HANOVER RESEARCH

ASSESSMENT AND GENERATIVE ARTIFICIAL INTELLIGENCE IN HIGHER EDUCATION

April 2024

In the following report, Hanover Research examines research literature, trade publications, and case studies highlighting emerging strategies in higher education to account for, or even incorporate, generative artificial intelligence when designing student assessments.

TABLE OF CONTENTS

EXECUTIVE SUMMARY

RECOMMENDATIONS

Based on an analysis of generative AI assessment case studies and best practices, Hanover recommends that institutions:

AVOID RELYING ON TOOLS LIKE GPTZERO OR OTHER AI DETECTION SERVICES, WHICH HAVE AN EXTREMELY POOR ACCURACY RECORD AND MAY MAKE THE CLASSROOM CLIMATE UNECSSARILY ADVERSARIAL.

OpenAI, the maker of ChatGPT, experimented with its own AI detection service, which achieved a success rate of just 26 percent. Experts are similarly skeptical of other third-party vendors selling detection systems.

REVAMP WRITING ASSIGNMENTS TO INCLUDE SCAFFOLDED DELIVERABLES OVER A PERIOD OF TIME AND FOCUS ON STUDENTS' DEVELOPMENT PROCESSES.

Writing assignments, which are the most common application for generative AI, should be modified to include intensive, iterative instructor feedback and revision over time. This shift to a process focus amplifies the use of formative assessment to develop students' skills, and somewhat de-emphasizes the final product.

INCORPORATE GENERATIVE AI PRODUCTS INTO STUDENTS' WRITING AND ANALYSIS ASSIGNMENTS.

Generative AI tools can be used to help students revise or critique their own writing, to generate and compare ideas and evidence, and as writing artifacts that allow students to contrast human writing with AI-produced work. Learning to brainstorm and write with AI, understand its limitations and weaknesses, and recognize its biases or blind spots are essential skills for navigating a world where generative AI will be commonplace.

KEY FINDINGS

In the year since OpenAI released ChatGPT 3.5, many of the technology's promised applications like formative assessment and personalized interactive tutoring remain under-realized. The leaders in applying these technologies for uses beyond writing student term papers are often technology companies, rather than colleges or universities. Recent reviews of GPT-powered tools like Duolingo's RolePlay and Explain My Answer and Khan Academy's Khanmigo suggest that they are useful, and more interactive than earlier chatbots, but still not close to replacing human instructors.

Survey data from the spring 2023 semester suggest that around 30 percent of postsecondary students had used generative AI in their schoolwork, and that 51 percent of respondents view it as a form of cheating or plagiarism. Of the subset of users that report "frequent" use (46 percent) only eight percent believed that generative AI had improved their grades, which may reflect the ongoing challenges these tools have with producing sharp arguments and convincing personal narratives. The ongoing debate among students about the ethical implications of using generative AI likely means that faculty will find most students receptive to conversations about academic honesty and appropriate use and citation.

To make assignments more generative AI-resistant, faculty are relying more heavily on process-focused formative assessments in writing assignments, in-class work, and group projects. Assignments that specifically ask students to use (and cite) generative AI resources and then evaluate their outputs are also a highly-recommended practice. The overall assumption among universities appears to be that these resources are going to be ubiquitous in higher education and the workplace, and that students must be equipped to interact with them critically and responsibly.

INTRODUCTION AND METHODOLOGY

Institutions anticipate significant shifts in the labor market, driven by the advent of ChatGPT. These changes necessitate a fundamental rethinking of higher education institutions' operations, including curriculum design and teaching methods. To better prepare for these impending changes, Hanover summarizes emerging trends in this area.

REPORT CONTENTS AND STRUCTURE

This report includes two sections:

- **Section I: Landscape Review – Generative AI In Assessment** provides an overview of generative AI and its impacts on assessment in a postsecondary education environment. The analysis draws on academic studies, trade publications, and news features to map out the current state of generative AI and its impacts on, and role within, assessment and evaluation of student progress.
- **Section II: Generative AI Case Studies and Assessment Use Cases** looks at three university examples and one education technology industry example of strategies for helping students to navigate the impacts of generative AI in the workforce and the world at large. It begins with a case study looking at AI as a tutoring and assessment resource in subscription-only versions of the language learning app Duolingo. The profiles of the University of Wisconsin and Montclair State University resources on generative AI offer insight into universities' evolving perspective on their role preparing students to function in a world and workplace with ubiquitous access to generative AI. These profiles, plus the third university case study featuring Australia's Monash University, also provide insights into how universities are rethinking assessment to mitigate students' unauthorized use of AI to complete assignments.

RESEARCH QUESTION

What are the emerging trends in discussions about how generative artificial intelligence (AI) is changing the way higher education institutions teach, particularly in relation to competency-based education and assessment?

SOURCE AND METHODOLOGY NOTE

Chat GPT 3.5 launched on November 30, 2022, making widespread public access to Large Language Model (LLM) artificial intelligence tools a very recent phenomenon. With this in mind, institutions of higher education, to say nothing of other industries, have been forced to adapt very quickly and are still navigating the complexities of generative AI and its impacts on education, commerce, and society at large. The rapid release of other, sometimes updated, resources such as Chat GPT-4, Bing (based on GPT-4), Dall-E 2 (which creates images), and Bard (Google) means that strategies that were sound a year ago may no longer be so today. Hanover expects that generative AI guides and resources produced by higher education institutions will continue to evolve in the coming year, and that we will also have much more insight into the performance and limitations of newly-released generative AI-powered tools with continued exposure. Given the novelty of the topic and the fact that many innovations are occurring outside of higher education, we have including non-university examples.

LANDSCAPE REVIEW – GENERATIVE AI IN ASSESSMENT

Overview of generative AI as an assessment tool in higher education and beyond.

HOW ARE STUDENTS USING GENERATIVE AI?

In a May 2023 opinion piece in *The Chronicle of Higher Education*, Owen Kichizo Terry suggests that most university students do not rely exclusively on Chat GPT and similar tools to write essays; instead they use it to craft a framework for their argument. His example recounts the use of Chat GPT (most likely the free 3.5 version available at the time of publication) to identify a debatable thesis for a six-page term paper on *The Iliad*, and to help him come up with supporting arguments and evidence from the text. For the most part he wrote the essay at the sentence level, but he recounts that "all that was left now was for me to follow these instructions, and perhaps modify the structure a bit where I deemed the computer's reasoning flawed or lackluster."

> ***"In reality, it's very easy to use AI to do the lion's share of the thinking while still submitting work that looks like your own. Once that becomes clear, it follows that massive structural change will be needed if our colleges are going to keep training students to think critically."***
>
> *The Chronicle of Higher Education, Owen Kichizo Terry, May 2023*

A September 2023 survey published by Intelligent.com found that 30 percent of students had used ChatGPT for schoolwork in the previous year. Among them, 46 percent said they "frequently" use the tool and that English, followed by hard sciences such as chemistry and biology, are the disciplines where they are most likely to deploy it. Only one in twelve chat GPA users credited the tool with raising their GPA, however. According to a March 2023 *Best Colleges* survey, 51 percent of college students agree that using generative AI on assignments is a form of cheating or plagiarism, but 22 percent had used the tool.

A RANGE OF FACULTY RESPONSES

Responses to the rise of generative AI range from alarm to limited concern and include calls to embrace the technology as a part of students' learning processes and to offer at least some forms of assessment where students cannot access the technology. On the less reactive end of the spectrum, John Warner's April 2023 *Inside Higher Ed* blog post, "ChatGPT and Writing Assessment, an Old Problem Made New," contends that generative AI writing achieves "surface-level fluency" that requires faculty to grade more stringently for the substance of students' arguments. He argues that in the post-GPT world writing assignments must be "tied to authentic occasions for learning" and repeatedly pushing students to sharpen and deepen their arguments over multiple drafts. As the case studies in Section II will show, this focus on process rather than product is a common strategy.

Meanwhile, Inara Scott's April 2023 *Inside Higher Ed* article sounds growing alarm about the pervasive use of generative AI in student work, and the impacts it is likely to have on their learning:

> Back in January, I, like many others, thought we could design our coursework to outwit students who would rely on AI to complete their assignments. I thought we could create personalized discussion questions, meaningful and engaged essay assignments, and quizzes that were sufficiently individualized to course materials that they would be AI-proof. Turns out, I was incorrect. Particularly with the arrival of GPT-4, there is very little I can assign to my undergraduates that the computer can't at least take a stab at. Students may have to fill in a few details and remember to delete or add some phrases, but they can avoid most of the thinking—and save a lot of time.

One of her proposed solutions, at least in the short term, is to have students revise work that an AI checker like ZeroGPT determines to be more than 50 percent AI-generated.

COMPETENCY-BASED EDUCATION AND GENERATIVE AI

GENERATIVE AI AND COMPETENCY-BASED EDUCATION

Data scientist J. Rogel-Salazar's February 2023 discussion of generative AI in competency-based education suggests that a collaborative approach to AI could allow the tool to function as a coach for students and a formative assessment tool for faculty. He defines "competency-based education" as "a student-centred approach that focuses on mastering skills and knowledge rather than accumulating credit hours" and which "focusses on the student's demonstration of desired learning outcomes as central to the learning process." Rogel-Salazar's primary example of such a tool and how it might work is quoted below. It derives from a November 2022 article in *Proceedings of the ACM on Human-Computer Interaction*.

USE-CASE SPOTLIGHT – CONVOWIZARD

Consider for instance the tool recently unveiled by researchers at Cornell University called ConvoWizard. The tool can detect when online debates are becoming heated and could lead to an irredeemable meltdown. ***ConvoWizard is a browser extension that uses a neural network trained on data pulled from the Change My View subreddit to warn users when their comments are likely to escalate tension.*** *You can read more about this in the paper here. In tests, more than half of participants reported that the warnings stopped them from posting a comment they would have regretted, while 68% felt the tool's estimates of risk were as good as or better than their own intuition.*

In other words, the tool lets users know when the conversation is getting tense as they write their replies and provides warnings as to whether their comment will escalate tension. ***One could argue that AI tools such as ConvoWizard can help support competency-based learning by enabling high-quality online discussions: students can receive real-time feedback that helps them develop communication skills that are critical for success in many careers.***

WHY COACHING IS ESSENTIAL TO COMPETENCY-BASED EDUCATION

Neither Salazar's proposal to use generative AI as a coaching and assessment tool nor the competency-based learning framework he attaches it too differ substantively from established best practices in postsecondary teaching. Teaching professional development service TeachThought, which provided the definition of Competency-Based Learning cited by Rogel-Salazar, compares it with traditional learning models which they say rely on summative assessment. The two definitions are provided below, and the key difference between the models is the purported "flexibility" of CBE, which enables students to progress at their own pace through content after achieving mastery. However, "its effectiveness…depends on the ecology it is embedded in," and CBE requires "diverse support systems, robust assessment forms, and clear and manageable learning outcomes," placing the burden of ongoing assessment on faculty. Again, a focus on learning as a process is central.

DEFINING COMPETENCY-BASED EDUCATION

Definitions quoted from TeachThought, 2016.

Competency-Based Learning	Traditional Education
"In a competency-based learning system, students are not allowed to continue until they have demonstrated mastery of the identified competencies (i.e., the desired learning outcomes to be demonstrated). In this way, the definition of competency-based learning is closely tied to mastery learning." "Its strengths lie in its flexibility, as learners are able to move at their own pace. This supports students with diverse knowledge backgrounds, literacy levels, and other related aptitudes."	"In other learning models, students are exposed to content–whether skills or concepts–over time, and success is measured summatively."

OpenAI BEST PRACTICES FOR FORMATIVE ASSESSMENT

OpenAI ON GENERATIVE AI AS A PEDAGOGICAL TOOL – CURRENT STATUS

OpenAI's own Educator FAQ resource contends that AI detector resources generally are not reliable enough to justify "judgments about students with potentially lasting consequences," and recommends alternative, instructor-sanctioned uses of AI that highlight the development of students' critical thinking faculties. They provide three examples of how ChatGPT can help instructors develop students' capabilities in ways that could be assessed (reproduced in the figure to the right). The expert consensus on AI detection tool like GPTZero is that they do not work, and even OpenAI's experimental AI Classifier achieved only a 26 percent accuracy rate, which is worse than a random guess.

At present, OpenAI's recommendations place the burden on students to be honest about their use of generative AI and faculty to analyze the ways in which students interact with AI-generated text to demonstrate critical thinking skills. Student interactions can be saved via Shared Links and then evaluated. John Warner's essay on pedagogical uses of ChatGPT warns against the temptation of using AI to "substitute for human response to student writing" even as he concedes that "there's some occasions where AI responses may be genuinely helpful to student learning." He agrees with the OpenAI best practices outlined to the right, which stress the value of formative assessment of students' processes as the most essentially human part of writing assessment:

> …because writing is an embodied process and writing is thinking, the best feedback on student writing is not summative—which LLMs will do passably—but formative, where the instructor can help the student reconsider and reflect upon some part of their process. Asking a writing teacher to do this where they have not read the student work is like asking a coach to work with a team where they know the score but have not watched the game itself.

THREE GOALS FOR EFFECTIVE AI USE IN THE CLASSROOM, WITH RELATED STRATEGIES

Figure reproduces content from OpenAI. The content shown in orange italic font is particularly relevant to questions of assessment.

1. Showing their Work and Formative Assessment	**Educators can analyze student interactions with ChatGPT** *to observe critical thinking and problem-solving approaches.* **Shared links can enable students to review each other's work,** fostering a collaborative environment. *[Shared links are a new feature that allow users to generate a unique URL for a ChatGPT conversation, which can then be shared with friends, colleagues, and collaborators. Shared links offer a new way for users to share their ChatGPT conversations, replacing the old and burdensome method of sharing screenshots.]* **By keeping a record of their conversations with AI, students can reflect on their progress over time.** They can see how their skills in asking questions, analyzing responses, and integrating information have developed. *Teachers can also use these records to provide personalized feedback and support individual growth.*
2. Information and AI Literacy	**Students can demonstrate their ability to interact with AI and their understanding of the shortcomings of AI systems.** *Educators can assess the quality of the questions asked, the relevance of the information obtained, and how well the student understood to challenge, double-check, and consider potential biases in that information.* **We anticipate a future where the use of AI tools like ChatGPT is commonplace.** *Encouraging responsible use helps students prepare for a future where they may be expected to leverage AI in different contexts.*
3. Creating Accountability	**Sharing interactions with the model ensures that students are held accountable for the way they use AI in their work.** *Educators can verify that students are engaging with the tool responsibly and meaningfully, rather than simply copying answers.*

OPTIMAL GENERATIVE AI AND FACULTY ROLES IN INSTRUCTION

GENERATIVE AI AS A TUTORING SERVICE

As demonstrated by nascent products like Khanmigo and Duolingo's RolePlay (discussed in more detail on subsequent pages), one of the most promising applications of generative AI is as a tutor to struggling students. MIT Horizon's December 2023 blog post, *Can Generative AI Unlock Technology-Enabled Tutoring, for Everyone?* Notes that a 2020 National Bureau of Economic Research metastudy found tutoring to be "among the most flexible and potentially transformative learning program types available at the PreK-12 levels" (57). The report, which focuses on the use of human tutors, notes that:

> With effect sizes averaging at over a third of a standard deviation and impacts consistently significant across a wide range of program and study characteristics, our review's meta-analytic findings demonstrate not only the power of tutoring, but its versatility. As customized learning grows in prominence across today's educational systems, there is little doubt that tutoring programs will constitute a key workhorse policy model. (57-58)

The MIT authors suggest that research links two elements of tutoring programs to student success, and then asks whether either are "things generative AI can do well." The drivers of success they identify are:

- Social connection and motivational support
- Cognitive scaffolding that unlocks a student's own thinking

While they acknowledge that ChatGPT and similar resources can "adopt particular tones and strategies that build up rapport" they contend that "it is not as appropriate for the longer-term social connection that comes from having a caring mentor who is invested in your success and who checks in with you." In terms of cognitive scaffolding, ChatGPT tools like Khanmigo are getting better at helping users think through how to come up with a correct answer or conclusion, rather than simply producing it.

CONTINUED RELEVANCE OF HUMAN INSTRUCTORS

Multiple sources recognize both the emerging promise of generative AI as an always-on tutoring and formative assessment tool, but suggest that the social connections found in teaching and learning relationships are essential for motivating many students. Two iterations of this idea are quoted below, and both of them note the continuing need for human oversight of the teaching and learning process, as well as the role of AI in providing targeted feedback and practice for students.

The idea of AI tutoring is compelling enough that the Defense Advanced Research Projects Agency (DARPA) is soliciting proposals "to create customized learning experiences that improve training of new skills in adults who have completed postsecondary education." Their focus is on "complex subjects required for national security, such as AI engineering."

AI TUTORS AND HUMAN TEACHERS

Figure reproduces content from MIT Horizon and Eric Hudson.

MIT Horizon

"…having social connections that help you, both by keeping you accountable and by rooting for you over the longer term, are critical. **Perhaps a productive model for the future involves a human 'meta-tutor' that helps guide a student's interactions with their personal robot tutor, with the human helping build your personalized path and stopping to help you engage with a generative AI tutor when you get stuck."**

Eric Hudson

"Timely and personalized feedback from a trusted expert is the most effective form of feedback. Yet, this level of support is onerous, if not unsustainable, for most teachers. **It might not be, however, with the support of AI."**

"A core tenet of competency-based education is that 'learning is the constant, time is the variable.' Personally, I don't yet have the confidence that AI can define, assess, and credential 'mastery' on its own. However, **I see immediate applications for AI to introduce more flexibility in time, place, and pace of learning for students, especially in formative assessments."**

OpenAI CURRENT ASSESSMENT PRODUCTS

ChatGPT USE CASES OUTSIDE OF HIGHER EDUCATION

According to OpenAI, partner organizations outside of higher education are building "AI powered education tools" with the potential to provide opportunities for practice as well as formative and summative assessments. The examples below derive from OpenAI's website.

OpenAI OVERVIEW OF ChatGPT-POWERED PEDAGOGICAL TOOLS

Figure reproduces content from OpenAI.

"Khan Academy, a nonprofit that offers online lessons to students of all ages, uses GPT-4 to power ***Khanmigo, a tool that functions as both a virtual tutor for students and a classroom assistant for teachers."***

"Canva, an online design platform, uses OpenAI's large language models to power Magic Write. It offers Magic Write for free to educators, who ***use the tool to create presentations, classroom activities and lesson plans."***

"Duolingo, a language online learning company, uses GPT-4 to power ***RolePlay, an AI conversation partner that practices real world conversation skills*** with learners, and ***Explain My Answer, which learners can use to gather deeper understanding on their mistakes."***

"edX, a global online learning platform, ***uses GPT4 and GPT3.5 to support digital tools that deliver real-time academic support and course discovery assistance*** to online learners."

EARLY RESPONSES TO THE ChatGPT-BASED TOOLS

In an October 2023 *Forbes* feature, Charles Towers-Clark evaluated the promise and shortcomings of Khanmigo, finding that it "only acts as a teaching assistant." Whether the tool is as interactive as promised remains to be seen, and Towers-Clark's experience suggests that the line between coaching students through a task and doing it for them was often blurry:

> It aids in grading papers, refreshing teachers' knowledge, and crafting lesson plans. Moreover, it provides personalized guidance to students, offering assessments for teachers to follow up on. When it comes to its intended functionality, Khanmigo appears to deliver and my experience was positive. It encouraged me to write, offering suggestions on how to restructure my content and incorporate additional relevant points. **However, it didn't strictly adhere to the Socratic method; I could prompt it to rewrite text on a paragraph-by-paragraph basis when I posed the right prompt.**

Nadia Bidarian's August 2023 profile of Khanmigo published by CNN suggests that its propensity to hallucinate, or generate incorrect or made-up answers, remains a central problem. This tendency appears when Khanmigo serves as an interactive writing or research coach and as a mathematics coach, where it can provide incorrect answers.

The new edX platform, called Xpert, is described as "a generative AI-powered learning assistant" and paired with an edX ChatGPT plugin "which enables ChatGPT Plus users to seamlessly discover higher education programs across edX's library of courses." Details about this platform and its impacts remain limited, but the platform is meant to provide customized academic assistance, course discovery services, customer service, and course content summaries.

GENERATIVE AI CASE STUDIES AND ASSESSMENT USE CASES

Examples of how universities globally and in the United States, as well as other educational entities, are using generative AI in assessment.

CASE STUDY – DUOLINGO *ROLEPLAY* AND *EXPLAIN MY ANSWER*

CURRENT DUOLINGO DEPLOYMENT OF AI

Language learning app Duolingo introduced two ChatGPT-powered tools in spring 2023 as part of their $14 per month Max subscription. As of August 2023 the two AI tools are the flagship features of its top-tier subscription level. They include RolePlay, which "gives users the opportunity to take everything they learn in their ordinary lessons and use it in a conversation with one of Duolingo's characters," and Explain My Answer, which "goes into detail on the answers you provide in your lessons, highlighting what you got right, where you've gone wrong, and where you can improve."

The ChatGPT-4-based RolePlay uses AI to converse with users in a simulated encounter (the main example given in reviews of the tool is ordering a drink in a café) and to "ensure the conversation is, broadly speaking, remaining focused on relevant topics" rather than veering off into irrelevant directions. Carina Chocano's April 2023 *New Yorker* article on Duolingo's use of AI describes it as follows:

> …as GPT-4 and the human user generate dialogue in RolePlay, a separate machine-learning model monitors the results, and registers whether they are within the projected range of appropriate conversation. 'If it's out of scope,' [Duolingo head of artificial intelligence Klinton Bicknell] said, 'then we just tell the learner, 'Hey, I think you're straying a little off topic.'"

Future deployments of AI at Duolingo will focus on generating course content and exercises according to an April 2023 *Forbes* article:

> Generating new course content has traditionally been a bottleneck – and this is one job in particular that the present-day generation of language models has proven to be highly efficient at.

DUOLINGO REVIEWS AND CRITIQUES

> ***"Right now, I don't think Max has enough to justify the current price point. Most users will be better off either upgrading to Super, or sticking with the free plan and spending their money elsewhere. Max isn't going replace human tutors any time soon, and, by itself, certainly isn't going to make you fluent."***
>
> *Duoplanet review of Duolingo Max AI-powered features, August 2023*

Duoplanet's review of RolePlay and Explain My Answer suggests that the new AI features are helpful for those seeking to learn French or Spanish (the two languages for which they are available) and that they surpass "the standard Duolingo experience" which "can sometimes leave you wanting something a little less gamey, and a little more authentic." In terms of value for the subscription cost, the reviewer remains unconvinced that the ChatGPT-powered features are worth the higher subscription cost to access them. Critiques of these tools suggest that while they are promising, they remain fairly limited in terms of interactivity.

CRITIQUES OF DUOLINGO AI TOOLS

Quotations derive from Duoplanet, August 2023.

"RolePlay is great. I'm sure it's going to be a massive help for a lot of users. But make no mistake – it's not a replacement for human interaction. ***As flexible as it already is, it can still feel a little cold and predictable, especially after repeating a scenario a couple of times.*** *Even if you offer a really fleshed-out answer, the character can still respond as if following a script. It could do with loosening up a bit."*

"Explain My Answer, although being a welcome addition, only seems to explain one part of the answer — and there's currently no way of selecting which part. ***Although it takes on a chatbot format, you can't actually ask any of your own questions.*** *Instead, you're limited to a set of three responses: yes, no, and show me an example."*

CASE STUDY – UNIVERSITY OF WISCONSIN GUIDING PRINCIPLES

CORE GOAL IS PREPARING STUDENTS FOR A WORLD WHERE AI IS UBIQUITOUS

Like most institutions, University of Wisconsin recognizes the need to prepare students to function in a world where generative AI tools are readily available. The university also assumes that "skills such as prompt engineering, problem-solving, bias detection and intellectual curiosity" will be essential, given the limitations and occasional unreliability of generative AI models. They distill the university's goals with respect to AI down to efforts to help students:

- Gain literacy in AI tools and learn to use them fluently, creatively and ethically
- Develop core competencies in conjunction with AI competencies (e.g. critical thinking, creativity, communication, citizenship, cultural sensitivity, ethics, etc.)
- Build capacity to live and work in tandem with evolving versions of AI

As shown in the figure to the right, AI introduces a range of concerns from academic dishonesty to bias to data privacy and theft of intellectual property. Especially important considerations are shown in boldface orange italic font and include the need to have explicit, recurring conversations about appropriate use of AI within the course and to scaffold writing assignments with ample instructor support to avoid situations where students are more likely to turn to generative AI as a shortcut in the face of academic challenges. Notably, the policy forbids faculty from outsourcing grading, assessment, and feedback to AI but permits its use in shaping feedback for students.

Laura Schmidli, et al., of University of Wisconsin published a guide to using AI in the classroom in early 2023 that includes specific activities designed to help students interface with AI and learn about its strengths and limitations within the context of their discipline. These exercises, which are unusually robust and specific, are reproduced on the next page.

GUIDING PRINCIPLES FOR INSTRUCTORS

Figure summarizes content from the University of Wisconsin generative AI guidelines, 2023.

Advance Accessibility and Equity *– Opportunities include AI assistance for students struggling with writing and entering a new discipline. Challenges include AI's potential to "generate racist, sexist, and other kinds of biased responses and potentially promote such biases," **accessibility barriers for students with disabilities, and unfair advantages for students who can afford access to premium generative AI tools.***

Protect Data and Intellectual Property *– Unless required by the assignment, students should not have to submit drafts of assignments created with AI or input personal information or course documents like prompts into AI tools. **Instructors should not input student work into AI tools to automate feedback and comments.***

Consider Educational Uses *– "Students can **dialogue with AI chat-bots in ways that generate new knowledge and insights** and that mimic conversations with peers or even instructors" and instructors can use AI to help design course materials, assignments, and accessible lessons, and to **inform their feedback on student work.***

Consider Adapting Learning Experiences and Assessments *– **Scaffolding assignments to reduce students' temptation to use AI,** using non-written assessments, focusing on metacognitive strategies like reflections, introducing group projects, and focusing on writing processes are essential recommendations.*

Discuss Course Expectations and Academic Integrity with Students *– **Communicate the expectations for AI usage early and often, with clear examples** of where it can and cannot be used. Avoid using a 'misconduct' approach to suspected uses of AI.*

Address Potential Misconduct Using Established Policies and Processes *– **Instructors should not rely on AI detectors, which have a history of false positive results and bias against non-native English speakers.** When illicit AI usage is suspected, faculty should follow standard university procedures for investigating academic misconduct.*

CASE STUDY – UNIVERSITY OF WISCONSIN AI-IMBUED ACTIVITIES

EXPLORING CAPABILITIES AND LIMITATIONS OF AI IN YOUR CLASSROOM

Figure reproduces content created by Schmidli, et al., University of Wisconsin, 2023. Bolding, formatting, and emphases derive from Hanover.

Ask 20 Questions of AI

- *In small groups, **students collaborate to write 20 questions for a text-generating AI about how it works.***
- *In a larger group, they **consider what the AI's responses mean for academic integrity, authority, validity, trust, or other important ideas** in your course.*

Predict Where AI Excels

- *Individually **students construct one question or prompt on a specific topic** that they think text-generating AI can respond to successfully, **and another prompt** or question they think AI responds to unsuccessfully.*
- *In a larger group, **students share their work to identify characteristics of prompts to which AI struggles to respond.***

Evaluate AI Output

- *The **instructor uses AI to generate work,** like a thesis, short analytical paper, theater dialogue, computer code, image, or even musical composition.*
- *In groups **students analyze the sample work created by AI,** with particular attention to evidence, sources, perceived bias, or other important elements for your course.*
- *Students can then **revise it for improvement** in groups and share back revisions for comparison.*

Analyze AP Perspectives

- *The **instructor asks the text-generating AI to respond to a prompt as a specific person** – e.g., a historical figure.*
- ***Students then critique the AI's response,** drawing on their interpretation of the person's perspective.*

Compare Exam Questions

- *Individually **students write one multiple choice exam question and ask text-generating AI to write a second.***
- *In a larger group, **students analyze which submitted questions are AI-written,** which are human-written, and **evaluate which provide a better assessment of learning.***

INTEGRATING AI INTO THE WRITING PROCESS IN YOUR CLASSROOM

Figure reproduces content created by Schmidli, et al., University of Wisconsin, 2023. Bolding, formatting, and emphases derive from Hanover.

***Rubric Calibration with AI Writing:** **The instructor uses AI to generate an essay, thesis, or other written work. Groups then use a rubric to evaluate the AI's work,** and suggest changes or improvement to the rubric. This can help students think about how they define high-quality work, and how a rubric might help identify AI-generated work.*

***AI Drafting Process:** **Students use AI to generate a draft of a simple writing assignment. Students then analyze the AI's writing,** focusing on accuracy, bias, or other characteristics important in your course. You may also ask students to improve the AI's draft to complete a second draft.*

***AI Thesis Revision:** **Students use AI to quickly generate thesis statements** on a variety of topics. Individually **students revise these statements** and share two that are strongest to a group for feedback, including what prompts they provided the AI and what revisions they made to each statement…*

***Planning and Evaluating AI Use:** **Students create a plan for using AI within a specific assignment,** like a scaffolded research paper, where they articulate for which steps it is valuable and appropriate to use AI, and when original thought and creativity are needed and why. This activity works best when sequenced with other activities that explore AI capabilities.*

***AI Feedback:** **Students ask AI to evaluate an initial draft of a short writing assignment,** asking AI to focus on a specific element. Students then incorporate any valuable feedback into their work, and share their revisions with a small group. This can help students get another perspective on writing quickly, while encouraging them to consider that feedback critically.*

***Citing AI:** **If students are using AI in your class assignments, discuss if and how AI should be cited.** For example, students may cite it as a source or disclose their use of it in a disclaimer, footnote, or appendix that includes the prompts they created…*

***Writing with Images:** Students or instructors **use image-generating AI as part of a reflective writing, freewriting, or creative writing process in any language**. Students can use AI to quickly visualize descriptions from their writing, or students or instructors can use AI to generate images that prompt elaboration in writing.*

CASE STUDY – MONTCLAIR STATE UNIVERSITY AI STRATEGIES

BEST PRACTICES BY STRATEGY TYPE

> ***"Teaching and learning in the era of generative AI does not need to be all about damage control. Plagiarism and ethics concerns are real, but on the other hand, the advance of generative AI gives us a special opportunity to focus on the challenges of teaching and work out strong solutions.***
>
> ***Generative AI itself can be a learning tool – as anyone who gets into the tool and starts inputting queries and studying the output knows. Your synapses are firing as you write and read the rapidly generated text. It's fun, and you're likely wide awake, judging, speculating, disagreeing, agreeing, and doing all those things that happen when an engaged reader encounters text. These potentials can be used in the classroom."***
>
> *Montclair State University Office for Faculty Excellence, "Practical Responses to Generative AI"*

Montclair State University's strategies for teaching in a post-Chat GPT world are divided into three basic categories: assigning tasks that AI cannot yet do convincingly, emphasizing the process of completing assignment over the product, and having students engage with and reflect on their use of ChatGPT. Examples of recommended assignments that can confound generative AI include the use of diverse media (e.g., having students produce videos), assignments that require students to connect the topic with personal experience, assignments that reference texts unavailable to generative AI, and handwritten assignments.

Process-focused approaches include group projects, flipped classrooms with an emphasis on in-class work, and scaffolded assignments with multiple components due as the project develops. Ideas that require students to interact with generative AI include "critical evaluation of AI outputs" and blind evaluations to determine whether AI or a human created a piece of writing, with attention to the features that distinguish AI-produced content.

AI WRITING DETECTION RED FLAGS

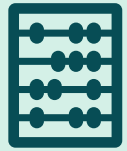
Figure reproduces guidelines from the Montclair State University Office for Faculty Excellence, which cautions that "no software is able to detect AI-generated content with 100% certainty" but provides the following signs of AI-inflected writing.

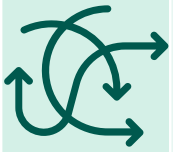
Factual Errors – "Generative AI models work by predicting the next word based on the previous context. They do not 'know' things. Because of that, they tend to output statements that look plausible but are factually incorrect. This phenomenon is known as AI hallucination." Models may also cite made-up sources.

Inconsistency with Assignment Guidelines – "A submission that is AI-generated may not be able to follow the instructions, especially if the assignment asks students to reference specific data and sources. If a submission references data and sources that are unexpected or unusual, that is a red flag."

Atypically Correct in Grammar, Voiceless and Impersonal Style – "It is correct and easy to read, but without any sense of a human person behind it — fallible, uneven, passionate, awkward. You will not be able to see a student behind the writing."

Predictable Style and Structure – "It follows predictable formations: strong topic sentences at the top of paragraphs; summary sentences at the end of paragraphs; even treatment of topics that reads a bit like patter: 'On the one hand, many people believe X is terrible; on the other hand, many people believe X is wonderful.'"





Directionless and Detached Arguments – AI "will shy away from expressing a strong opinion, taking a position on an issue, self-reflecting, or envisioning a future. With the right kind of prompting, it can be coaxed to do some of those things but only to an extent..., as it will continue sounding unnaturally cautious, detached, and empty of direction/content."

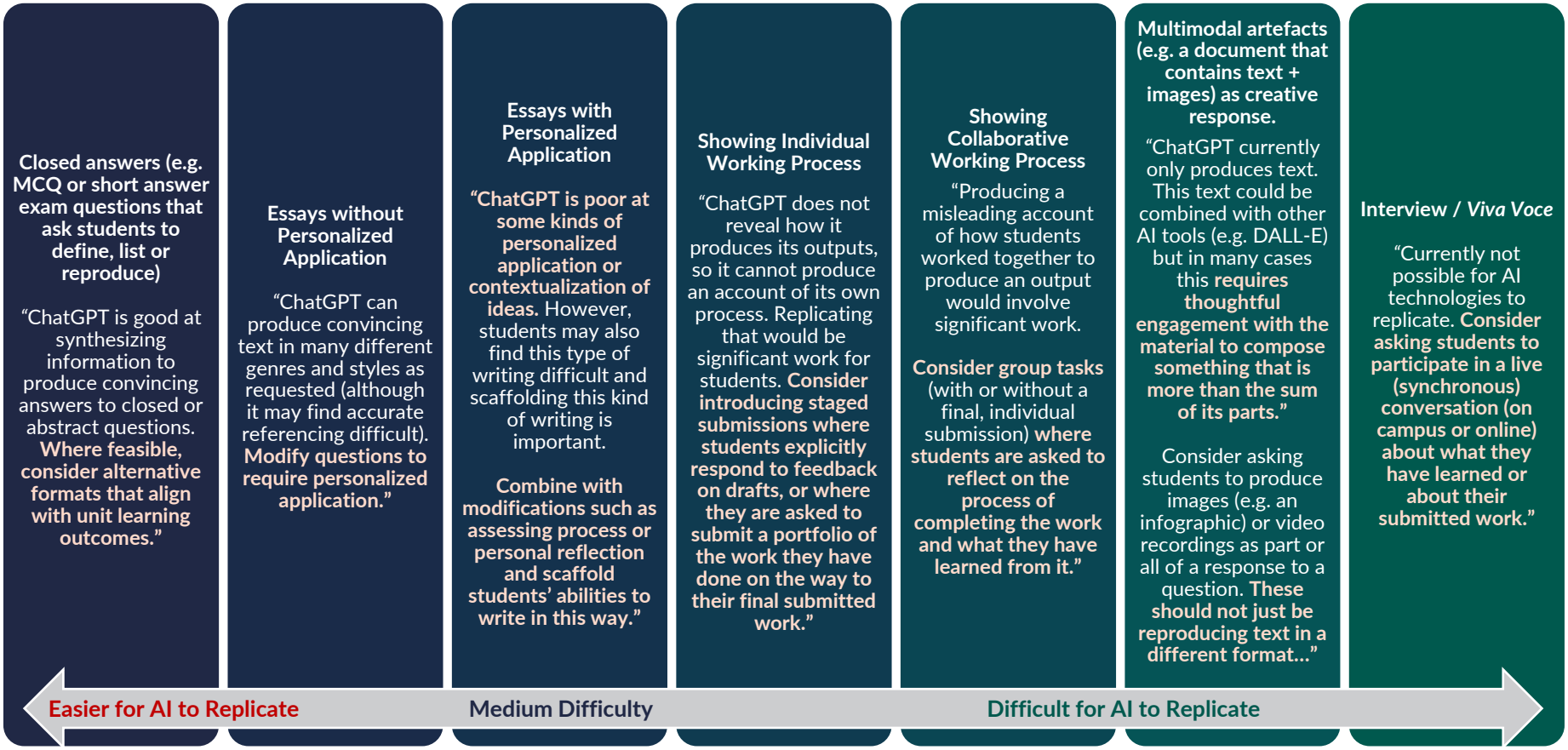
CASE STUDY – MONASH UNIVERSITY ASSESSMENT STRATEGIES

OVERVIEW

Monash University's guide to Generative AI and Assessment urges faculty to "target higher order thinking" when designing assessments, while conceding that limitations on instructor time could make implementing more AI-resistant assignments difficult. The university argues that "it may now be necessary to target forms of knowledge and expression that are more difficult for generative AI technologies - critical thinking, evaluation or creativity, for example."

ASSESSMENT ACTIVITIES BY LEVEL OF DIFFICULTY FOR CHATGPT TO REPLICATE

The three easier activities for generative AI are shown on the left, with more difficult, AI-resistant activities on the right. Table reproduces content from Monash University.



HANOVER
RESEARCH