INSTITUTE FOR ETHICS AND EMERGING TECHNOLOGIES

UMASS BOSTON APPLIED ETHICS CENTER

AI Agents in Higher Education

2025

WHITE PAPER ON AI AGENTS IN HIGHER EDUCATION

AUTHORS
Jake Burley, UMass Boston Applied Ethics Center

PUBLISHED BY
Institute for Ethics and Emerging Technologies
info@ieet.org
56 Daleville School Rd
Willington CT 06279 USA

GRAPHIC DESIGN
Steven Umbrello, IEET

IEET 2025

ISBN 979-8-9879599-4-7

ABOUT THE IEET & AEC

This white paper has been drafted by the Institute for Ethics and Emerging Technologies in cooperation with the Applied Ethics Center at UMass Boston

The Institute for Ethics and Emerging Technologies is a nonprofit think tank which promotes ideas about how technological progress can increase freedom, happiness, and human flourishing in democratic societies. We believe that technological progress can be a catalyst for positive human development so long as we ensure that technologies are safe and equitably distributed. We call this a "technoprogressive" orientation.

The Applied Ethics Center promotes research, teaching, and awareness of ethics in public life. Our current projects are concerned with the ethics of emerging technologies.

INSTITUTE FOR ETHICS AND EMERGING TECHNOLOGIES

Applied Ethics Centre

Table of Contents

CONTENTS

01. EXECUTIVE SUMMARY

Artificial intelligence (AI) systems are no longer speculative tools of the future; they are already shaping the operations and experiences of higher education institutions. From automated course registration to personalized learning platforms and AI-generated study guides, increasingly autonomous systems are being deployed across both the administrative and academic domains of university life.

This white paper outlines a framework for understanding how higher education is integrating AI. It examines the ethical and social questions that emerge as these systems grow more autonomous and visible to students and faculty.
We distinguish between three broad categories of AI integration:

- Non-autonomous systems, such as those used for purchasing, registration, and scheduling, are utilized in various contexts. These systems operate in the background and are largely task-specific. They utilize AI technologies to process large datasets, identify patterns, and provide insights, while human agents establish their goals.

- Hybrid systems encompass a range of tools, including AI-assisted tutoring, personalized feedback tools, automated writing support, and chatbots. These systems integrate human and machine functions, interacting directly with students, instructors, and administrators. They often rely on generative AI technologies (especially LLMs). While human agents set their overall goals, the intermediate steps to achieve those goals are often not specified.

- Autonomous agents, still emerging, have the potential to act as research collaborators, teaching assistants, or even principal investigators in scientific settings.

It is essential to note that these categories—non-autonomous, hybrid, and autonomous—are intended to serve as useful points along a continuum, rather than distinct stages that neatly categorize specific technologies. As we will see, many technologies appear to straddle the line between hybrid and autonomous, or exhibit both non-autonomous and hybrid characteristics. As a result, we are more interested in considering the kinds of ethical concerns that emerge as systems become increasingly autonomous, rather than in classifying systems into rigid categories.

We emphasize that ethical and social concerns tend to intensify as systems become more autonomous and more front-facing. While back-end systems may raise questions about transparency and control, front-facing agents implicate deeper concerns about academic integrity, mentorship, intellectual labor, and the cultivation of critical thinking. At the same time, AI systems may offer real benefits, such as reducing administrative burdens, expanding access to learning support, and accelerating research.

Our goal is not to advocate for or against the use of AI in higher education, but to offer a structured and fair-minded account of what is at stake. We urge academic leaders, educators, and policymakers to consider both the promise and perils of AI in the university context—and to make informed choices that reflect the full complexity of this technological transformation.

02. DEFINITIONS

Before analyzing the ethical and institutional implications of AI in higher education, it is important to clarify what we mean by *artificial intelligence*, *agency*, and *autonomy*. These terms are often used imprecisely, but they carry significant weight in both technical and public discourse.

Artificial Intelligence

We use the term "artificial intelligence" (AI) broadly to refer to computational systems capable of performing tasks that would typically require human intelligence. It includes, but is not limited to, pattern recognition, language generation, evaluation, and decision-making. While AI encompasses a wide range of techniques—from simple rule-based systems to advanced machine learning—our analysis focuses on AI systems that are probabilistic and supported by large data sets, such as LLMs.

Agency

The term “agency” can refer to several divergent concepts across distinct intellectual disciplines and traditions. We distinguish between two broad understandings of agency relevant to this paper. We term the first intentional agency and the second non-intentional agency.

By intentional agency, we simply mean what one might consider a “common-sense” notion of agency. We all understand the experience of intending to do something or of deliberating about a decision. This kind of agency underwrites normative responsibilities and treats subjects as entities capable of forming intentions, reasoning about actions, and taking ownership of outcomes.

In contrast, we understand non-intentional agency as capturing a sense of agency that doesn’t require conscious deliberation or phenomenal properties. For example, when we speak of institutions, firms, or nations acting, we don’t view these entities as subjects with intentions. By non-intentional agency, we simply mean the capacity of entities—whether human, artificial, or institutional—to produce effects within a system.

Literature in both philosophy and psychology provides us with reason to doubt that all actions performed by agents should be considered intentional in the rational, deliberative manner outlined above. Classic experiments in psychology demonstrate that subjects confidently offer reasons for choices whose real determinants they cannot access—for example, preferring a stocking because of its position rather than any intrinsic feature (Nisbett & Wilson 1977), or justifying a decision about a face they did not in fact choose (Johansson et al. 2008). In philosophy, some have argued that the traditional intentional accounts (e.g., Davidson 1980, Anscombe 1957) are insufficient to account for automatic or non-deliberative actions, with some (e.g., Burge 2009, Steward 2009, Okasha 2018) seeking to characterise agency in ways that could apply to non-human animals. Concurrently, Floridi (2025) proposes that agency itself need not depend on intelligence, intentionality, or mental states, arguing instead that artificial systems exemplify a novel form of “agency without intelligence.”

The distinction between intentional and non-intentional agency is not meant to provide a comprehensive theory of agency. Rather, it offers a straightforward way of conceptualizing artificial forms of agency without relying on assumptions about minds or consciousness. When we describe AI and other emerging technologies as "agents," the claim is not that they possess the phenomenal or subjective qualities of human agents, but that they exhibit patterns of interaction, autonomy, and adaptability that allow them to bring about outcomes in the world. In this sense, their agency lies in the capacity to initiate and shape processes, even in the absence of understanding or intention.

Autonomy

Similarly, the term autonomy is used in multiple, often imprecise ways, frequently carrying different implications depending on the context. In much of the philosophical literature, particularly in ethics and moral psychology, autonomy is closely linked to self-governance, rational deliberation, and the ability to choose in accordance with one's values or reasons. On this view, to be autonomous is to be free in a robust, normatively significant sense—capable of moral responsibility, deliberative choice, and free will. We might refer to this as moral or deliberative autonomy, and it is closely linked to the conception of intentional agency described above.

In contrast, when we speak of an autonomous system in the context of artificial intelligence —such as an autonomous vehicle or an autonomous agent—we typically mean something quite different. Here, autonomy refers not to moral or rational self-governance but to the capacity to operate without moment-to-moment human intervention. It is a functional notion, concerned with the system's ability to carry out tasks, adapt to inputs, and respond to its environment in a semi-independent way. An autonomous agent in this sense is not essentially self-aware or morally responsible, but rather procedurally independent: capable of initiating or completing actions based on preprogrammed goals, probabilistic models, or real-time data.

In this paper, we adopt the latter conception. Our concern is not with whether AI systems are autonomous in the moral or philosophical sense, but instead with the functional autonomy they exhibit within educational institutions. That is, we are interested in how these systems can act, interact, and affect outcomes in higher education without direct or continuous human control, and what ethical and social questions arise as their autonomy increases.

03. NON-AUTONOMOUS SYSTEMS

While much of the public conversation around AI in education focuses on high-visibility applications such as chatbots, tutoring platforms, or generative tools, a quieter transformation is already well underway. Higher education institutions have already begun integrating non-autonomous AI systems into their day-to-day operations. These systems typically do not make independent decisions; however, they perform complex, often large-scale tasks that would have previously required significant human labor or would not have been possible at all.

Importantly, broader structural forces shape the adoption of non-autonomous AI in higher education. Like other large-scale institutions, universities face mounting pressures to streamline operations, reduce administrative costs, and make data-driven decisions. In fact, a 2024 survey found that 80% of higher-ed administrators are motivated to adopt AI for improved efficiency, and AI adoption in higher education has more than doubled in a single year (AI in Higher Education: Understanding the Present and Shaping the Future, 2024). These pressures help explain why even relatively modest AI systems—ones that do not teach, advise, or interact directly with students—are being adopted at an accelerating pace.

Examples include algorithmic systems used in admissions review, purchasing, academic advising, and institutional risk assessment. Some universities have implemented early alert systems that flag students at risk of academic failure based on attendance, grades, or behavioral indicators. A 2012 survey revealed that 93% of four-year institutions had such "early alert" systems in place, and recent studies have linked these alerts to improved grades and retention (Alonso, 2025). Others have implemented AI in enrollment management, financial aid processing, and record-keeping processes (Khairullah et al., 2025). Research indicates that AI can significantly streamline these administrative tasks by automating processes like admissions and records management (ibid). These technologies operate largely in the background, functioning as digital infrastructure that supports the institution.

Ethical Concerns

Although these systems may not meet the threshold of autonomy as defined in the previous section, they are far from trivial. Many are built on sophisticated machine learning models or rule-based systems that process vast amounts of institutional data. Their growing presence reflects not only technical possibility but also an institutional logic oriented toward efficiency, standardization, and predictive oversight.

However, while non-autonomous systems may be powerful, the ethical concerns they present are largely well-understood within current computer and AI ethics frameworks. Concerns about data security, privacy, surveillance capitalist logics, and the like are all relevant when considering how non-autonomous systems are implemented and managed. Furthermore, AI-specific concerns, such as algorithmic biases, hallucinations, explainability, and reliability, are also relevant to these systems. This is not to dismiss these concerns as trivial—they are not—but rather to acknowledge that existing ethical frameworks are well-equipped to guide the implementation of these technologies in higher education institutions [1].

04. HYBRID SYSTEMS

Hybrid AI systems occupy a growing and increasingly visible role within higher education. Unlike non-autonomous systems, which operate behind the scenes, often hybrid systems engage directly with students, instructors, and researchers. They combine AI tools with human oversight, delivering personalized feedback, instructional support, or adaptive services in real time. These tools are not fully autonomous, but they are already having a meaningful impact on higher education.

At the course level, hybrid systems are already seeing use in a wide range of tasks, both for instructors and students. For example, writing assistants like Grammarly and Turnitin Draft Coach offer real-time feedback on student writing. AI-powered tutors for university students are also seeing positive early results, and commercial platforms like Macmillan Learning now offer AI tutors for STEM subjects (Kestin et al., 2025; Mowreader, 2025). Individual universities such as MIT and Harvard are beginning to roll out AI-powered tutors for specific courses (Feijo, 2025; Manning, 2024). Meanwhile, AI-supported tools like Cognii and Century Tech are designed to assist instructors in building syllabi, generating assignments, and more. In addition to these specialized tools, more general-purpose LLMs, such as ChatGPT, Google's Gemini, and Anthropic's Claude, offer enormous flexibility, capable of acting as a tutor, writing coach, or assistant as needed.

Researchers are also utilizing hybrid systems to summarize existing literature, present study results, and suggest future experiments. Tools like Elicit, Semantic Scholar, and Connected Papers assist in literature review and citation tracing. AI tools, such as GPTs, can also help summarize academic texts, identify relevant datasets, or generate preliminary research designs. These tools do not replace the researcher but function as analytical collaborators, offering leads, suggestions, or structures that accelerate the early stages of scientific and scholarly work.

At the institutional level, hybrid AI tools support tasks such as scheduling, student retention, and academic advising. Chatbots like Georgia State's "Pounce" are being deployed in hybrid workflows where staff can step in when the AI encounters queries it cannot resolve (Inman, 2022). In practice, human monitors oversee these bots and take over complex questions, as seen with other university chatbots that are monitored by staff and hand off unanswered queries to human experts. These systems complement rather than replace professional staff, but they are beginning to change how universities manage scale and allocate human attention.

Ethical Concerns

When considering non-autonomous systems, many of the central ethical issues are serious but relatively well understood. Frameworks in computer and data ethics already help us think about problems of privacy, security, and bias. As systems become more autonomous, however, they take on a different character of risk. With tools like AI tutors, research assistants, and automated evaluators, we begin to encounter a different set of questions, such as:

- **Transparency** — Natural language systems already show how hard it can be to tell when you are interacting with a non-human agent, and the challenge will only grow as these tools become more sophisticated. For students, teachers, and administrators, clarity about who—or what—they are engaging with is essential. Without it, confusion and mistrust may foster a sense of alienation—both from the university deploying the technology and from the broader educational project itself. Transparency, therefore, matters not only for trust but also for understanding who is grading work, making decisions, or offering guidance.

- **Accountability and intellectual credit** — Much of the discourse around AI in education has focused on student plagiarism and academic integrity. However, hybrid systems raise broader questions: when evaluation or feedback is a shared effort between a human and an artificial agent, who should be credited—or held accountable—for the result? In research, as AI contributes more directly to literature reviews, experimental design, or analysis, how should we understand the authorship of research products? Moreover, when errors or disagreements occur, where does responsibility ultimately lie?

- **Cognitive offloading** — For students, early evidence suggests that frequent reliance on AI can lead to reduced critical engagement and problem-solving. Studies in 2025 indicate that over-reliance on AI tools may lead students to offload cognitive effort and engage less critically with the content (Favero et al., 2025). However, the concern extends beyond the classroom. Teachers may lose valuable opportunities for granular engagement with student work if they routinely delegate feedback to AI. Researchers may lose practice in basic scholarly skills, such as literature reviews, data handling, or experimental design, even when advanced tools are available. Administrators, too, may come to depend on AI-driven analytics in ways that discourage deeper institutional judgment. Across all these roles, the risk is not simply the loss of work, but the erosion of the skills and habits that come from doing the work.

While the ethical concerns of non-autonomous systems align well with existing computer and data ethics frameworks, the issues raised here have a distinct character. They are less about technical compliance and more about relationships, ascription, and cultivating personal and professional skills.

05. AUTONOMOUS SYSTEMS

While most AI systems in higher education today fall into the non-autonomous or hybrid categories, increasingly, fully (or near-fully) autonomous agents have captured the imagination of researchers, developers, and industry. It is important to be clear at the outset: we do not yet have strong examples of truly autonomous systems in this space. What exists are small-scale cases and early-stage prototypes that hint at what might be possible. The appeal of autonomy is straightforward—systems that could operate independently promise to increase productivity, free up time, and handle tasks that now require significant human labor. However, at present, autonomous, intelligent systems in higher education remain more of an aspiration than a reality.

On the institutional side, universities are seeking ways to utilize AI to enhance decision-making, streamline operations, and glean insights from budgets and data. The longer-term goal is to develop systems that can manage parts of the bureaucracy with minimal human involvement. While current tools fall short of that goal, there are, nevertheless, instructive early examples. Georgia State University's Pounce chatbot can automatically answer the vast majority of student enrollment and financial aid questions, with only a small number requiring escalation to staff. Washington University's TRACE-cs scheduling system can generate course timetables and explain its reasoning to students. These examples illustrate what partial autonomy looks like, but they remain tightly bound in scope. A system capable of handling complex and policy-sensitive decisions independently remains a projected, not an actual, development.

In teaching and learning, we see a similar picture. Systems like Georgia Tech's Jill Watson or Macmillan Learning's AI tutor give students responsive feedback and answers. However, these still operate more like semi-autonomous tools: they engage only when prompted and remain limited to specific domains. MIT's OpenLearning platform offers another promising example, with selected courses featuring AI tutors that assist students with problem sets and questions, without providing them with the correct answers (Feijo, 2025). Khan Academy's Khanmigo demonstrates a more ambitious form of conversational tutoring, yet it, too, is best understood as a preliminary step. The larger vision in this area is of a fully autonomous personal tutor—an AI capable of working continuously with a student across subjects and domains. At a minimum, autonomous systems would be like a "grad student in a box," taking on tasks such as grading, routine communication, and other support functions. In principle, such systems could extend to course design, syllabus preparation, and even assessment strategies, offering comprehensive instructional support. At the limit, one can imagine systems that handle nearly all routine pedagogical tasks, leaving instructors to focus on mentorship, advanced content, and shaping the broader learning experience.

Research is the area where autonomous systems have arguably advanced furthest. Examples of robotic laboratories, such as Lawrence Berkeley National Lab's A-Lab and the University of Liverpool's robot chemist, already automate large portions of the experimental cycle, running continuously and selecting new tests based on prior results. Initiatives like Sakana AI and the Chan Zuckerberg Initiative's virtual lab push in the same direction. Here, it is helpful to distinguish two different levels of ambition.

The first is an "AI grad student" model, where a human investigator sets a problem and the AI designs and conducts the necessary experiments. The second is an "AI Principal Investigator" model, in which the system itself identifies knowledge gaps, formulates research questions, and carries projects through to results and publication. The latter remains largely speculative, but it captures the trajectory many envision: AI systems that not only support research but also carry out the process of discovery itself.

Ethical Concerns

In the preceding sections, we traced how ethical concerns shift as systems become more sophisticated. Considering non-autonomous systems, the key issues (i.e., privacy, data security, bias) are serious but broadly understood within existing computer and data ethics frameworks. Hybrid systems, we claimed, raise additional concerns about transparency, accountability, and cognitive offloading—concerns that tend to surface at the level of individual experience for students, teachers, and researchers.

With autonomous systems, however, the risks take on a different character. These tools have the potential not only to automate individual experiences of teaching or advising but to restructure pedagogy, research, and even the organizational life of universities themselves. At this level, the ethical questions are institutional and social: how automation may contribute to deprofessionalization, disrupt apprenticeship pathways, or reorganize higher education around logics of efficiency and scale. The following sections highlight three such concerns: opportunity costs and downstream effects, rationalization and de-skilling, and stratification and platform decay.

Opportunity Costs and Downstream Effects

As higher education becomes increasingly "automatable" with the development and implementation of more intelligent and autonomous technologies, we should be attentive to the potential opportunity costs and disruptions to our knowledge-producing ecosystems. Universities are not simply job training programs—they are hubs of knowledge transfer, apprenticeship, and discovery. Their value lies not only in the content of what is taught but also in the practices, relationships, and tacit forms of expertise that emerge when people learn and work alongside one another.

As AI systems assume more responsibilities across the university, we may see a subtle reorganization of who is responsible for what. Faculty may still "teach," but more and more of the day-to-day work, such as responding to students, reviewing assignments, or structuring a syllabus, can be handed off to systems optimized for efficiency and scale. Principal investigators may still lead research, but automated agents can handle the concrete tasks that sustain inquiry.

Graduate students and early-career academics, who have historically relied on these tasks as training grounds, may find themselves with fewer opportunities to learn how to teach, mentor, or think pedagogically. The epistemic and network costs of disrupting this pipeline also deserve attention. Roles such as internships, research assistantships, and teaching assistantships do more than provide labor; they generate mentorship and network effects between experts and those at earlier stages of their careers. These close-proximity arrangements create channels of tacit knowledge transfer. Much of the history of science and scholarship suggests that discoveries and intellectual traditions cluster precisely because expertise localizes in particular labs, departments, or schools, and because students and mentors embed themselves in sustained, situated practice.

In addition to the social opportunities outlined above, undergraduate students may face a more subtle opportunity cost. Now equipped with tools that can draft, revise, or explain just about anything, they may not be asked, or may not ask themselves, to engage with the difficult, formative parts of learning. Struggle is often essential, not incidental, to the process of skill acquisition [2].

Struggle is often essential, not incidental, to the process of skill acquisition. Research on the testing effect demonstrates that effortful retrieval (i.e., being tested) yields significantly better long-term retention than additional study, even when no feedback is provided (Roediger et al. 2006). Likewise, in writing, research indicates that students who iteratively self-correct achieve substantially higher-quality drafts than those who rely on first-pass fluency; the cognitive effort of diagnosing and fixing one's own prose is a primary driver of growth rather than a dispensable add-on (Graham & Perin, 2007).

The pressures on students to perform well are enormous, and assessment in higher education remains primarily oriented around outputs, such as grades, test scores, and other measurable benchmarks, rather than the process of "learning to learn." As long as education systems define success in terms of these outputs, students will face strong incentives to offload the difficult, less visible work of learning to AI systems and focus instead on meeting the standards that determine their outcomes. This dynamic risks reinforcing a cycle in which efficiency and short-term performance overshadow the slower, less tangible, yet ultimately more transformative aspects of education.

Taken together, these developments suggest that the greatest risk posed by automation in higher education is not simply the loss of particular tasks, but the erosion of the ecosystem of practice that has long sustained teaching, research, and learning. When we reduce opportunities for apprenticeship, mentorship, and productive struggle, the downstream effects may be profound: fewer scholars and students equipped with the tacit expertise, network connections, and habits of "learning to learn" that emerge only through active participation in these processes. As emerging technologies promise to make higher education more efficient, we should be equally attentive to the opportunity costs that efficiency may conceal.

Rationalization and De-skilling

The growing role of intelligent tutoring systems and other AI tools in education is not only a technical development but also the continuation of long-running trends in the rationalization of professional work. As James Hughes argues, teaching—like medicine, law, and other professions—has historically claimed a kind of autonomy rooted in expertise, judgment, and personal relationships. However, pressures for efficiency, standardization, and cost reduction have steadily eroded that autonomy, setting the stage for automation (Hughes, 2021).

A critical element of this trajectory is curricular standardization. For AI tutoring systems to compete with or replace components of human teaching, developers must systematize knowledge into standardized competencies and assessments. This process has advanced furthest in STEM fields and in K–12 education, where national curricula and standardized testing are the norm. In higher education, outcomes assessment, accreditation, and competency-based models have similarly encouraged the unbundling of faculty work into discrete, rationalized tasks that can be reassigned or automated.

The result is a form of deprofessionalization. As tasks such as grading, curricular design, and assessment are standardized and digitized, they are increasingly shifted to contingent faculty, paraprofessionals, or, in some cases, software. The decline of tenure and the rise of adjunctification accelerate this process, weakening the professional authority of faculty and making universities more receptive to automated tools. Intelligent tutoring systems, auto-graders, predictive analytics, and AI-based advising are not appearing in a vacuum—they are being adopted by institutions already primed to reduce costs and subdivide the work of teaching.

For students, the appeal of these systems lies in personalization, efficiency, and flexibility. For institutions, the appeal lies in scale and cost savings. However, for teachers, the danger is the erosion of the very skills that once defined their role. When instruction is increasingly mediated by software, what remains of the core craft of teaching—shaping curricula, interpreting student needs, and building intellectual relationships—risks being hollowed out. Rationalization and de-skilling thus frame not just a technical shift, but a profound reorganization of the academic profession, one in which AI is positioned less as a tool and more as a substitute for practices once thought to be irreducibly human.

Moreover, because the systems that are taking on these roles are optimized for measurable outcomes, there is a risk of reshaping education itself to fit what can be tracked, scaled, and delivered. This is not to dismiss the value of evidence-based assessment; empirical measures of learning and performance can indeed play a crucial role in informing decisions.

Increased Stratification of Higher Education

Another risk posed by the expansion of AI in higher education is increased stratification between institutions. As automation advances, less well-resourced universities will likely rely more heavily on AI to deliver instruction and student services. In contrast, elite universities emphasize human-intensive elements, such as mentorship, seminars, and research collaboration, as a premium product.

At the same time, it would be a mistake to treat human instruction as inherently more valuable in every context. In some domains, automation is the preferred or even premium option. Consider commercial aviation: few passengers would consider a pilot manually flying the plane from start to finish as a boutique luxury. The reassurance comes from robust autopilot systems that increase safety and reliability. In higher education, the same logic can apply. Automation may provide greater consistency, responsiveness, and efficiency than human-led alternatives. For example, automated grading systems that deliver results more quickly and uniformly than fatigued graders or chatbots that resolve thousands of student queries instantly.

What elite institutions are likely to preserve are not ineffable qualities of "the human touch," but concrete relational goods: mentorship, networking opportunities, and the social and intellectual capital that comes from being taught and advised by particular individuals. These forms of value are difficult to replicate in automated systems and remain deeply tied to institutional prestige and community. Elite universities will almost certainly adopt AI technologies, but they will also continue to market these enduring advantages. The risk, then, is not a simple binary between automated mass education and boutique human instruction, but a more complex differentiation where institutions adopt automation in uneven ways, reinforcing existing hierarchies rather than leveling them.

A further complication is the risk of platform decay, or what some commentators have called "enshitification" (Doctorow, 2023). As tech platforms grow, they often follow a familiar trajectory: early iterations are optimized to attract users, but over time, incentives shift toward cost-cutting, monetization, and efficiency. The result can be a decline in quality and trust in the tools themselves. For less-resourced institutions that heavily rely on third-party AI platforms, this decay could exacerbate existing inequities. Students at these schools may not only receive more automated instruction, but instruction that itself worsens over time as platforms prioritize profit over pedagogy. In this light, stratification is not just about who uses AI and who relies on human-intensive practices, but also about how unevenly platform decay effects are distributed across the higher education landscape.

06. CONCLUSION

Throughout this paper, we have emphasized that the integration of increasingly autonomous AI systems into higher education raises not only technical and operational challenges but also more profound questions about the institutional fabric of universities. The concerns outlined above—disruptions to opportunity structures, the rationalization of pedagogy, and the potential for new forms of stratification—are not discrete issues, but interconnected signals of a broader tension.

At the center of this tension is a question that higher education has long skirted: what do we expect universities to do? For some, the primary function of higher education is to equip students with marketable skills, preparing them for a changing labor market. For others, the university is a site of personal formation and civic cultivation, a place where the value lies as much in mentorship, community, and “learning to learn” as in the production of credentials. These functions are not always compatible, and automation forces us to face their divergence more starkly.

In short, automation does not simply accelerate existing practices in higher education; it compels us to clarify what higher education is for. Is the university primarily a training ground for the labor market, or a space for broader human and civic development? Should its value be measured in scalable outcomes, or in the harder-to-quantify processes of mentorship, community, and discovery? These are not speculative questions, but practical ones that cut to the heart of how we design, regulate, and adopt new technologies.

As we look toward a future where knowledge work itself becomes increasingly automated, the guiding question we must ask is: What is the purpose of higher education in such a world? The answer to this question will shape not only how we evaluate the risks and opportunities of AI, but also how we preserve the most essential, and perhaps irreplaceable, functions of the university.

07. RECOMMENDATIONS

01

Recognize the Risks of Increasingly Autonomous Systems

Universities should distinguish between non-autonomous, hybrid, and autonomous systems in their governance frameworks. Non-autonomous tools (e.g., admissions analytics) can be managed within existing frameworks for privacy and bias. In contrast, hybrid and autonomous systems require heightened oversight due to their direct impact on student learning, trust, and mentorship.

02

Preserve Struggle and Apprenticeship in Learning

Institutions should resist the temptation to automate formative learning tasks fully. AI can scaffold learning, but students must continue to engage in productive struggle—drafting, revising, problem-solving—that underpins genuine skill acquisition. Faculty and graduate student roles in mentorship and apprenticeship should be safeguarded, ensuring that automation does not erode pathways of tacit knowledge transfer.

03

Ensure Transparency in Hybrid Interactions

Universities should adopt clear disclosure policies so that students, faculty, and staff are aware when they are interacting with an AI agent. Transparency is essential to avoid alienation and mistrust, particularly as hybrid systems blur the line between human and machine roles.

04

Develop Accountability Protocols for Shared Work

Institutions should clarify authorship and responsibility when AI systems contribute to grading, research, or advising. Clear policies should assign accountability for errors or misconduct, preventing ambiguity about whether humans or AI systems are ultimately responsible.

05

Safeguard Against Rationalization and De-skilling

Faculty development programs should emphasize the irreplaceable elements of teaching—judgment, interpretation, mentorship—even as routine tasks become automated. Universities should view automation as a complement to current teaching strategies rather than a substitution, thereby preserving the craft of teaching and the professional authority of faculty.

06 AUDIT FOR EQUITY AND GUARD AGAINST STRATIFICATION

AI adoption must be monitored for its potential to exacerbate educational hierarchies. Less-resourced institutions may rely more heavily on automated systems, while elite universities continue to offer mentorship and community as premium goods. Equity audits should ensure that automation expands access rather than deepening divides.

07 PLAN FOR PLATFORM DECAY AND VENDOR DEPENDENCE

Universities relying on third-party AI platforms must anticipate the risk of “enshitification” as vendors prioritize profit over pedagogy. Institutions should negotiate robust contractual safeguards, support open-source alternatives, and maintain the institutional capacity to adapt if the quality of these platforms declines.

NOTES

1. See, for example, Zuboff, 2019 on surveillance capitalism.
2. Early studies suggest that frequent AI use can correlate with diminished critical thinking, particularly among younger learners (see Gerlich 2025).

REFERENCES

Alonso, J. (2025, January 22). Study: Academic Alerts Slash Course Withdrawal Rates by 30%. Inside Higher Ed. https://www.insidehighered.com/news/students/retention/2025/01/22/research-shows-academic-alerts-impact-grades-withdrawals

Anscombe, G. E. M. (1957). Intention. Cornell University Press.

Burge, T. (2009). Primitive Agency and Natural Norms*. Philosophy and Phenomenological Research, 79(2), 251–278.

Davidson, D. (1980). Essays on actions and events. Oxford University Press.
Doctorow, C. (2023, January 2). Commentary: Cory Doctorow: Social Quitting. Locus Online. https://locusmag.com/2023/01/commentary-cory-doctorow-social-quitting/

Favero, L., Pérez-Ortiz, J.-A., Käser, T., & Oliver, N. (2025). Do AI tutors empower or enslave learners? Toward a critical use of AI in education (No. arXiv:2507.06878; Version 1). arXiv. https://doi.org/10.48550/arXiv.2507.06878

Feijo, S. (2025, July 21). MIT Learn offers "a whole new front door to the Institute." MIT News | Massachusetts Institute of Technology. https://news.mit.edu/2025/mit-learn-offers-whole-new-front-door-institute-0721

Floridi, L. (2025). AI as Agency without Intelligence: On Artificial Intelligence as a New Form of Artificial Agency and the Multiple Realisability of Agency Thesis. Philosophy and Technology, 38(1), 1–27. https://doi.org/10.1007/s13347-025-00858-9

Gerlich, M. (2025). AI Tools in Society: Impacts on Cognitive Offloading and the Future of Critical Thinking. Societies, 15(1), 6. https://doi.org/10.3390/soc15010006

Graham, S., & Perin, D. (2007). Effective Strategies to Improve Writing of Adolescents in Middle and High Schools.

Hughes, J. (2021). The Deskilling of Teaching and the Case for Intelligent Tutoring Systems. Journal of Ethics and Emerging Technologies, 31(2), 1–16. https://doi.org/10.55613/jeet.v31i2.90

Inman, W. (2022, March 21). Classroom Chatbot Improves Student Performance, Study Says. Georgia State News Hub. https://news.gsu.edu/2022/03/21/classroom-chatbot-improves-student-performance-study-says/

Johansson, P., Hall, L., & Sikström, S. (2008). From Change Blindness to Choice Blindness. Psychologia, 51(2), 142–155. https://doi.org/10.2117/psysoc.2008.142

Kestin, G., Miller, K., Klales, A., Milbourne, T., & Ponti, G. (2025). AI tutoring outperforms in-class active learning: An RCT introducing a novel research-based design in an authentic educational setting. Scientific Reports, 15(1), 17458. https://doi.org/10.1038/s41598-025-97652-6

Khairullah, S. A., Harris, S., Hadi, H. J., Sandhu, R. A., Ahmad, N., & Alshara, M. A. (2025). Implementing artificial intelligence in academic and administrative processes through responsible strategic leadership in the higher education institutions. Frontiers in Education, 10. https://doi.org/10.3389/feduc.2025.1548104

Manning, A. J. (2024, September 5). Professor tailored AI tutor to physics course. Engagement doubled. Harvard Gazette. https://news.harvard.edu/gazette/story/2024/09/professor-tailored-ai-tutor-to-physics-course-engagement-doubled/

Mowreader, A. (2025, January 22). Students and Instructors Say AI Tool Helps With Understanding, Confidence in Course Materials. Inside Higher Ed. https://www.insidehighered.com/news/student-success/academic-life/2025/01/22/survey-college-students-enjoy-using-generative-ai

Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. Psychological Review, 84(3), 231–259. https://doi.org/10.1037/0033-295X.84.3.231
Okasha, S. (2018). Agents and Goals in Evolution. Oxford University Press. https://doi.org/10.1093/oso/9780198815082.001.0001

Henry L. Roediger, I. I. I., & Karpicke, J. D. (2006). Test-Enhanced Learning. Psychological Science. https://journals.sagepub.com/doi/10.1111/j.1467-9280.2006.01693.x

Steward, H. (2009). Sub-intentional Actions and the Over-mentalization of Agency. In New Essays on the Explanation of Action (pp. 295–312). Palgrave Macmillan.

ACKNOWLEDGEMENTS

The views of the authors of this report do not necessarily reflect the views of their affiliated institutions.

CONTACT

IEET
56 Daleville School Rd
Willington CT 06279
USA

info@ieet.org
+1-617-287-5420