Gallagher

AI Governance That Works for Higher Education

How institutions can manage risk, enable innovation and build trust across the AI lifecycle

Joey Sylvester, AIGP
Senior Vice President
Gallagher's Cyber Practice

Introduction

Artificial intelligence (AI) is rapidly reshaping higher education, offering opportunities to enhance student services, instructional support, research productivity and administrative operations. However, these benefits pose significant risks, ranging from privacy and civil rights concerns to academic integrity issues, model inaccuracies and reputational harm.

As AI becomes more complex and embedded in institutional processes, effective governance becomes essential. Universities must ensure responsible, transparent AI deployment that aligns with academic values, legal obligations and institutional strategy.

Key challenges include ensuring accountability for automated decisions, maintaining fairness and avoiding biased outcomes, protecting sensitive student data and preventing overreliance on AI tools by faculty or students. Generative AI (GenAI) introduces additional risks, including confabulations, concerns about information integrity and exposure to harmful or inappropriate content.

A structured AI governance framework supported by cross-functional expertise, consistent policies and lifecycle-based risk management enables institutions to adopt AI safely and effectively.

This paper outlines the key risks facing higher education, the essential components of a governance program and practical steps for embedding risk management throughout the AI lifecycle.

Improper use or deployment of AI can result in significant harm. According to a recent Gallagher study, 54% of businesses believe they understand the risks of AI "very well," while another 39% say they understand the risks "quite well." However, the NIST AI Risk Management Framework (RMF) and GenAI Risk Profile highlight several significant risks unique to, or exacerbated by, GenAI:[1, 2]

- CBRN (chemical, biological, radiological, nuclear) information or capabilities
- Confabulation (also known as hallucinations)
- Dangerous, violent or hateful content
- Data privacy
- Environmental impacts
- Harmful bias or homogenization
- Human-AI configuration
- Information integrity
- Information security
- Intellectual property
- Obscene, degrading and/or abusive content
- Value chain component integration

Moreover, higher education institutions face various AI risks that are unique to their environments:[3]

- Laws and regulations related to student privacy and fairness, such as FERPA or Title IX
- Academic integrity concerns and plagiarism
- Intellectual property (IP) infringement
- Overreliance on AI tools leading to loss of critical thinking and writing skills
- AI-literacy concerns
- Reputational damage resulting from AI misuse by students, faculty or staff

For higher education institutions, AI has been deployed in multiple ways, each offering unique opportunities. For example:

- AI chatbots for student engagement and queries
- Admissions tools for reviewing applicants at scale
- Research tools that support academic research
- Systems for grading and reviewing students' work, such as term papers, or for generating educational content, such as presentations
- General office/administrative use
- HR tools that assist with employee queries and job applicant screening/personnel selection
- AI-enhanced cybersecurity or public safety products

However, each of these use cases also presents potential risks to individuals, groups and the institution.

In admissions, AI can enable rapid, large-scale application reviews, freeing staff to focus on other responsibilities. However, biased or flawed outputs may raise Title IX or other discrimination concerns. The Family Educational Rights and Privacy Act (FERPA) may also become a factor in the use or development of an AI model if student records are used at any point in the AI system, and privacy is compromised.

Allowing the use of AI in the classroom can raise academic integrity concerns and may lead to overreliance on large language models (LLMs) for generating academic works. Additionally, falsely accusing students of generating work can cause a lack of trust and transparency in the academic process. In research settings, AI can support rapid, large-scale research. However, persistent hallucinations pose reputational risks if inaccurate information is incorporated without proper verification.

Student interactions with large language models, such as chatbots, can sometimes affect their mental health. A recent OpenAI report indicated that 15% of their weekly active users "in a given week have conversations that include explicit indicators of potential suicidal planning or intent."[3] This highlights the need for additional safeguards and oversight.[4] Given the widespread use of AI among students, representation from student life on an AI-governance committee may be warranted.

Universities are complex institutions, often characterized by decentralized decision-making, academic freedom and flexible instructional policies.[4] Many universities impose strict rules regarding the use of AI by students, staff and faculty, while others adopt more open approaches to how these models can be used.[5] Moreover, student use of AI in coursework may be encouraged or restricted depending on the course or instructor, underscoring the lack of a one size fits all governance approach.

While a particular university may allow a professor to decide how students may or may not use AI in the classroom, a different university, operating under a different set of rules, could implement a strict campus-wide approach that prohibits students from using AI and limits faculty and staff to specific use cases. The permissibility of AI use on campus varies widely from institution to institution. Notably, existing university policies can be adapted for individual institutional use.

A recent Gallagher survey of university risk managers highlighted several key questions and concerns:

1. What liability exposures should we consider for our degree and incubator programs?
2. How can we best leverage AI while maintaining the provision of verifiable information?
3. How do we manage overreliance on AI?
4. How should we address the risk of false accusations related to AI use?
5. We exhibit a lack of awareness of AI standards and acceptable use.
6. There may be compliance issues related to FERPA or other laws when student data becomes part of these models.

In virtually all cases, effective governance can establish clear guidelines and address many of these concerns.

Institutions may adopt centralized or decentralized governance models. While centralized models offer consistency and oversight, decentralized models allow flexibility across departments. Successful governance frameworks often include cross-functional teams, clear policies and regular audits.

Governance Models

Governing AI?

AI Governance goes hand in hand with institutional strategy and objectives. While our research shows that risk management frameworks, ethical-impact assessments and written incident response plans are among the least adopted measures, they remain essential for effective AI risk management. Effective governance framework implementations can assist a university with meeting overall objectives while remaining compliant with increasing legal and regulatory requirements regarding the use of AI.

Establishing a dedicated **AI Governance Committee** and a formal Responsible AI function enables the university to proactively manage AI-related risks. Centralizing oversight enables consistent policies, clear accountability and rapid response to emerging issues across teaching, research, student services and operations. It also embeds ethical principles, transparency and explainability into AI selection, development and deployment, supporting trust with students, faculty, institutional leadership and external partners while aligning with evolving laws and accreditation standards.

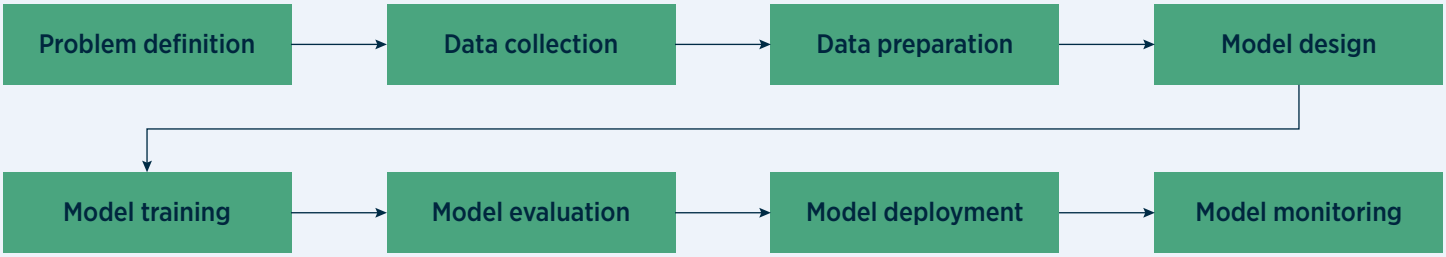
To be effective, the committee should be cross-functional and represent the university's full risk profile. Membership can include: legal counsel and compliance, risk management and internal audit, information security and data privacy. Additional considerations for membership can include faculty from computer/data science as well as social sciences/humanities with ethics expertise, student affairs and academic integrity leaders, accessibility specialists, IT, enterprise architecture, and data governance, procurement and vendor management, HR (for workforce uses of AI) and labor relations, communications and public affairs, and, where applicable, representatives from healthcare/medical centers and international programs. This mix ensures risk owners are fully engaged and relevant policies are practical, equitable, secure and aligned with the university's mission, strategic priorities, academic values and community impact.

The AI Governance Committee defines the principles and guardrails that ensure AI initiatives deliver academic and operational value while protecting privacy, equity and academic integrity. Through systematic AI use case reviews, the committee evaluates proposed applications for fit, feasibility and ethical considerations, supported by structured AI impact assessments to identify and mitigate potential harms to students, faculty, staff and the broader community. Complementing this, AI readiness assessments gauge the university's technical, data and cultural capacity to deploy solutions safely and effectively, guiding investment in infrastructure, talent and change management.

In parallel, the committee provides risk oversight across the **AI lifecycle,** implementing a consistent framework for governance, controls and accountability that integrates with existing risk, compliance and data governance structures. It leads policy development covering procurement, model selection, data usage, transparency, human oversight, accessibility, academic integrity and establishes AI incident response plans to rapidly address misuse, performance failures or security/privacy events. Ongoing monitoring and reporting illuminate AI performance, fairness and compliance metrics, fostering continuous improvement and informed decision-making at the executive level. Finally, the committee drives education and awareness across the campus community, equipping stakeholders with practical guidance, training and resources to use AI responsibly and confidently in teaching, research and operations.

Incorporate risk management into the AI lifecycle

Palo Alto Networks describes the AI development lifecycle as a continuous process that spans problem definition, data handling, model design and evaluation, deployment and ongoing monitoring.[8] For institutions of higher education, this lifecycle provides a practical structure for applying governance and risk management in a consistent and operational way. Even when universities procure AI capabilities rather than develop models internally, they still participate in the lifecycle through use-case selection, data decisions, system configuration, policy enforcement and post-deployment oversight.



The lifecycle begins with **problem definition**, where institutions determine whether AI is appropriate for a given use and how its outputs will be applied. Governance at this stage establishes accountability by clarifying the intended purpose, defining success criteria, identifying affected stakeholders and assigning ownership. Early risk considerations include student privacy, fairness in decision-making, academic integrity and reputational exposure resulting from inaccurate or misleading outputs. Weakly defined use cases create risk that persists throughout the lifecycle, regardless of model quality.

As the lifecycle progresses into **data collection** and **data preparation**, governance becomes particularly important due to the sensitivity of institutional data. Universities manage large volumes of student records, research data, employee information and other regulated datasets. Controls at this stage should address data ownership, access restrictions, retention requirements and documentation of data lineage. Risk considerations include FERPA compliance, unintended secondary use of data, re-identification risks when datasets are combined and bias introduced through incomplete or unrepresentative data. Decisions made during data preparation, including labeling and filtering, can materially affect fairness and reliability and should be treated as formal governance checkpoints.

During **model design, training and evaluation**, governance should focus on transparency, oversight and documentation of limitations.[9] Higher-impact use cases such as admissions support, grading assistance, student conduct analysis or employment decisions require stronger review and explainability standards than lower-risk productivity tools. Evaluation should assess not only accuracy but also fairness, robustness and predictable failure modes, including the risk of confabulation in instructional or research settings. This phase is also where AI-specific incident response planning should be established, as traditional cybersecurity response processes may not fully address harms related to bias, misinformation or inappropriate outputs. Read more here: Risk Bulletin: AI Incident Response Plans — Proactive Strategies for Emerging Threats.

AI model cards, which are provided by model developers, can be provided which give crucial information on the development of the model, testing and so forth.[10] They can serve as a practical tool to meet transparency requirements of the institution or regulations like the EU AI Act.

Many commercial models also have critical dependencies within what NIST refers to as the Value Chain. It is one of the key risks highlighted in AI 600-1, the GenAI Risk profile, which reads as follows:

"**Value chain and component integration:** Non-transparent or untraceable integration of upstream third-party components, including data that has been improperly obtained or not processed and cleaned due to increased automation from GAI; improper supplier vetting across the AI lifecycle; or other issues that diminish transparency or accountability for downstream users."

In short, many AI models rely on critical third-party dependencies and components, not all of which are well known, understood, or disclosed on AI model cards. As such, calls for the so-called **Artificial Intelligence Bill of Materials (AIBOM)** are rising in order to provide greater transparency in the value chain. Obtaining such information will lend greater oversight into AI supply chain risk.[7, 10]

Deployment marks the transition from controlled testing to operational use. Prior to deployment, institutions should confirm that both the AI system and the organization are prepared. This includes validating alignment with institutional policies, defining acceptable and prohibited uses, establishing human oversight expectations, and ensuring audit logging and access controls are in place. Given the decentralized nature of many campuses, deployment governance also supports consistency across departments while allowing for appropriate academic flexibility.

Before deploying, an **AI Impact Assessment** should be performed which gathers details on intended use and limitations, applicable laws, key stakeholders that may be impacted by AI, the identification of potential harms, alignment of security, privacy, and transparency to internal policies, and accountability measures.[12]

Read more about AI Impact Assessments here:
AI Impact Assessments.

An AI Readiness Assessment should also be undertaken. This can accomplish two main goals:

- Determine readiness of the AI model itself before deployment. Is the AI actually ready? Is it performing as expected? Is the output in line with the baselines established?
- Is our organization ready for this specific use? Are our staff and faculty ready and have they been made aware of how to use the model effectively? Are they aware of information that can be used in the model and what cannot? Have we effectively intended and prohibited use? What additional steps do we need to take to ensure this model will be used successfully?

The lifecycle continues through ongoing **monitoring and maintenance**, where governance becomes an ongoing responsibility rather than a one-time approval. Post-deployment monitoring allows institutions to detect performance degradation, model drift, rising hallucination rates, misuse patterns or indicators of compromise. Monitoring should be conducted against established baselines and supported by clear escalation paths and the ability to suspend or disable AI capabilities when risk exceeds acceptable thresholds. Periodic audits and policy reviews remain necessary as legal requirements, institutional expectations and AI technologies evolve.

Taken together, the AI development lifecycle offers institutions of higher education a practical framework for embedding governance and risk management into routine AI decision-making. Aligning controls and accountability with each phase of the lifecycle supports responsible adoption while protecting students, faculty, staff and institutional reputation.

Summary

AI incidents occur with some regularity, with public databases documenting incidents and related litigation in detail.[6] These incidents range from allegations of discrimination and toxicity to misinformation, and can sometimes lead to unwanted consequences.[7] Improper AI use can raise privacy and safety concerns and adversely affect institutional reputation.

Establishing effective governance teams and governance frameworks early in the AI-adoption process can help institutions manage this rapidly evolving risk.

Contact Gallagher for more information.

About the Author:

Joey Sylvester, AIGP, is an area senior vice president of Gallagher's Cyber practice with a focus on institutions of higher education and public entities. He is an artificial intelligence governance professional.

Sources

[1]"AI Risk Management Framework," *NIST*, accessed 1 Apr 2026.

[2]"Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile," *NIST*, Jul 2024. PDF file.

[3]"Strengthening ChatGPT's responses in sensitive conversations," *OpenAi.*

[4]Duffy, Clare. "Parents of 16-year-old Sue OpenAI, Claiming ChatGPT Advised on His Suicide," *CNN Business*, 27 Aug 2025.

[5]"The AI Wake-Up Call for Universities: Key Insights from the 2025 EDUCAUSE Survey," *University of Massachusetts Amherst*, 28 Apr 2025.

[6]"Welcome to the AI Incident Database," *AI Incident Database*, accessed 1 Apr 2026.

[7]"How is AI harming us?" *MIT AI Risk Initiative*, accessed 1 Apr 2026.

[8]"Making AI Systems Transparent, Auditable, and Secure," *OWASP AIBOM*, accessed 1 Apr 2026.

[9]"Model Cards," *Google.*

[10]"Welcome to the AI Incident Database," *AI Incident Database*, accessed 1 Apr 2026.