-enabled harassment—that require proactive institutional safeguards rather than reactive responses. At the same time, rising concerns about student mental health, screen exposure, and digital wellbeing underscore that education policy must account not only for data protection and bias mitigation, but also for developmental and psychosocial impacts.

The pathway forward is therefore clear: smart education must be governed as a public-interest transformation. This requires enforceable standards that protect minors' rights, constrain surveillance and biometric monitoring, institutionalize contestability and redress, and preserve teacher agency as a central pillar of educational integrity. It also requires investment in ethical infrastructure—AI literacy, secure data governance systems, and trusted education technology ecosystems that align innovation with accountability. If these conditions are met, AI can strengthen learning outcomes and equity while reinforcing public trust in education systems. If they are not, smart education risks entrenching inequality, normalizing surveillance, and weakening the human foundations of learning.

List of Authors

- **Yanis Ben Amor**
  Center for Sustainable Development, Columbia University. USA
- **Antoine Blondelle**
  Université Catholique de Lille, France
- **Xiuyan Fei**
  East China University of Political Science and Law, China
- **Mehdi Ghassemi**
  ISTC/Université Catholique de Lille, France
- **Cristina Godoy**
  University of São Paulo, Brazil
- **Masa Inakage**
  Keio University, Tokyo, Japan
- **Scott Martin**
  Columbia University/The New School for Social Research, USA
- **Nour Naim**
  Yeni Yüzyıl University, Istanbul, Turkey

The authors would like to thank Yatin Jain at Columbia University for his invaluable research and writing contributions, which were instrumental in the development of this report.

Appendix: FII PRIORITY Compass 2025

Education Findings

FII Institute conducted the FII PRIORITY Global Survey in partnership with Ipsos between July and August 2025, engaging more than 60,000 respondents across 32 countries (representing 66% of the global population). The participants represented regions and diverse demographics across the planet, encompassing all walks of life (age, gender, residential status, employment status, occupation, Et Al). Their voices reflect the true mosaic of humanity and capture the shared hopes, challenges, and aspirations that define our global story today.

→ FII PRIORITY Compass Navigator

Findings:

1. **In the top 10 global priorities, the Education system ranks #6.**
2. **Among the global top 6 concerns, #5 is "Good education for your children":** 76% worry that AI in education will increase the digital divide, favoring students with better access to technology
3. **Education System ranking in the top 5 priorities per region:**

- **Africa:** education system ranked #4 priority
- **Asia:** ranked #5 priority
- **Europe:** didn't make it to top 5 priorities
- **MENA:** ranked #3 priority
- **North America:** didn't make it to top 5 priorities
- **Oceania:** didn't make it to top 5 priorities
- **South America:** didn't make it to top 5 priorities

4. Responses to the statement **"Our education systems are unable to keep pace with technological change":**

COUNTRY	STRONGLY AGREE	MOSTLY AGREE	MOSTLY DISAGREE	STRONGLY DISAGREE
Brazil	34%	45%	17%	4%
China	19%	39%	33%	9%
France	22%	50%	23%	5%
Japan	10%	54%	31%	5%
Qatar	N/A	N/A	N/A	N/A
United States	24%	45%	26%	5%

5. Responses to the statement **"The use of AI in education will increase the digital divide, favoring students with better access to technology":**

COUNTRY	STRONGLY AGREE	MOSTLY AGREE	MOSTLY DISAGREE	STRONGLY DISAGREE
Brazil	24%	45%	24%	7%
China	25%	60%	13%	2%
France	22%	49%	22%	7%
Japan	9%	62%	24%	5%
Qatar	N/A	N/A	N/A	N/A
United States	24%	47%	21%	8%

6. Responses to the statement **"There should be a global framework/rules to regulate artificial intelligence":**

COUNTRY	STRONGLY AGREE	MOSTLY AGREE	MOSTLY DISAGREE	STRONGLY DISAGREE
Brazil	22%	39%	26%	13%
China	23%	48%	24%	5%
France	20%	32%	33%	15%
Japan	8%	40%	38%	14%
Qatar	N/A	N/A	N/A	N/A
United States	27%	36%	23%	14%

7. Responses to the statement **"My country is prepared for the digital age (e.g., education, skills, training, broadband coverage)":**

COUNTRY	STRONGLY AGREE	MOSTLY AGREE	MOSTLY DISAGREE	STRONGLY DISAGREE
Brazil	25%	48%	21%	6%
China	14%	38%	40%	8%
France	25%	47%	24%	4%
Japan	7%	44%	43%	6%
Qatar	N/A	N/A	N/A	N/A
United States	33%	44%	20%	3%