INSIGHTS REPORT

AI's Effect on Grading, Assessment, and Course Evaluation

THE CHRONICLE OF HIGHER EDUCATION

WITH SUPPORT FROM watermark

Balancing human-centered pedagogy with the power of AI

As someone who transitioned from a student services professional and lecturer to the role of product visionary, I deeply understand the tension central to this report: the desire to provide meaningful, personal feedback to students against the reality of limited time and resources. The introduction of generative AI into grading and assessment presents a flurry of new opportunities.

At Watermark, we believe the path forward isn't about replacing the educator. We believe in solving educator's most pressing challenges so they can support, guide, and teach their students to the best of their ability.

Our approach is rooted in the belief that innovation demands an ethical, principled framework. We are committed to a human-in-the-loop approach, using AI to speed up analysis and uncover insights, while ensuring educators and administrators always have final oversight and control. By prioritizing transparency and fairness in our applications of AI, we aim to provide tools that increase efficiency without compromising the integrity of the assessment process or institutional data privacy.

We are proud to support *The Chronicle of Higher Education's* Insights Report as it explores how institutions are navigating this new terrain.

Brian

Brian Robinson
SVP Product, Watermark

TABLE OF CONTENTS

AI's Effect on Grading, Assessment, and Course Evaluation

By Sarah Lindenfeld Hall

Contact CI@chronicle.com with questions or comments.

AI's Effect on Grading, Assessment, and Course Evaluation was written by Sarah Lindenfeld Hall and underwritten by Watermark. *The Chronicle* is fully responsible for the report's editorial content.

Cover image: iStock

INTRODUCTION

ISTOCK

In spring 2024, Cameron Blevins, an associate teaching professor in history at the University of Colorado at Denver, decided to test the hype around generative artificial intelligence. He took three essays written in 2019, anonymized them, and then ran them, along with the assignment's rubric, through a custom AI tool trained to grade student essays.

The results surprised him. With some exceptions, the AI tool and Blevins, who also serves as his campus's digital-initiatives director, reached similar assessments. Same grades. Similar written feedback. The tool, he wrote in a blog post, could shave precious hours from his workload. But Blevins has yet to deploy a robot grader even as AI capabilities have leaped forward.

"My strong assumption would be that the actual technical capacities of these things to mechanically grade, assess, and give feedback on student work has probably gotten much better," he says. "But that is only one piece of the puzzle in terms of thinking about whether or not we should be using these and what the purpose of feedback and grading is."

While much attention has focused on students using AI to cheat or plagiarize, faculty members like Blevins are grappling with whether and how they might use generative AI in their own teaching practices.

"It is the most sweeping technological shift of my teaching career, and it's happened very quickly,"

Blevins says. "In conversations with colleagues, the biggest kind of sentiment I feel is really just confusion — in a sense of almost drowning under the pace and scale of these changes."

For faculty and staff juggling teaching, research, and administrative work, generative AI's promise is compelling. Tools could provide faster feedback to students, sift through and implement suggestions in course evaluations, and streamline accreditation obligations.

"My strong assumption would be that the actual technical capacities of these things to mechanically grade, assess, and give feedback on student work has probably gotten much better."

These efficiencies could offer genuine relief for harried faculty and administrators. Yet campus conversations around deploying AI also are tinged with words and phrases like "fear," "shame," and "existential crisis." Critics cite concerns about fairness, accuracy, and eroded student relationships, alongside AI's impact on the environment, mental health, and student privacy.

In a survey of 1,057 U.S. faculty members in November 2025 by the American Association of Colleges and Universities and Elon University's Imagining the Digital Future Center, respondents expressed deep concerns about the technology, and some 68 percent say their institutions haven't prepared faculty to use it for effective teaching.

"We are still very much in an early stage of any type of AI integration," says Susan Purrington, Connecticut College's generative-AI teaching and learning fellow, about higher ed.

But, ready or not, the technology is moving fast. This report examines how institutions are navigating these tensions and what policies and practices are emerging.

SECTION 1

Current Uses: Generative AI in Practice

ISTOCK

Using AI to evaluate student work isn't entirely new. In STEM classes, pattern-recognition software has streamlined grading for more than a decade. But generative AI expands those capabilities. Powered by large language models, it can read and evaluate content, including suggesting feedback and grades on student work, or update syllabi based on new teaching frameworks. So far, using AI for these purposes remains limited among faculty members. In an analysis of 74,000 Claude conversations from users with higher-education email addresses in May and June 2025, Anthropic found that only 7 percent involved assessing student performance. Similarly, the AAC&U survey found 11 percent of faculty use AI to grade at some level. Yet those numbers likely will grow as AI-embedded grading tools roll out in popular learning-management systems.

Using AI for grading and assessment varies drastically by discipline, pedagogy, and philosophy. What one instructor sees as cheating, another views as a critical-thinking practice, says C. Edward Watson, AAC&U's vice president for digital innovation and co-author of the recently updated *Teaching With AI: A Practical Guide to a New Era of Human Learning*. That spectrum is visible in how faculty are using AI today.

HIGH-STAKES ASSESSMENT: EAGER ADOPTERS

Some instructors are pushing the boundaries. Last fall, Panagiotis G. Ipeirotis, a professor at New York University's Leonard N. Stern School of Business, turned his traditional final exam into an oral exam, delivered via an AI voice agent, and convened a "council" of three large language models to grade his undergraduate students' work.

In his class on AI and machine-learning product development, AI use was encouraged, but it wasn't supposed to replace actual learning. Yet Ipeirotis found some students turning in top-notch work they couldn't verbally defend in class. So he and his co-teacher built the AI voice agent to conduct personalized oral exams for all 36 students, questioning them about their capstone projects and course case studies. The AI council graded the transcripts as did Ipeirotis and a teaching assistant.

The humans mostly agreed with the AI graders, he wrote in a blog post titled "Fighting Fire With Fire." In fact, he found the council graded more consistently — more strictly, but more fairly — than the humans. It also diagnosed where the instructor had rushed through the topic too quickly for the students to gain mastery.

Students didn't love the experience. Some 83 percent found the format more stressful than written exams. But Ipeirotis, whose expertise is integrating human and machine intelligence, is taking it further. This semester, students will design the quiz agent that gives them their final exam, he says.

In a few years, if it's set up with the right workflows, AI might eclipse his ability to teach a class or grade papers, he says, because the technology can adapt and generate "excellent learning material." He says, "I don't know if there will be a future for humans, but the education will be better," adding, "I'm a very glass-half-full type of guy."

FORMATIVE FEEDBACK: SUPPORTING STUDENT LEARNING

Most faculty members aren't so ready to be replaced. But some are finding benefits in using generative AI to enhance feedback, instructional design, and learning support. At the University of California at Davis, instructors and researchers have developed Peer & AI Review + Reflection, or PAIRR. Combining traditional peer review with AI-generated feedback, PAIRR prompts students to reflect on both as they revise. Researchers found that students felt the AI feedback was helpful when it was paired with peer feedback, and that they appreciated the complementary perspectives.

In a Harvard physics class, prompts were built with "comprehensive, step-by-step answers" to avoid hallucinations from the generative-AI tool.

Evidence also is emerging that AI tutors can support student learning. In January 2026 a Brookings report summarized the findings of four randomized controlled trials that found learning gains with AI-based tutors. Safeguards, however, are important, the report says. For example, in one of those studies, which took place in a Harvard physics class, prompts were built with "comprehensive, step-by-step answers" to avoid hallucinations from the generative-AI tool, or confident-sounding but entirely inaccurate generative-AI claims.

Liza Long is an eager adopter of AI tools and has taught the PAIRR model as an English instructor at the College of Western Idaho. She's also now an academic technology program manager for the Idaho State Board of Education and a Ph.D. student in English at Idaho State University. In her own graduate studies, she has created a bot

Faculty members use AI to redesign courses and assessment, which they say benefits students by creating clearer assignments, more complete rubrics, and better materials.

to help her practice translations and check her work in a Middle English class. She is working on another for her study of *Beowulf*. "By just thinking through the creation of that bot, that's an interesting learning problem," says Long, author of *Cyborgs and Centaurs: Academic Writing in the Age of Generative Artificial Intelligence*. "[It's] personalized to me, and I can go in and tweak it, and [it] learns from me what my mistakes are, what things I'm struggling with."

COURSE DESIGN: IMPROVING TEACHING PRACTICE

Other faculty members use AI to redesign courses and assessment, which they say benefits students by creating clearer assignments, more complete rubrics, and better materials.

Trey Conatser, assistant provost for teaching and learning at the University of Kentucky, uses AI to organize his thoughts to produce rubrics or better assignment prompts that follow frameworks such as Transparency in Learning and Teaching, or TILT.

"I found it helpful for cognitive assistance and putting my ideas together that ultimately will be clearer for students," says Conatser, who also is director of Kentucky's Center for the Enhancement of Learning and Teaching.

But any AI output is just a starting point, Conatser says. An AI-generated rubric for an essay might overemphasize grammar or polish versus content or disciplinary-specific skills. "The default assumptions of AI may be a bit reductive or off focus from what we think is important," he says.

At Berkeley College, Jason Gulya, a professor of English and media communications, has carefully waded into AI deployment in writing classes. He's had students use a chatbot to help build self-assessment rubrics and developed a custom "pressure bot" with prompts designed to ask students questions about a text and push for evidence. Students then turn the transcript in as part of their assignment, so he can see their thinking in action or the moment when the AI bot took over the direction of the conversation.

"I literally tell my students, 'Don't let AI push you around,'" says Gulya, who has consulted with other campuses on the role of AI in higher education. "It will pop you into a corner. It will direct things. It will take over. And you might not even realize it, especially because they're sycophantic by design."

COURSE EVALUATION, ACCREDITATION SUPPORT, AND FACULTY CAUTION

Some are exploring AI's benefits for more administrative tasks. Take course evaluations. It's easy enough to sift through 10 or 15 student-written assessments at the end of a semester. For instructors with large teaching loads or 500-student lecture classes, however, it's much more difficult to uncover any insights.

Though Conatser hasn't seen much real-world adoption of AI for distilling course evaluations, he sees the potential for tools to help process the qualitative feedback, identifying trends, filtering out unhelpful comments, and surfacing feedback that's actionable for instructors.

"I can see how, given the proper guardrails, governance, and so on, that you could have AI

built into the technical evaluation systems at universities to better manage the data," he says.

Paul Shovlin, an assistant English professor at Ohio University specializing in AI and digital rhetorics and an AI faculty fellow, says instructors could even build a chatbot trained on accumulated evaluations over time.

"If we had a chatbot that had a year and a half's worth of student evals," Shovlin adds, "we could say, 'Hey, I want you to take a look at the most current syllabus and then give me feedback based on student evaluations for the last year and a half. What are three concrete improvements I can make to this course?'"

Some instructors are using AI to distill student comments. Before she moved to Connecticut College, Purrington was teaching at Northern Arizona University, where she used a generative-AI tool to analyze, among other data points, student-feedback surveys over several years, alongside her own reflections in journals.

The exercise revealed, for example, when an assignment needed an extended deadline given other coursework or campus activities, and pulled together recurring themes from students' open-ended comments, highlighting both success and growth areas, she says.

Purrington is now consulting with another college to create a bot to help with cyclical accreditation obligations. The tool will cross-reference existing institutional data with accreditation requirements, identifying gaps, and pointing faculty toward potential sources.

The common thread in each of these cases, faculty caution, is that AI shouldn't replace human judgment, it should enhance it.

The common thread in each of these cases, faculty caution, is that AI shouldn't replace human judgment, it should enhance it. Shovlin opposes using chatbots as substitutes for peer feedback in writing courses. The practice of reading someone else's writing and providing observations is valuable and can't be replicated by AI, he argues.

Yet he also recognizes a practical reality. Sometimes students are working at midnight when a peer or professor isn't available for support. "Is it better to not give them anything, or is it better for them to have access to a tool that they might elect to use?" he says.

It's one of the questions that higher-education leaders are navigating as they consider the pros and cons of the technology.

SECTION 2

Controversies and Concerns: The Birth of the 'Dead University'?

ISTOCK

Concerns swirling around AI extend to issues far greater than what happens in the classroom. Critics point to stolen intellectual property used to train large language models, massive environmental impacts from energy-intensive data centers, and troubling mental-health implications. Chatbots have been linked to suicides.

"I tell folks I struggle with my position every single day," Purrington says. "From a practitioner side, I am fearful of people turning to chatbots for companionship. It terrifies me to know that children are killing themselves because they are interacting with something they think is real."

Some educators fear we're witnessing the birth of the so-called dead university — a place where AI replaces human interaction at every stage of learning, from AI bots delivering lectures to students using AI to complete assignments to AI tools grading student work.

"It's kind of like two Teddy Ruxpins talking to each other," says Amanda Sturgill, an associate professor of journalism at Elon University, referring to the animatronic teddy bear. "And then we've really not met our mission to society as higher education. Higher education is supposed to make better people, and we're all just pretending."

While the scenario might sound extreme, it raises fundamental questions. "It gets to not only the question of why are we here in the first place but our value proposition as a university," Conatser says.

Many others echo his concern, considering this an existential crisis that threatens the future of higher education just as it confronts mounting

challenges, including growing budget pressures, enrollment drops, and declining public confidence in the value of a college degree.

The potential hit to institutional credibility is especially sharp given what we know about how students learn best and how colleges market themselves, says Kevin Gannon, a history professor and director of the Center for the Advancement of Faculty Excellence at Queens College of Charlotte.

Decades of research demonstrate the importance of student-faculty relationships in learning, Gannon says. He points to the work of Peter Felten on relationship-rich education, which shows that mentoring relationships with both faculty and staff can improve student satisfaction and outcomes. College-admissions offices regularly play up that personal touch in their marketing to prospective students.

"And then now we're like, 'Nah, we'll let the robots do this,'" Gannon says. "It's a bait and switch."

From the student perspective, the calculus becomes simple: "Why would I write something that someone else wouldn't read?" Gannon says. In spring 2025, one Northeastern University student asked for her tuition back — and made headlines in *The New York Times* — after she realized that her instructor was using generative AI for course materials.

Beyond these broader concerns, critics point out that AI grading presents practical problems too.

For now, it's not reliable. "I'm going to get something that looks like a reasonable facsimile of what a good set of feedback or assessment data might look like," Gannon says, "but it may or may not be correct."

When instructors offload assessments to algorithms, they lose important insights about student progress. "It prevents us from telling the

"And then now we're like, 'Nah, we'll let the robots do this.' It's a bait and switch."

story of student learning," he says. "If I'm using generative AI to assess student writing, and if writing is a process, I'm not familiar with what my students are doing from step to step."

And then there is the potential loss of learning for students and faculty. In 2025, early research has found that outsourcing work to AI could homogenize writing and lead to cognitive losses for those who use it.

An MIT Media Lab study asked 54 individuals to write several SAT essays using either a generative-AI tool, a Google search engine, or their own brainpower. Researchers measured the subjects' brain activity using an EEG, or electroencephalogram. AI users "consistently underperformed at neural, linguistic, and behavioral levels."

"It offloads the deeper processes that are actually necessary to learn something meaningfully," Gannon says. "And, of course, the more you offload, the less good you're going to be at it. It's like I stopped going to the gym and then, all of a sudden, wonder why I can't lift 10 pounds anymore."

Even cautious approaches carry risks. Using AI as a "first pass" for grading that faculty can adjust could trigger anchoring bias, a tendency to rely too much on initial information. "The research says, if I go in and I see an 89, I can't unsee it," Gulya says. "There is a way in which it's already started to nudge my judgment."

> **"Looking to AI as the magic-bullet solution strikes me as kind of a failure of leadership."**

Some worry that even well-intentioned AI-assisted use will erode over time. Human nature, after all, is to take the easy way out. And instructors aren't being trained for proper use.

Shovlin points to how quickly professors came to rely on AI writing-detection tools, even though he sees those systems as flawed and ethically fraught. As learning-management systems integrate more AI, that makes Shovlin nervous. "I don't think faculty are adequately trained on how to use those features," he says. "There's going to be problems."

The quick acceptance of AI-writing detectors raises concerns about what could happen next, especially as budget pressures mount, the technology improves, teaching loads grow, and the system includes adjunct professors attempting to manage hundreds of students at multiple campuses.

"Looking to AI as the magic-bullet solution strikes me as kind of a failure of leadership," Gannon says. "It's the easy way out. You've sort of accepted the sale. You've taken off your critical-thinking hat. You haven't addressed some of the complex ethical questions that surround it."

The path forward will require more than just deploying AI for deployments' sake, he says.

"I'm an impatient person. I'm an early tech adopter on everything, and every instinct I have is saying, 'Pump the brakes.'"

SECTION 3

The Path Forward: Deliberative Innovation

ISTOCK

Higher education's traditional pace of decision-making is struggling to keep up with the rapid integration of AI into campus systems. Yet, the need for careful, thoughtful deliberation has never been greater.

Across the country, a small number of early adopters are beginning to grapple with how to move forward responsibly. Connecticut College, for example, brought in Purrington to support intentional AI implementation. Many campuses, especially smaller ones, however, have not set aside resources for dedicated AI leadership.

Small institutions "are really just grasping what this technology is, and how it really is impacting individual disciplines," Purrington says.

Some larger campuses, such as the University of Georgia, are moving more quickly. In spring 2023, Georgia convened a faculty symposium on generative AI to draft recommendations and identify where faculty members felt excited, nervous, or stuck, says Megan L. Mittelstadt, assistant vice president for instruction and director of Georgia's Center for Teaching and Learning.

The university then created a Provost's Committee on AI, which includes a subcommittee on AI and instruction with representation from students, the faculty, and the information-technology and student-affairs departments. That work produced an AI Hub website with institutionwide guidance. It's a work in progress. In January 2026, the guidance was updated with detailed recommendations on the ethical use of AI for grading and feedback.

The new guidance permits generative AI for developing grading rubrics, assessment criteria, and some high-level grading as long as no student-specific information is entered.

It prohibits fully automated grading without any instructor review and the use of publicly available generative-AI tools without student consent. Georgia offers three AI tools integrated into its secure IT environment for students, faculty, and staff.

"One of the things that we stumbled on early on as a theme for our conception of responsible use of generative AI, broadly writ across disciplinary boundaries, is the conception of [the] human in the loop," Mittelstadt says. "That idea governs everything. Every piece of guidance that we provide, there should always be a human brain in the loop."

As campuses consider what's next, experts say the path forward should include these highlights.

INCLUDE ALL VOICES

For Flower Darby, associate director of the Teaching for Learning Center at the University of Missouri, but who is speaking from her own experience working with other campuses on the topic, the most important path forward is to meet people where they are.

"AI is the most highly individualized topic that I have ever seen in 30 years of higher education," she says. "People have strong opinions, and it's very individual. Many faculty are not really using it yet. They're worried about it. They're afraid of it. We have super users who come in with their enthusiasm and actually just overwhelm beginners or skeptics, and I don't know that that's really productive either."

That means campuses must foster constructive dialogue across differences and respect and wrestle with the concerns that faculty have, Darby says. "In academics, the superpower that we have is critical thinking, poking holes, diving deep into the philosophy of teaching and learning. Those are our superpowers. We need to bring them to bear on this situation that we find ourselves in."

"One of the things that we stumbled on early on as a theme for our conception of responsible use of generative AI, broadly writ across disciplinary boundaries, is the conception of [the] human in the loop."

Indeed, Shovlin says, AI skeptics shouldn't be treated as obstacles to progress. "I want those people to feel like they are not extraneous, vestigial organs that are going to wither up and die," he says. "They need to be central parts of discussions."

BUILD SUPPORT INFRASTRUCTURE

Generative AI has been easy to access for nearly four years, but AI literacy is still lacking among every stakeholder group on campus. Thoughtful AI integration requires sustained investment in both people and processes. Rather than relying entirely on slow-moving committees for decision-making to guide AI deployment, institutions should invest in individuals dedicated full time to campuswide AI or emerging-technologies integration, Purrington recommends.

"It doesn't need to be a computer scientist to do the work," says Purrington, whose Ph.D. is in leisure behavior. "But really allow that person to have some autonomy, to make decisions in a quick manner, or have the chain of command be really short to the person who can, so that you can keep pace." In many cases, that expertise already exists in campus teaching and learning centers. "Those

are the multidisciplinary people that are engaged in evolving technologies and how they impact teaching and learning," Mittelstadt says. "They're thinking about this every day."

Still, one-off trainings aren't enough. Faculty members need regular "microdoses" of information — quick updates on new features and opportunities to share what's working and what's not, Gulya says. Sustained support will help instructors and staff stay current with tools that are constantly changing.

To encourage participation, institutions must also grapple with time and workload. That may mean course releases, stipends, or dedicated time for exploration.

"If colleges and universities had the capacity to foster all kinds of conversations, that would mean workload reallocation," Darby says. "That would mean maybe we would find money for incentives — free lunch or something — just to help people take the time to figure this out and learn from and with each other."

DEFINE WHAT SUCCESS LOOKS LIKE

Higher education was already at a crossroads long before the first bot entered the chat. Some instructors were shifting away from high-stakes final exams toward process-based assessments. Institutions were searching for better ways to engage, educate, and retain students.

"The problems that AI is exposing were here way before AI," says José Antonio Bowen, former president of Goucher College and co-author of the book *Teaching With AI: A Practical Guide to a New Era of Human Learning*. "The problem with stand and deliver to 700 freshmen — that is not a problem AI created. AI is shining a light on the fact that some of what we did was stupid."

The question, Gulya argues, is whether colleges will use AI to tackle those longstanding problems or accelerate them. One of the biggest risks, Gulya says, is assuming generative AI automatically "disrupts" the system for the better.

"I actually don't think that's true," he says. "It can be an opportunity to disrupt and rethink things, but it can also just as easily become a tool for doubling down on what we already do."

In that scenario, AI could simply accelerate inauthentic learning — students completing assignments laser-focused only on their final grade, faculty grading just to meet requirements. Lecture-style teaching might become faster and more efficient, but also less effective.

"If this technology is going to be truly disruptive, we have to do a lot of work into figuring out how it can be disruptive, and what exactly it's disrupting," Gulya says. "If we're disrupting something, what are we replacing it with?"

Higher education was already at a crossroads long before the first bot entered the chat.

CONCLUSION

SECTION 4

‘We Need to Be in the Game’

CONCLUSION

ISTOCK

Of course, generative AI isn't standing still. The technology's capabilities are accelerating, and adoption is spreading whether or not institutions are prepared. Consider this: Incoming college freshmen this fall will have had access to ChatGPT since ninth grade. Generative-AI tools have been a part of their academic life — used by students to study or complete assignments and by teachers to grade or provide feedback.

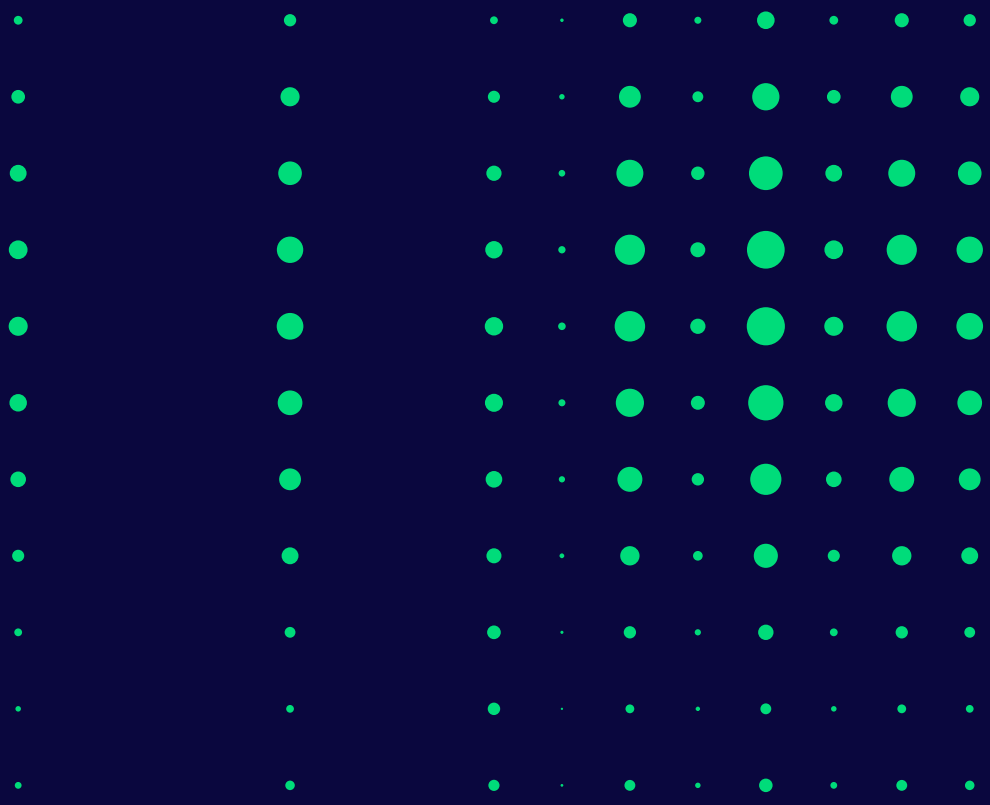
"We're so used to thinking about this technology as something that just happened. And, to some extent, that's true," Gulya says. "But if you're an 18-year-old who's had access to this since they were 14 or 15, those three years are extremely formative, especially when it comes to schoolwork."

Professors and administrators say they are now navigating a disruption they never asked for, arriving at a speed far faster than institutions are built to absorb. Yet amid the division over how to move forward, many see an opening.

As she contemplates what's next, Mittelstadt remains focused on keeping "art of teaching" at the center of the conversation. "That human connection, that human feedback, that ability to look into someone's eyes and see when they're struggling and react on a human level," is key, she says.

For Mittelstadt, this is an energizing time, but also clarifying. "It's a real opportunity to re-interrogate our teaching values and what makes teaching effective and lean more heavily into that," she says. "The technology changes over time. It's a constant set of learning priorities [and] … going back to basics: What is good teaching? What is good learning? In what way should the generative AI be supportive? In what way should it not be?"

The choice, Bowen warns, is stark. "AI has the potential to do enormous damage. … But if we do nothing, it will happen to us. We need to be in the game."



THE CHRONICLE
OF HIGHER EDUCATION

900 19th Street, N.W.
Washington, D.C. 20006
202.466.1000 | Chronicle.com