inspera

Leading the Way in Digital Assessment

Trustworthy AI

What is it, how can we use it and create effective policy?

inspera.com

Contents

Introduction

The potential impact of AI in education can't be overstated. Whether it's the implementation of AI, generative or otherwise, adaptations to assessment and policy, competing views on the virtues of AI, or the rapid advancement of AI inserting itself into almost every conversation, it has been an outsized part of academic discourse over the last two years.

We know that what we choose to offer in this space has to solve meaningful problems rather than just be bolted onto what you do today. With all of that in mind, this white paper set out our position: Trustworthy AI.

To this end, we look at what Trustworthy AI is, how AI can be deployed with purpose, how we might redesign assessments in the face of it, ethical student use of AI, and finally, how institutions might draft their AI policy.

What is Trustworthy AI?

Trustworthy AI is about adopting an intentional, measured, and human-centric approach to the design, implementation, and use of artificial intelligence. It's the acknowledgement that AI is not a solution by itself, but instead that its value comes from how it works in harmony with educators and their expertise, not as their replacement. This places emphasis on the skill and expertise of the educator as primary and paramount, which cannot be understated.

Four fundamental principles underpin Trustworthy AI: choice, purpose, transparency and adaptability.

Choice ensures that AI is not forced upon the unwilling, or mandated where its use would be inappropriate.

Purpose ensures that AI is deployed in response to address a specific challenge, because simply adding a generative AI interface to an existing technology solution is purposeless due to an educator having the ability to open in another tab, then copy and paste their results. A clearly articulated purpose is, for example, improving integrity through surfacing performance data on a question or assisting a candidate as they take an assessment.

Transparency underlines the need for technology providers to explain how their technology works and for educators to understand the meaning and impact of it.

Finally, adaptability in how AI is deployed, which is something much broader than choice. Where an educator chooses to use it, having the ability to 'dial up and dial down' the AI or add context such as a syllabus makes it a sharper tool instead of a blunt instrument.

By framing AI through the lens of trustworthiness, with principles that underpin the concept, educators can engage with the technology through their customary critical and constructive lens that is central to education. In turn, this empowers educators to make informed decisions that align with their pedagogy.

Choice: We Don't See AI as Mandatory

A fundamental principle of Trustworthy AI is the belief that adoption of AI should never be mandatory. Educators have always used their cognate expertise and wider pedagogical knowledge to create learning, teaching, and assessment materials suitable for their context and cohort. Valid and reliable assessments existed long before AI and will continue to do so with and without AI's input.

This aligns with a general guiding principle at Inspera; that we offer technology with an ability for an institution to decide if they want to deploy it or not, and furthermore, for the ability of the institution to decide if their educators use it in their context, or not.

We Understand the Needs of Education

Technology Decision
Inspera chooses what we develop and make available to the market in consultation with our customers. We can decide this based on market needs and based on our expertise in working with educators to enable them to deliver better student outcomes through assessment

Institution Decision
From what Inspera chooses to make available, an institution can decide what to switch on. They do that in collaboration with us; we as tech experts, them as pedagogic experts ("off" is a viable option)

Academic Judgment
From what the institution allow faculty to use, an academic can decide what to use on their assessment based on the specificity of the purpose of the assessment ("off" is a viable option)

We encourage institutions to ensure that educators have the final say. Where they choose at an institutional level to make AI, or indeed any capability available, the academic freedom to decide whether AI is right in context aligns with the trust reposed in them to design curricula and teaching methods.

This also fosters a culture of innovation. Educators who choose to adopt AI will do so out of genuine interest and perceived value, which in turn leads to a more thoughtful, effective, and impactful use of technology. Aligned with that is the ability for educators to 'switch on' AI in the background or after the event to see what impact AI would have had on their assessment without impacting real data. These no-stakes methods of seeing the impact of AI are important to the macro choices made by educators, and the subsequent outcomes in their learning environments.

Purpose: In The Service Of a Community

One of the most persistent concerns about AI is that it seeks to replace human roles. In the context of education, this concern can create significant resistance to AI adoption. AI should be an amplifier rather than a replacement; in the service of the community not running it. It is a tool designed to enhance the capabilities of educators, not to diminish their role or expertise. Where it can't or won't meet that standard, it won't and should not be used.

Education is and will remain a human endeavor. It is about relationships, communication, and the sharing of knowledge and values. While AI can assist in many aspects of this process, it cannot replicate the empathy, intuition, and creativity that educators bring to their work.

Finding a Balance Between Educator and AI

For example, an AI tool might analyze student performance data to identify areas for improvement, but it is the educator who interprets this data and decides how to address the identified issues.

AI has the potential to free up time so that an educator can use it for other tasks. Trite, but that requires AI to actually free up time. Absent purpose, context, and trust, AI generates data not insights, words not feedback, text not questions.

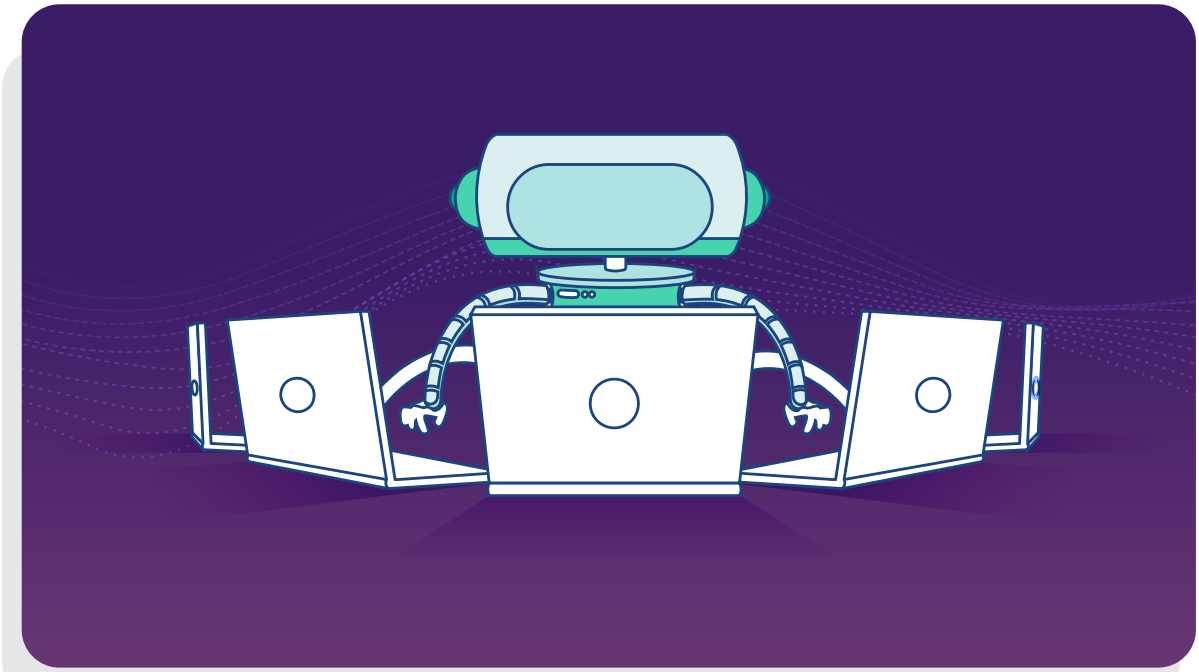
AI enablement comes with a corresponding responsibility to improve AI literacy as a subset of digital literacy. Without support, potential benefits remain untapped or worse, inadvertently misused. The trustworthiness of AI flows all the way through from the availability of a capability to the moment of use.

Educators need to understand how AI they might use, operates to the extent of what it can and cannot do in their context, and the implications of its use. This doesn't require educators to become AI experts to any greater extent than they wish to do so. A driver of a car needs to know how to operate it safely while not necessarily needing to know the internal workings of a combustion engine. But they are not prevented from becoming experts should they choose.

This understanding in context enables them to make informed decisions about where and how to deploy AI. For example, an educator might choose to use AI for administrative tasks but choose between wholly manual, partly AI or full AI for marking. Where AI is used, providing clear, accessible information about AI's capabilities and limitations empowers educators to use it effectively and responsibly.

Control also extends to data usage and privacy. Educators should have visibility of how much data is shared with AI systems. This builds trust and ensures that AI aligns with the ethical standards of the educational community.

It is critical educators retain ultimate control over assessment. AI can provide valuable recommendations or automate certain tasks, but the final decisions on what is used, should rest with the institution and educator. This approach ensures the balance of the human elements of education: empathy, ethics, and context, are primary. Ultimately, supporting the assertion that pedagogy comes before technology.



Adaptability: Context is Key

It is essential educators are able to undo the most recent action taken by AI. Whether it's a grading decision, content generation or modification, or an automated suggestion, educators must have the option to reverse AI-driven changes. This ensures that the human expertise of the educator remains the final authority, maintaining their professional autonomy and safeguarding the integrity of their work. It also allows for experimentation without fear of irreversible decision making.

In a similar vein, AI-generated recommendations such as feedback or a modification to a question, must be editable by the educator. This allows the expert to refine and adapt AI generated content to better fit their context and cohort. This collaborative relationship between educators and AI ensures that the technology supports, rather than dictates, pedagogy. Where possible, the ability for educators to adjust AI settings to meet their teaching context and curriculum provides a vastly greater benefit than a simple on/off switch. This includes being able to use a syllabus, curriculum, and other artefacts as a source from which AI works.

The ability to scale AI usage up or down is vital. As educators become more comfortable with AI, they may choose to expand its role in their workflows. Conversely, they might find that certain AI applications are not as effective as anticipated and change or end it's use. This flexibility is essential for creating a sustainable, long-term relationship between educators and AI technology.

Transparency: We Know You'll Only Use AI if it Meets Your Ethical and Data Privacy Standards

Educators want to understand what AI is doing and how it operates. This requires an understanding of the algorithms and processes that drive an AI capability. For example, if an AI tool is used to grade an assessment or generate feedback, educators need a high-level explanation of how those outputs were determined; e.g. the submission was parsed against the rubric. Transparency builds trust and ensures that educators can confidently rely on AI while understanding its limitations. Without this transparency, the risk of mistrust, misinterpretation or misplaced reliance on AI increases.

To build confidence in data practices, it is essential to provide educators with clear, accessible information about how their data is handled. This includes outlining what data is collected, how it is stored, who has access to it, and how long it is retained. It also involves providing educators with control over their data, allowing them to opt in or out of certain features or data-sharing arrangements.

Data ownership is a key part of this conversation, and educators need assurance that their data is secure and that it will not be used for purposes beyond those explicitly agreed upon. For example, where a syllabus and curriculum are used to inform and better contextualize an AI capability, it (the syllabus and curriculum) should not be available outside of the institution without explicit permission, nor should any derivative of it, e.g. an improvement to a large language model that comes from the data or use of it. Similarly, with student data. The data itself and subsequent treatment by AI should be accessible to the institution only, unless it chooses otherwise.

Where Do We See AI Having a Potential Purpose?

AI Authoring Assistant

The process of creating assessment questions is both an art and a science. Educators must craft questions that accurately measure learning outcomes, align with curriculum, and are valid and reliable. This process is necessarily time-intensive, and AI is capable of serving a purpose here.

With the important caveat of it having the right context, ie the syllabus and curriculum, AI can generate questions for an educator. Their efficacy depends on the level of instruction. Educators are already using existing LLMs to generate multiple choice questions (MCQ). Question generation however, is just one use case, and arguably the most simple. It's also one where there's a strong feeling that AI is only appropriate at the lowest level of instruction for simple MCQs. That may change in time as AI develops and educators are more confident adding their curriculum to an engine.

The more interesting use cases do not require a question to be generated by AI. In fact, they are predicated on the educator writing the question with AI providing analysis. As always, with the right contextual background, AI can serve as an invaluable assistant, offering insights that enhance human-authored questions without compromising the educator's control and creativity. An AI authoring assistant does not replace the educator's role in question design, instead, it serves as a collaborative partner, providing data-driven insights and creative suggestions while leaving the final decision in the hands of the educator. This approach respects the skill, expertise and judgment of an educator, ensuring that the questions remain aligned with their pedagogical goals as well as the needs of their students.

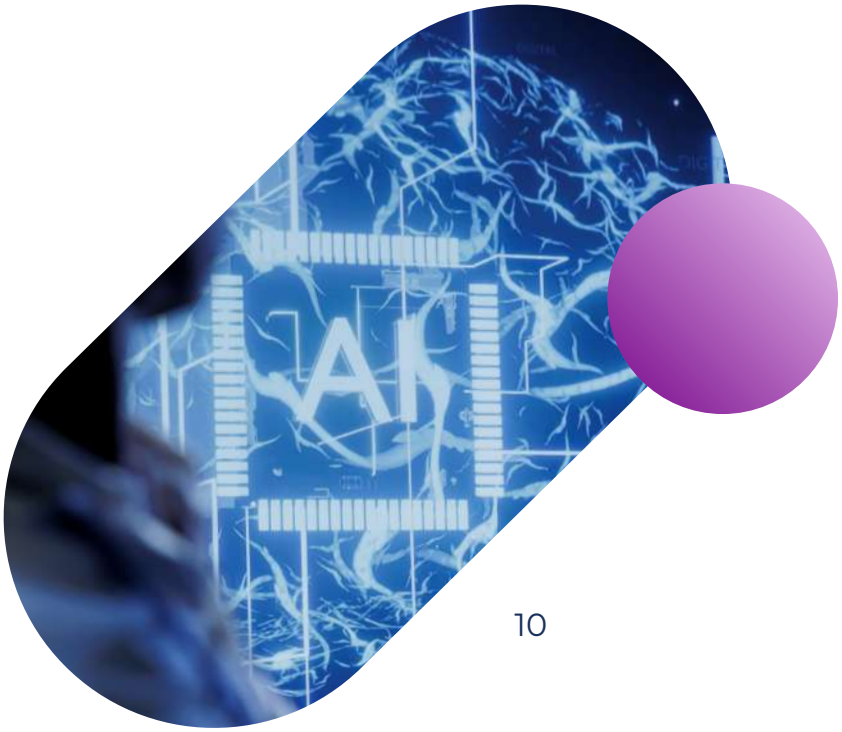
An AI authoring assistant can analyze how questions have performed in the past, providing insights on the difficulty level of a question, identifying whether it tends to challenge high-achieving students or needs adjustment to avoid being not sufficiently challenging. It could also highlight patterns of common misconceptions, allowing educators to adjust the question wording or associated learning materials to address gaps in understanding. This feedback loop ensures that each iteration of a question becomes more effective at assessing student learning and mastery.

Another capability of an AI authoring assistant is its ability to suggest alternative question types. For instance, an educator might write an essay question and receive AI-generated suggestions for converting it into a short-answer question, multiple-choice question, or even a numerical simulation. This feature allows educators to diversify their assessment strategies and cater to different contexts. A question that works well as an essay in one context could be broken down into diagnostic MCQs. By offering these alternatives, the AI authoring assistant empowers educators to explore new possibilities without starting from scratch.

Finally, integrity concerns can be addressed by an authoring assistant by analyzing a question for potential integrity risks based on past instances of misconduct or cheating. For example, it could flag questions that have appeared on the internet, or identify question types that are particularly susceptible to contract cheating or AI-generated answers. This insight allows educators to proactively address these vulnerabilities, whether by rephrasing the question, altering its format, or pairing it with additional integrity measures.

How an AI Authoring Assistant Could Benefit You

- Multiple choice question generation
- Providing data-driven insights and suggestions
- Assessing a question's difficulty based on previous data
- Addressing common misconceptions
- Suggesting alternative question types
- Flagging questions that have appeared online.



AI Marking Assistant

Marking assessments and providing rich feedback is both a crucial aspect of the assessment lifecycle and at the same time, hugely labor intensive. Any form of AI marking has to be cognisant of both points. Saving time at the expense of accuracy and meaningful feedback, serves no purpose and is damaging. An AI marking tool must provide the educator with valuable outputs that can be edited in order to be useful. For example, in a mathematics course, AI could automatically grade numerical problems, while the educator focuses on evaluating written explanations or proofs. This division of labor allows the educator to allocate their time where it is needed most.

AI marking is not currently capable of replacing human judgment in areas where subjectivity and nuance are paramount. For example, while AI might be able to identify grammatical errors or suggest improvements to an essay's structure, it cannot fully appreciate the creativity or originality of a student's work. These elements require the empathy and insight of a human educator.

Beyond full AI marking there are other applications of AI within this sphere of assessment, where, like an AI authoring assistant, technology acts in concert with the educator to improve outcomes. AI has the potential to serve as an additional marker, offer suggested feedback, and enhance the quality of educator-written feedback. These capabilities are not only capable of improving efficiency but also assist in students receiving timely, rich feedback.

A practical application of AI as an assistant in marking, is as an extra marker alongside human educators. In this role, AI is not replacing the marker but producing an additional perspective for moderation.

Second, AI can also assist with feedback by providing suggestions that educators can review and edit. For instance, the AI might analyze a student's essay and generate preliminary feedback on aspects such as structure, clarity, and alignment with the question. The educator can then refine this feedback, ensuring it reflects their perspective and understanding of the student's work. This approach streamlines the feedback process without sacrificing its quality or personalisation.

A final application of AI in marking is its ability to enhance educator-written feedback. An educator could write a comment on a student's performance, and AI could expand this feedback to include references to specific sections of the syllabus or curriculum. This added context helps students understand not only what they need to improve but also why it matters in the broader scope of their learning.

Metadata Analysis: Providing AI Insights for Assessment Integrity

Metadata, a set of data that describes and gives information about other data, can be quickly and effectively analyzed by AI. In the ambit of assessment it includes a wide range of information such as timestamps, submission patterns, activity logs, and interaction histories. Analyzed by AI, it can reveal insights that assist educators making decisions about academic integrity, identify anomalies, and ultimately make informed decisions about the conduct of assessments.

Crucially, it's the educator making the decision, not the AI. The division of labor is the AI doing the data crunching, a task it can do repeatedly at speed. The human expert can see the nuance and context to make a decision; what the human is best at.

How Metadata Could be Utilized

AI-powered metadata analysis can provide valuable context to help educators make informed decisions. Current static data on assessment access and interaction can be revisualized as a timeline of activity. This level of detail ensures that any investigations are thorough and based on clear evidence rather than assumptions. To this end, it also provides context on the following:

- AI-driven metadata analysis enables educators to uncover patterns in student behavior during assessments. It can detect unusual submission timings, such as students consistently submitting work within seconds of each other, which may indicate impermissible collaboration.
- Similarly, metadata might highlight instances where a student frequently revisits a question before submitting a final answer, raising questions about potential access to external resources.
- Anomalies in activity logs can also point to integrity concerns. Metadata analysis might identify students accessing the platform from multiple devices during a single exam session or logging in from unusual geographic locations.

All of the above is the sort of analysis that takes place today. AI is not making a decision, but instead, speeding up the process of bringing the data into one place. By flagging these irregularities, AI helps educators focus their attention on cases that merit further investigation.

Beyond identifying potential misconduct, metadata analysis can be used proactively to design more secure assessments. By studying historical metadata, educators can identify questions or formats that may need revisiting in their design.

Armed with this information, educators can take steps to redesign assessments. This can include using adaptive testing, randomization, pulling from a bank of questions or a wholescale change in question or assessment format. These insights allow for continuous improvement in assessment design, enhancing both integrity and fairness.

To ensure trust in AI-driven metadata analysis, we go back to transparency as a key part of the process. Educators need to understand what the AI processes, how it correlates data, and how outputs are derived. This transparency not only drives confidence but also ensures that students are treated fairly. Institutions must also establish clear policies around the use of metadata analysis, including guidelines for how flagged cases are reviewed and resolved, with educators making decisions, not the technology.

Assessment in the Age of AI

Trustworthy AI isn't just about the use of AI but what we might do with assessment design given AI's ubiquity. Even if you choose not to use AI, others will. The availability of AI, particularly generative AI for students, continues unabated, leaving assessment redesign as a necessity rather than an option. Assessment hasn't stood still and AI is not the only catalyst for its evolution - though perhaps it will increase the velocity of change.
Before considering how assessments might be adapted, student understanding of AI, its benefits, and drawbacks, need to be part of the same education as their program of study, wider skills, and digital literacy.

Student AI Literacy

Whether a subset of evolving digital literacy or a topic of its own, student AI literacy is a necessity both as part of their wider education but also as an inherent part of their educational journey. Students might, but can't be expected, to use AI appropriately without guidance. With unprecedented access to tools that are capable of enhancing their learning experience ranging from personalized study aids to sophisticated research assistants, the potential of AI to empower students is immense. However, alongside these opportunities come significant challenges. Particularly the risk of misuse. To fully harness the potential of AI while maintaining academic integrity, students must understand how to use AI ethically and responsibly.

The first step to using AI ethically is understanding its role in education. AI should be viewed as a tool to enhance learning, not as a substitute for effort or critical thinking. When students use AI responsibly, it can support their academic growth. AI-powered tools are capable of breaking down difficult concepts into digestible chunks. AI writing assistants can help students identify grammatical errors, suggest clearer phrasing, or refine their arguments. Applications that track learning progress and recommend study plans can help students prioritize areas where they need the most improvement.

But all of these potential positives have to be handled with care. The distinction between a machine correcting your work and understanding the feedback is vast; the latter fosters learning and advancement, the former does not. AI tools are not infallible; they can produce incorrect, biased, or misleading information. Blindly accepting AI-generated outputs can lead to errors in assessments and a lack of critical engagement with the material.

Disclosing AI Use

- If institutions decide to let their students use AI to assist with an assignment, they should disclose it.

- If an AI tool helped refine the structure of an essay or suggest citations, this should be acknowledged, just as they would cite a source in a bibliography.

Students should feel comfortable discussing their use of AI tools with their instructors. This openness allows educators to provide guidance and ensures that students use AI in a way that aligns with course objectives. By integrating the use of AI into these programmes, students can deepen their understanding of ethical practices and become advocates for integrity in their academic communities.

To help students navigate the ethical use of AI, the following practical guidelines can serve as a framework. Before using AI for a task, students should ask: Am I using this tool to learn or to shortcut my work? Am I following my institution's policies? Will my use of AI enhance my understanding of the subject? AI literacy is required to help them accurately answer these questions.

Students need to collaborate with AI instead of delegating to it. Instead of relying on AI to produce essays or projects, students can use it to generate outlines, suggest improvements, or critique their work such as using an AI writing assistant to refine grammar and clarity, while ensuring that the core ideas and arguments are entirely your own.

Responsible use of AI extends beyond academic work. Students should also consider the broader ethical implications of AI, such as its impact on privacy, fairness, and society.

Redesigning and Adapting Assessments

With students who are both AI literate and aware of the wider consequences of AI use, assessments can be redesigned to differing levels by incorporating the following strategies.

- Assessments should require students to document their thought processes, such as outlining their approach to a problem or providing drafts of an essay. This makes it harder for students to rely solely on AI-generated outputs.

- Oral presentations or viva examinations can be used to verify students' understanding of their work. For instance, after submitting an essay, a student might be asked to discuss its key arguments or defend their conclusions. This approach not only deters misuse but also helps students develop valuable communication skills.

- Group projects and peer assessments can reduce the likelihood of individual misuse by requiring students to work collaboratively and hold each other accountable.

- Assignments that require students to engage with real-world problems or apply knowledge to unique, personalized contexts are less susceptible to AI-generated solutions. Students might be asked to analyze local data, conduct interviews, or reflect on personal experiences, all of which require original input.

- Assignments can explicitly allow or even require the use of AI tools, provided students document how they used them and reflect on the results. A written assessment might include a section where students critique the suggestions made by an AI assistant and explain what they chose to accept or reject them.

- Encourage students to critically evaluate AI-generated outputs. A history assignment might involve using AI to generate a historical analysis ,which students then compare to primary sources and refine based on their own research.

None of these are immune from AI misuse. Not all of them are practical in every assessment. But that is equally true of any form of potential misconduct.

Developing a Comprehensive AI Policy

Developing a comprehensive AI policy for universities requires a thoughtful and focused approach that considers the unique challenges and opportunities presented by artificial intelligence in the educational environment. The policy must address both the **practical and ethical dimensions** of AI use, ensuring that its implementation enhances teaching and learning while preserving trust, academic integrity, and fairness.

One of the central considerations is the scope of AI use within the institution. Institutions must define the specific contexts in which AI tools will be applied, whether for administrative tasks, assessment design, grading, or student support. This specificity prevents ambiguity and ensures that educators and students alike understand the intended applications of AI. By **clearly defining the boundaries of AI use**, and training to that end, the policy helps align expectations and provides a framework for consistent implementation.

The ethical use of AI is a cornerstone of any effective policy. Institutions must ensure that the tools they endorse are free from bias and operate in ways that **uphold the principles of equity and fairness.** For instance, if AI is used in grading, it is critical to demonstrate that the algorithm treats all students equally, regardless of background or demographic factors. Regular audits of AI systems should be mandated to identify and rectify potential biases, with a clear commitment to transparency about how the tools function and make decisions.

Data privacy is another vital element of the policy. The use of AI often involves the collection and analysis of large amounts of data, including sensitive student information. Institutions must establish clear guidelines on how this data is collected, stored, and used, ensuring compliance with relevant legal frameworks such as GDPR, FERPA, or other regional privacy laws. Students and educators should be informed about what data is being collected and how it will be used, and they should have the opportunity to provide consent or opt out of certain data-driven functionalities. A transparent approach to data management not only builds trust but also ensures the ethical deployment of AI tools.

The policy must also **address the role of academic integrity** in the age of AI. For students, this means setting clear expectations about the acceptable use of AI tools in assessments. Institutions should distinguish between ethical uses, such as leveraging AI to refine writing or generate study plans, and unethical practices, such as submitting AI-generated essays as original work.
For educators, the policy should provide guidelines on how to use AI responsibly in assessment design and grading, ensuring that AI augments rather than undermines their professional expertise.

Training and education are key components of a comprehensive AI policy. Both educators and students need to be equipped with the skills and knowledge to use AI tools effectively and ethically. Institutions should offer workshops, resources, and ongoing support to help their community understand the capabilities and limitations of AI. Educators may need training on how to interpret AI-generated insights or how to adjust AI settings to align with their pedagogical goals. Similarly, students should be educated on the ethical implications of AI and how to use these tools responsibly in their academic work.

To ensure the policy remains relevant and effective, it must include **mechanisms for regular review** and adaptation. AI is a rapidly evolving field, and what works today may not be suitable tomorrow. Institutions should establish a process for gathering feedback from educators and students about the implementation of AI tools and their impact on teaching and learning. This feedback loop allows for continuous improvement, ensuring that the policy evolves alongside advancements in AI technology and changing institutional needs.

Finally, the policy should promote **transparency and accountability** at every level. Educators and students must understand how AI tools are chosen, implemented, and monitored. Institutions should be open about their partnerships with AI vendors, the criteria used to select tools, and the steps taken to ensure their ethical use. By fostering a culture of openness, universities can build trust in their AI initiatives and encourage responsible engagement with these emerging technologies.

About the Author

Ishan Kolhatkar is a Global Client Evangelist at Inspera and an education technology leader.

With over a decade years of experience in the intersection of technology and education, Ishan is committed to driving positive change and inspiring the adoption of modern assessment methodologies worldwide.

Before joining Inspera, Ishan was a legal academic following a career as a Barrister. He was Deputy Dean of Learning and Teaching and Director of EdTech which led to him purchasing and implementing Inspera.

He has a deep understanding of how pedagogy needs to be enabled by technology through experience before and while at Inspera.

With editorial contributions by Brent Mundy (Chief Product Officer) and Jessica Awtrey (Head of Americas).

Contact Us

If you're interested in learning more, book a demo with us.

Book at demo today: inspera.com/demo