AI, Academic Integrity, and Authentic Assessment:
An Ethical Path Forward for Education
## Contents

- Introduction ........................................................................................................... 3
- Outlining Ethical Boundaries for AI Plagiarism ......................................................... 4
- Understanding How AI Detectors Work .................................................................... 6
- Anti-Plagiarism Tools vs. AI Detectors: What’s the Difference? ............................... 7
- Anthology’s Testing Supports Research-Identified Issues with AI Detection ............. 9
- A More Ethical and Effective Approach: Empowering Instructors ......................... 12
- An Eye to the Future: What Comes Next for AI, Detection, and Academic Integrity? . 14
- References List ........................................................................................................ 15
Introduction

Legendary basketball player Allen Iverson went by a couple of nicknames across his storied career: “A.I.” and, most famously, “The Answer”. More than a decade after Iverson’s retirement, learners across the globe are turning to AI to find The Answer for common test and essay questions, while instructors and institutions look on in shock as formerly difficult assessment tasks are turned instantly into free throws.

Time out on the basketball puns. The wide availability of generative artificial intelligence (AI) tools for learners has ushered in a new era for academic integrity, with established assessment policies and practices needing urgent revision now that students can generate responses to many tasks with just a couple of clicks. This white paper provides Anthology’s perspective on best practices to maintain academic integrity, outlining what we believe is the most ethical response to the rise of generative AI, and detailing both the possibilities and limitations of learning technology to support the use of AI in higher education.

Academic policies related to plagiarism were devised at a time before humans chose to share intelligence with machines. Consequently, common anti-plagiarism tools were built to compare learners’ work with existing texts, meaning educators always had sources to review and clear evidence to assess the extent of any academic integrity violation. Generative AI, on the other hand, creates original text and media that appear authentic and do not match with any existing work. As a result, tools have been established to detect AI-generated work—and the integration of this functionality into anti-plagiarism software is being widely explored—however, early results have shown lower levels of accuracy than anticipated.

Further, there is an emerging body of research that challenges how accurately generative AI use can be detected by AI detection software (Weber-Wulff et al., 2023; Sadasivan et al., 2023). As well as failing to adequately differentiate between human and AI authorship, studies have raised concerns that AI detection is biased against non-English speakers (Liang et al., 2023). At Anthology, we have conducted a thorough beta test in collaboration with market-leading AI detection tools, from which both Anthology and participating clients concluded that AI detection is not currently fit for purpose in education.

The ethical adoption of AI requires the empowerment of instructors. Rather than hoping that AI detectors score an unlikely victory in an arms race against generative AI tools, institutions should focus on providing instructors the support and resources they need to embrace authentic assessment practices. Learning technologies—including those powered by AI—should help to facilitate these kinds of assessments, as well as provide efficiency for educators by reducing time spent on administrative and production tasks. With their focus on the application of skills rather than the accrual of knowledge, authentic assessments elicit the nuances, experiences, and even fallibilities that define human beings, but evade generative AI.

Such as, for example, including a gratuitous sporting reference in the introduction of a white paper on artificial intelligence!
Outlining Ethical Boundaries for AI Plagiarism

Before we look at AI plagiarism specifically, let us take a step back and review plagiarism as a broader concept. The Oxford Learner’s Dictionary defines plagiarism as “the practice of copying another person’s ideas, words, or work and pretending that they are your own.”

This, in turn, has shaped institutions’ own definitions for academic practices. Staying at Oxford, for example, their university policy begins “[Plagiarism is] presenting work or ideas from another source as your own, with or without consent of the original author, by incorporating it into your work without full acknowledgement.” Across the Atlantic at Penn State University, their definition is essentially identical: “Plagiarism is defined as the appropriation of another person’s ideas, processes, results, or words without giving appropriate credit.”

An obvious shortcoming emerges: these definitions were written when the only source intelligent enough to warrant copying was another human being. This brings us to artificial intelligence, or AI. But before we put the dictionary away, let us look at a few of the related terms we will be using throughout this paper:

- **Artificial intelligence (AI)** is defined as the development of computer systems that can copy intelligent human behavior.

- **Generative AI**, more specifically, refers to a particular subset of AI technologies that can create original content, including text and images.

- **AI plagiarism** is the process of using generative AI to produce content that students submit as their own work for assessment tasks. This is the primary focus of this paper, as well as the efficacy of AI detection software that has been developed to try and identify where AI plagiarism occurs.

Now that generative AI technologies are widely available to learners and AI plagiarism is growing as a consequence, clearly an ethical approach to plagiarism must be reconsidered. Many institutions are updating their definitions and policies to emphasize originality, including phrases like “the presentation of work that is not your own” to prohibit student submissions copied directly from either human or AI sources.

To step away from definitions and into the classroom, however, is to acknowledge that generative AI tools that create seemingly original work present a far more complex challenge for academic integrity. Consider a learner writing an essay: on one hand, it was not long ago that it would have been questionable behavior for them to leverage AI spelling and grammar tools to improve their submission, yet now this is widely considered acceptable for most tasks. At the other extreme, clearly it is a breach of academic integrity for the learner to copy an answer directly from ChatGPT and submit it as their own, but where does the line exist between these two applications of AI? Which technologies are helpful or even empowering for learners, and which should we categorize as cheating?

As higher education looks to grapple with these questions, often generative AI technologies are advancing faster than institutional policies can account for. A recent UNESCO study found less than 10% of universities globally have a formalized AI policy, leaving their instructors and learners looking for more...
guidance on what is classified as responsible practice.

To inform leaders, Cecilia Ka Yuk Chan, head of professional development at the University of Hong Kong, recently published a global study on best practices for institutional AI policies. The study found that encapsulating all the potential use cases for AI in higher education in a single policy was too complex, and instead recommended separate policies for Pedagogy, Governance, and Operational.

Teachers are the initiator for the Pedagogical Dimension. Teachers need to foster ethical use of AI, for example through proper attribution, to acknowledge the contributions of AI technologies in student work, and develop assessment tasks that require critical and analytical thinking to avoid AI-assisted plagiarism.

Cecilia Ka Yuk Chan¹, Head of Professional Development at the University of Hong Kong

At Anthology, we agree with Chan that empowering instructors is an essential guiding principle for best pedagogical practice in the age of generative AI, including the task of combatting AI plagiarism. A second crucial consideration, as detailed in our Anthology Trustworthy AI Approach¹⁹, is fairness: as academic integrity policies are designed to create a level playing field for all students, it’s important that attempts to tackle AI plagiarism don’t create new biases or practices that disadvantage some demographics more than others.
Understanding How AI Detectors Work

Given the threat that generative AI represents to academic integrity, it is unsurprising that a lot of attention has turned to software programs that aid in its detection. Several software vendors have released AI detectors, all of which use—you guessed it—AI capabilities to deduce which learner submissions are theirs and which are created by generative AI technology.

This raises an immediate red flag. With the huge investment that’s going into generative AI technologies and the rate at which their output is evolving and improving, it’s going to be incredibly difficult for AI detectors to match their speed of development. “What’s so concerning now is that [ChatGPT has] really impressive outputs. Detection models just can’t keep up. You’re playing catch-up the whole time,” notes Muhammed Adbul-Mageed from the University of British Columbia in the *MIT Technology Review* (Heikkilä, 2022).

While the options vary, all AI detectors operate in a similar manner:

1. **Data Collection and Learning:** AI gathers a diverse range of writing samples, including both human and AI-generated texts. Through this extensive dataset, they learn to recognize patterns that differentiate AI writing from human writing.

2. **Training the Algorithms:** Using this data, the algorithms undergo rigorous training to effectively identify the distinct characteristics of AI-generated texts. As noted above, a key challenge here is that those characteristics are constantly changing.

3. **Analyzing New Texts:** When presented with a new piece of writing, the AI detection tool compares it to the knowledge acquired during training. It examines the structure, word usage, grammar, and other features to determine the likelihood of human or AI authorship.

4. **“Perplexity” and “Burstiness” Scoring:** These are further measures to assess a text’s authenticity. “Perplexity” refers to the complexity and coherence of the text, while “burstiness” evaluates repetition of words and phrases that might indicate that it was not compiled by a human.

5. **Presentation of Results:** After analyzing the text, the tool generates a report showing the likelihood of AI involvement in the writing. This includes both an overall result as well as analysis of a range of factors which may indicate that the text was AI-generated.
Anti-Plagiarism Tools vs. AI Detectors: What’s the Difference?

Traditional anti-plagiarism tools have been widely used by educational institutions to identify academic dishonesty. Within our flagship learning management system, Blackboard® Learn, our clients have access to our proprietary anti-plagiarism tool, SafeAssign², as well as native integrations with tools from major anti-plagiarism providers.

With AI now at the forefront of the discussion around plagiarism, institutions are reasonably questioning whether traditional anti-plagiarism tools can expand to protect against AI plagiarism as well. Before elaborating on the viability of this approach, let us briefly review how anti-plagiarism tools work, using SafeAssign as an example:

1. **Document Submission**: Learners submit their assignments or essays through the educational platform integrated with SafeAssign.

2. **Database Comparison**: SafeAssign has an extensive database containing academic papers, publications, articles, and internet sources.

3. **Text Analysis**: The tool analyzes the submitted document and breaks it down into smaller sections for comparison.

4. **Search for Similarities**: SafeAssign then searches its database and the internet to find matching content for each section of the document.

5. **Scoring and Highlighting**: Based on the identified matches, SafeAssign generates an overall probability score for the submitted document. It also highlights the specific sections that match existing sources.

6. **Originality Report**: Instructors receive a detailed originality report, which presents the probability of the document being plagiarized, along with the sources where similarities were identified.
Of particular importance are points two and four. The traditional anti-plagiarism model is based on building a large database of existing texts to compare with the learner’s submission, rather than on the ability to protect against texts created by generative AI.

Providers of anti-plagiarism tools are now assessing how effectively AI detection can be included in their solutions. Early results have showed these integrations have lower levels of accuracy than anticipated, causing some major players to discontinue their development of AI detection tools.
Anthology’s Testing Supports Research-Identified Issues with AI Detection

As AI detection tools become available to institutions, so too does a growing body of research casting doubt on their efficacy. This starts with concerns on the accuracy of AI detection; for example, five computer scientists from the University of Maryland conducted a detailed study titled “Can AI-Generated Text Be Reliably Detected?”, and concluded emphatically that it cannot, with simple paraphrasing sufficient to evade detection (Sadasivan et al., 2023). A separate study of 14 AI detectors conducted by academic researchers in six countries found the accuracy of these tools ranges from just 33 to 81 percent depending on provider and methodology (Weber-Wulff et al., 2023).

Another emerging concern relates to bias. As generative AI tools are most effective in English due to the predominance of English texts on the internet, this inherent bias can lead to inaccurate identification of AI-generated content in other languages or from non-English speakers. A study conducted at Stanford, to provide just one example, found that more than half of authored TOEFL (Test of English as a Foreign Language) essays were incorrectly classified as AI-generated. “Our findings reveal that these detectors consistently misclassify non-native English writing samples as AI-generated, whereas native writing samples are accurately identified,” the researchers concluded (Liang et al., 2023).

At Anthology, inclusion and accessibility are central pillars for the development of all our solutions. As part of an in-depth process to evaluate the inclusion of AI detection within SafeAssign, we have recently conducted a detailed beta test in collaboration with market leading AI detection solutions to understand the extent to which these fears around AI detectors were replicated when applied within our solutions. The test took place across May and June 2023, and involved 65 of our client partners submitting more than 1000 texts—some authentic, others generated by AI—and then responding to a survey on how accurately these documents were assessed by the AI detection tool.

The results of the test supported the conclusions from broader research on the weaknesses of AI detection.
Accuracy

The participants left the test with very low confidence in the ability of AI detectors to distinguish between AI and human generated texts. 80% of respondents felt that detectors were—at best—only able to “sometimes” identify texts correctly.

If you look at your overall experience with this beta test, how often do you think the AI detection tool correctly identified whether text was written by AI?

This sentiment came through in anecdotal feedback, also. Responses included:

- “I used ChatGPT to generate five different texts. Four introduction paragraphs and two entire essays, and the AI detector completely misidentified all of them as being human generated.”
- “The AI detector completely failed to identify any of the five AI generated texts I submitted as being AI generated.”
- “It is easy to manipulate the tool into thinking the text was written entirely by humans.”

The final comment raises a crucial point and supports the research of Sadasivan et al. (2023) mentioned above. Even if AI detection tools develop to the point where they can accurately identify work that was generated artificially, the challenge of students making basic modifications to avoid detection still remains.
Bias

Participants were also asked to submit authentic work from a range of different learner demographics so that any biases could be identified. Here again the results from broader research were supported, with respondents often finding a lower degree of accuracy for English as a second language (ESL) students. Furthermore, multiple responses suggested that AI detection was incorrectly marking the work of learners with autism spectrum disorder (ASD) as plagiarized, raising another potential issue that warrants further investigation. Responses included:

“ESL and ASD student work more likely to be flagged than actual AI.”

“Work from individuals with ASD and non-native English speakers was more likely to be flagged than actual AI generated content.”

“After observing a curious pattern, I intentionally submitted a series of responses from writers with known autism spectrum disorder. For each of these submissions, the likelihood of AI was indicated to be stronger although I can confirm with absolute certainty they were written by humans.”

Other possible issues with AI detection, which have surfaced both in our beta test and elsewhere, include confusion around terminology and metrics, inadequacy of detection for tasks with highly predictable responses, easy manipulation of AI-generated content, and more.

What is clear, however, is that institutions cannot currently say with any degree of certainty that results from AI detection are accurate and non-biased—and nothing is more important for academic integrity than ensuring a learner’s work is assessed accurately and without prejudice. To employ AI detection tools at their current point of development is to risk incorrect allegations of AI plagiarism, likely directed at certain learner demographics more commonly than others, and we don’t believe that this represents an ethical approach. As a respondent in our beta test noted, “I would not be comfortable using this tool on real student submissions at this time.”
A More Ethical and Effective Approach: Empowering Instructors

We believe there is a better approach to tackling AI’s threat to academic integrity than pushing all the chips in on AI detection software. At Anthology, we feel strongly that empowering instructors is not only the best defense against plagiarism, but also provides institutions the opportunity to make AI-powered technologies beneficial to education, rather than detrimental.

Crucial in the defense against plagiarism, as outlined in the work of Chan (2023) referenced earlier in the paper, is the adoption of authentic assessment. As defined by prominent academic and assessment researcher Grant Wiggins, who coined the phrase in the 1990s, “authentic assessments are representative challenges within a given discipline. They are designed to emphasize realistic (but fair) complexity; they stress depth more than breadth.” In its simplest form, authentic assessment moves away from accrued knowledge to focus on the practical application of skills, prioritizing complex tasks over binary right and wrong questions, and parting with many of the traditional staples of assessment, such as time limits, single-submissions, and grading curves. For example, a business course looking to teach negotiation skills might consider replacing a traditional exam task with a live interview, wherein the learner needs to apply the theory presented in the class to achieve a desired business outcome.

Even prior to the arrival of generative AI, authentic assessment practices have been shown to have a number of benefits, including making learning more enjoyable and facilitating the path from education to the workforce across a range of fields. Anderson et al. (2022) researched the use of authentic assessment in health promotion courses and found it an effective way to evaluate both practical and ethical skills. In a separate study, Ozan (2019) looked at the use of authentic assessment in teaching degrees and concluded that it is “an approach that can serve to provide cooperation between the theory and practice which is a major problem in the field of teacher training.”

Authentic assessment takes on even more importance in the era of generative AI. Put simply, tasks that focus on critical thinking, personal perspectives, and self-reflection are much harder for generative AI technologies to produce in a way that appears genuine. Activities might also look to explore subject areas where these tools do not have as much historical data with which to work, such as current and local events, personal experiences, and future predictions.

ChatGPT has sparked a whole new level of awareness and proactivity from our faculty with regards to authentic assessment. Knowing that students can easily generate responses to standard questions and tasks has placed greater emphasis on looking for original ways to test different subject areas. It might even be that AI ends up being a net-positive for assessment in higher ed.

Jacob Spradlin, Director of Online Instructional Development and Support, Sam Houston State University
It is essential that learning technology facilitates instructors’ adoption of authentic assessment practices. Our learning management system, Blackboard Learn, empowers teachers to do just that in many ways:

- The recently announced AI Design Assistant, developed in partnership with Microsoft, uses generative AI to provide suggestions on course structure, test questions, and even grading rubrics when instructors are building their courses. By streamlining some of these time-consuming tasks, instructors have more time to focus on authentic assessments and other ways to make their courses engaging and beneficial. The same applies for instructional designers, who are often caught up in production tasks rather than higher-level strategic development. It is essential that in the discussion around the risks of AI we don’t lose sight of the potential benefits that it can hold for the efficiency of institutional processes, including those of instructors and instructional designers.

- Peer assessment and group work are also great ways of adding authenticity to tasks. Fostering collaboration within student groups not only allows them to learn from their colleagues, it also creates a culture of shared responsibility which makes learners less likely to consider unethical practices. Blackboard Learn comes with all the necessary functionality to facilitate group tasks, collaboration via discussion, and peer assessments, helping instructors to apply the principles of authentic assessments within online and hybrid learning modalities.

- Personalized assessments are also essential. Every learner is different, and assessments that allow them to express their ideas and perspectives will make them feel included while requiring an output that is difficult for AI to replicate. The other benefit for instructors in applying personalized tasks—for instance, regular reflection as part of a group discussion within the LMS—is that it allows them to observe each learner’s writing style, making it easier to spot any AI-generated work that the student may submit at a later time.

Beyond the adoption of authentic assessments, empowering instructors also requires detailed training around AI technologies and—as mentioned in this paper’s first section—there is currently a lack of guidance available to instructors on responsible use of generative AI. As recently reported in Inside Higher Ed, a survey of instructors across the United States found that only 14% had been provided guidelines for use of AI in the classroom—and, consequently, only 18% had set guidelines for their students (D'Agostino, 2023). Naturally, this causes uncertainty among both instructors and learners, and academic integrity requires a shared understanding of what is and is not ethical across all stakeholders.

My experience has been that faculty come to us usually from one of three perspectives. A ‘fight it’ perspective where they’re really concerned, with good reason, about what happens to academic dishonesty with the entry of easily accessed AI tools. Or maybe they come to us with the opposite ‘use it’ perspective and they’re ready to jump in. Or maybe they’re in the open-minded middle, where they’re watching to see what happens. [...] We have this new paradigm that we all have to learn about [AI] and figure out what to do with it, and where we enter into the conversation.

Suzanne Tapp, Assistant Vice Provost of Faculty Success and Executive Director of the Teaching, Learning, and Professional Development Center at Texas Tech University.

Learn more from Susan and other leaders in this recent Anthology webinar.
An Eye to the Future: What Comes Next for AI, Detection, and Academic Integrity?

With the amount of change that we have seen in just the last few months, it would take a brave person to claim to know exactly what will happen next. When it comes to AI, an old adage applies: change will be the only constant.

It is not our expectation that AI detection will reach the required level of accuracy for ethical use any time soon. The eye-watering investment that major tech companies are making in generative AI—including Microsoft, our partner in developing the AI Design Assistant within Blackboard Learn, who have announced a $10 billion investment—means that the chances of AI detectors being able to keep pace are very small. Nonetheless, we will continue to monitor this closely and consider its application within our learning solutions based on the latest evidence.

This paper has touched on four principles that we believe can shape an ethical approach to AI for institutions. Once again, these are:

1. **AI is here to stay – embrace it**: Do not get caught up in the negatives. AI is going to provide opportunities for all industries, education included, and institutions need to focus on maximizing the benefits across campus, not just restricting its threats.

2. **It will bring about significant change**: The initial conversations around AI in education have often elicited comparisons with previous technologies that changed the learning experience, the calculator being one example. These are apt in that they reflect the sentiment in point one above—we have handled change before and will do so again—but risk understating the significance of generative AI. A fundamental re-thinking of academic integrity, and indeed many other parts of the learner’s journey, is essential for success.

3. **Flexible policy and practices will be required**: Academic integrity, and related terms like plagiarism, have always been very clearly defined, with a shared understanding of what is considered ethical practice. AI changes this. It will no longer be possible to maintain strict, institution-wide policies, as the impact of AI will change for different study areas. While having clear guidelines is essential, instructors should also have autonomy to apply these relative to their subject matter. Co-creation with students is also a worthy approach and helps to create a culture of trust across the institution.

4. **Empowering instructors is central to an ethical approach**: Teachers are, ultimately, the gatekeepers of education, and supporting them is fundamental to providing great experiences for learners. In the era of AI, this means that instructors need to be empowered—by both their institution and technology—to embrace authentic assessment practices, including leveraging the power of AI to make administrative and production tasks more efficient.

At Anthology, we are committed to the responsible adoption of AI within our technologies. Based on guidance received from our global learning community, we have published our **Trustworthy AI Approach** to provide full clarity on how we are approaching the legal, responsible, and ethical use of AI as a company, including in relation to product development. We are excited to continue our close collaboration with institutions around the world to map out an ethical and effective path forward for education.
References List


10.) “GPT detectors are biased against non-native English writers,” by Liang, Yuksekgonul, Mao, Wu, and Zou, 2023, p. 6.


16.) “Plagiarism,” University of Oxford, 2023, https://www.ox.ac.uk/students/academic/guidance/skills/plagiarism


23.) “What is Grammarly, and is it cheating?” by David Tomar, Academic Influence, 2023, https://academicinfluence.com/inflection/study-guides/what-is-grammar

About Anthology

Anthology offers the largest EdTech ecosystem on a global scale, supporting over 150 million users in 80 countries. The company’s mission is to provide dynamic, data-informed experiences to the global education community so that learners and educators can achieve their goals.

Through Anthology Intelligent Experiences™ and over 60 SaaS products and services, Anthology advances learning in partnership with education, business and government institutions. Tapping into this unmatched portfolio of solutions, only Anthology can leverage data from across the EdTech ecosystem to create Intelligent Experiences that lead to better outcomes.

Learn more about our mission at www.anthology.com.

©2023 Anthology Inc. and its affiliates. All rights reserved.