AI Detectors' Bias Against English Language Learners (ELL)

Policy Mitigation using the General AI Detector Assistance (GAIDA) Policy

Thomas Zucker-Scharff

BOSTON UNIVERSITY – RESPONSIBLE AND ETHICAL DATA SCIENCE

Current AI-detectors systematically disadvantage English-Language-Learners (ELL), violating Title IV civil-rights protections; therefore, institutions must adopt a policy that couples detector use with mandatory human review, transparency, and an appeals process. Currently, there are few, if any, policies specifically addressing ELL and how instructors should utilize AI detectors in these cases—other than the guideline that AI detectors should not be the sole basis for determining if a student's work has been produced with AI assistance (Dang & Wang, 2024; Najarro, 2023).

One does not need to look hard to find mentions of using AI detectors with ELL student writing in education forums. Many repeated issues highlight potential problems that can happen if there is even one false positive, especially since false positives are more common with ELL learners. (D'Agostino, 2023; Woelfel, 2023) As the Hirsch reference describes, AI detectors are an "ethical minefield." It is morally and ethically incumbent upon the instructors working with ELL students to find ways to complete tasks without putting their students at risk. This policy addresses that need and fills these gaps.

From a deontological perspective, the institution has a duty to treat all students as equals according to the 14th Amendment of the US Constitution, which accords equal protection under the law, and to uphold their civil-rights protections under Title IV of the Civil Rights Act. (*The 14th Amendment and the Evolution of Title IX*, 2025; "Student rights in U.S. higher education," 2025). FERPA, although primarily about privacy, obligates institutions to be sure that their records are accurate and fairly administered; this would include "integrity" determinations (*34 CFR Part 99 -- Family Educational Rights and Privacy*, 2025; *FERPA | Protecting Student Privacy*, 2011). Any policy that relies solely on a tool that systematically disadvantages a protected group violates that duty, regardless of the tool's overall accuracy. Consequently, the

use of AI-detectors must be constrained by procedural safeguards (human review, transparent error rates, and an appeals process) that respect the inherent rights of ELL learners. Without these safeguards, the institution violates the deontological and legal obligation of equal treatment.

Although the primary approach involves the nature of the institution and its obligations, one could also argue that, from a secondary consequentialist perspective, using AI-detectors without safeguards creates measurable harms — such as inflated false-positive counts (Wood, 2024), wasted faculty time (Alsharefeen, 2025), and potential legal liability for the institution (McLogan, 2025). These possible harms outweigh what may be the ultimate purpose of AI detectors, academic integrity.

The very basis by which one would justify the use of AI-detectors, *to ensure the integrity of any written work and to help students achieve their best possible work*, is the reason why one should **not** use AI-detectors. When using these detectors, users are more likely to misidentify ELL-generated human text as AI-generated (a false positive). The results of such a false-positive could be disastrous to the student. The student will, in turn, not be doing their best possible work, but instead defending work they have already done. Therefore, in a pure ethical/logical sense, one should not use AI-detectors alone when evaluating ELL students' writing ("Academic Integrity," 2025; Cornell University, 2023; Dang & Wang, 2024). A way to deal with this problem directly is to enact an optional policy.

The question of why one would use AI-detection tools might be a key issue. Generally, AI-Detection tools, when used with ELL students, cause more problems than they fix(Check, 2025; Giray, 2024). Since evidence shows that AI detectors will identify more ELL students' writing as false positives compared to native English speakers' writing (Mathewson, 2023;

Myers, 2023), it would seem ethically and morally prudent to use such tools only in extreme cases. Instructors would be better off following some basic principles, such as using baseline sampling (Cambridge, 2025; Kratzer, 2020), initial review scanning ("Assessing Student Writing," 2025), and informing students in advance about GenAI and AI-detection policies. These practices will put any instructor on a better footing when making decisions about AI usage.

One of the many issues that needs to be addressed is the legal concern related to students' use of AI when AI-detectors are **not** employed. (Weaver, 2025) Although AI-generated text may contain errors (Bohannon, 2023; Marcus, 2022; Weaver, 2023a, 2023b) if one has access to the necessary tools and is aware of the problem, likely, the finished product will no longer contain those AI-generated errors or "hallucinations". As Furze pointed out(Furze, 2024), AI tools tend to favor native English speakers with higher incomes. This means students who are ELL and lack easy access to computers and/or AI tools are less likely to perform well on written academic submissions. Conversely, students *with* such access will be able to use AI and AI humanizers to produce submissions, which may receive higher grades and be less likely to be identified as AI-generated.

This issue emphasizes the need to establish something like the General AI-Detector Assistance (GAIDA) policy which would be applicable to all students to ensure that AI tools do not become another avenue for discrimination based on socio-economic status.

Several institutions do have policy examples, for instructors to use on a per-course basis, regarding students' use of generative AI (BU, 2023; CELT, 2025; Cornell University, 2023; Palmer-Clarke, 2025; Vanderbilt_Univ, 2023). There are existing or complete policies on

students' use of generative AI. However, they do not specifically address use by ELL or English as a Second Language (ESL) students.

AI detector tools are being used in many higher education classrooms despite increasing evidence that they produce disproportionately high false-positive rates for English Language Learner (ELL) students, as shown by the Stanford study (Furze, 2024; Mathewson, 2023; Myers, 2023; Najarro, 2023), which reported up to a 61% false-positive rate for ELL essays. This exposed non-native writers to unnecessary academic integrity investigations, emotional distress, and—in extreme cases—visa status risks that could lead to deportation (Castellanos-Canales, 2025; University_of_Rochester, 2025).

Many academic institutions currently advise instructors NOT to use AI detectors (Univ_Pittsburgh, 2023; Vanderbilt_Univ, 2023), due to the high number of false positives they have experienced. Some provide general guidelines on students' AI use and offer templates for creating AI usage or detector policies for each class. ("Academic Integrity," 2025; Cornell University, 2023)[1]. While this approach has been the best so far at addressing most legal concerns about AI detector use, including those related to civil rights, it has not addressed the problem directly.

Civil rights issues are specifically addressed because if a false positive disproportionately affects ELLs, then Title IV of the Civil Rights Act concerning "disparate treatment," "disparate impact," or "hostile learning environment" could be invoked. This could have more serious implications than initially expected. (Markey; Orr, 2022).

---

[1] Sample AI policies

In some instances, AI detection tools may be deemed to be useful. Instructors should follow these guidelines when selecting the best tools to minimize the likelihood of falsely identifying text written by a human as AI-generated. (Mahmood, 2023)

A few of the possible objections and critique can be easily summarized and an example of how the policy addresses them may be instructive. I have created a table below which shows what I believe to be possible critique and how this policy addresses them.[2]

Potential Objection/Critique	Response by Plan
Increased Faculty workload	• **human-review workflow pilot** • provide **training modules** • **compensation**
Potential reintroduction of bias due to human review	• **bias-training** • **standardized rubrics**
Impact on visa status for international students	• **T**ransparent **appeals** process • **Legal counsel** involvement
Technical feasibility / cost	• **Cost-benefit analysis**
Risk of over-reliance on detectors	• **AI detectors will be only one indicator among many**

Use extreme caution when employing AI detection tools ("Employability-Survey-Report," 2024). Instructors should refer to the latest version of Bloom's taxonomy, which has been updated to include AI (Jones, 2025). A detection score should be just one of many indicators used to assess whether a student's written work has been assisted by AI.

Supporting evidence can be found in the seminal study from Stanford (Myers, 2023; Tian, 2023; Woelfel et al., 2023). The study tested seven commercial AI detectors using two datasets: (a) essays written by native-born U.S. students, and (b) TOEFL essays written by non-native speakers. The findings showed that non-native writers are significantly over-represented among false positives, leading to unnecessary academic integrity accusations.

While the AI detectors performed nearly flawlessly on essays written by native speakers, 61% of the 91 TOEFL essays were incorrectly flagged as AI-generated. All seven detectors

---

[2] How to choose an AI detector suggestions

unanimously misclassified 18 of the essays (19%). At least one detector flagged 97% of the

TOEFL essays as AI-generated. More evidence can be found in case studies from a

ResearchGate book chapter (2023) – AI Detection's High False Positive Rates and the

Psychological and Material Impacts on Students (Hirsch, 2024). This study examined the

problem of false accusations by conducting interviews with students focusing on neurodivergent

and ELL writers. ELL writers reported greater anxiety, stigma, and extra administrative burden

after a false positive ID. The book also notes that "neurodivergent writers, along with [ELL]

writers, are flagged at higher rates." A Markup investigative article, "AI Detection Tools Falsely

Accuse International Students of Cheating" (Mathewson, 2023), which consisted of interviews

with faculty and students from several U.S. universities, plus replication of the Stanford

experiment, came to the same conclusions. International/ESL students reported being singled out

after detectors flagged their work. The article mentions the 61% false-positive rate for non-native

essays and highlights real-world disciplinary cases (e.g., a student placed on academic probation

after a false flag).

 Although a policy created and revised by those involved is probably the best solution,

alternative options include using technology to reduce AI detector bias toward ESL students

(Pangram Labs AI Detection, 2025; Tian, 2023) and actions similar to Vanderbilt University's

(Coley, 2023) decision to disable the Turnitin AI detection module campus-wide. Although one

could advocate a more technical solution, this begs the question of how accurate such a solution

would be (if one could even procure the mostly proprietary algorithms to evaluate).

 Because of the reasons previously stated, institutions should implement safeguards in a

policy statement, including manual review, a transparent appeals process, and increased educator

awareness, to minimize these risks. Using AI-detectors voluntarily without a policy may cause

unforeseen problems. In a survey by the Center for Democracy & Technology (Prothero, 2024), 68% of instructors reported using an AI detector on student work without a policy. To prevent unexpected outcomes, establishing a policy to regulate AI-detectors is crucial.

The policy addresses behavioral changes by the users of AI detectors, and specific changes in institutional procedures that will contribute to better outcomes for ELL students. The need to change how instructors of ELL students process their students' work can be succinctly stated as the fundamental tenet that *any final judgment must always involve human review*. AI-Detection may be best suited to highlight those items that require closer examination.

Although the policy mostly applies to ELL instructors, the implementation will involve many stakeholders (see the stakeholders and timeline sections). It is recommended for use in higher education institutions as a starting point. This must be implemented as a "top-down" approach to the use of AI detectors with ELL.

Because AI detectors tend to be biased against ELL students, schools and universities must not rely **solely** on these tools for high-stakes decisions involving ELLs, such as academic integrity violations, admissions, or grading. To avoid confusion about AI detector use, institutions need **clear guidelines**, including mandatory human review and an accessible, transparent appeal process, especially when ELLs are involved. They should also implement basic transparency measures, such as informing students when AI detection is used and sharing information about error rates and known biases. The GAIDA policy advocates for civil rights protections for ELLs, including banning denial of enrollment, visa revocations, or punitive measures based solely on AI detector results. Educators should evaluate student work **holistically** and provide support that is culturally and linguistically responsive, rather than resorting to punitive actions if AI use is suspected.

It is significant and extremely important that these measures indicate that AI detectors should not be the sole agent by which any action or decision is made regarding student work. This is to acknowledge the realistic use of AI detectors in the framework of human review of any work.

How can institutions enforce any of the requirements put forth herein? Some instituions may be bound by local laws (*AB 2013- CHAPTERED*, 2024), While others are bound by federal laws such as FERPA. Enforcing current policies includes faculty hearings, honor codes (for both faculty and students), and AI usage boards meant to determine whether an accusation is true or false. Some institutions have found that the best way to enforce AI usage is by asking students to adhere to academic integrity policies ("Code of Academic Integrity," 2025). If these policies are not followed, the result can be as serious as expulsion. This decision, along with such a result, may also lead to the student being deported (due to loss of visa status). This policy recommends keeping these enforcement policies but applying them to the current policy.

The policy implementation timeline[3] aims to ensure this is well-received and avoids conflicts with stakeholders. The timeline is divided into *"Phases"*, each of which is iterative to ensure that every stakeholder involved in that phase understands the implications of the policy and feels comfortable progressing to the next phase. The suggested *"Duration"* is a recommendation only. However, each time frame is intended as a guideline for that specific phase. The *"Core goal"* section of the timeline describes the desired outcomes for this phase. Producing the best language for the next step may be challenging, but the eventual result – the deliverable, will be more than worth it. The *"Key Milestones & Deliverables"* column contains suggested items that should be the result of that particular phase. The *"Owners"* section assigns

---

[3] Policy implementation suggested timeline

those people who will have primary ownership of this step in the process. This is essential to make hand-offs from one phase to the next as transparent as possible while making it obvious how to escalate if a deadline is missed. The "*Success Metric*" section contains suggested metrics which can be implemented to measure that particular phase's success. This is also a criterion to meet in order to proceed to the next phase. The total implementation period is recommended to span one (1) academic year and be approximately divided in line with the suggested timeline. This will allow for both a gradual adoption of the policy and, if started at the beginning of an academic year, will involve a new group of students who may be more amenable to change than an existing cohort.

The various stakeholder categories (**stakeholder**) must be evaluated as to how much effect they can have on the implementation of this policy (**power**) and how much interest (**interest**) they may have in the implementation and import of the policy. These will inform the actions advised for that stakeholder category (**action items**). One can create a Power/interest stakeholder matrix from this information. This will also indicate which **quadrant** of a grid each category should lie in (ranging from high power/high interest to low power/low interest).[4]

Stakeholders include, but are not limited to 1) International students and English language learners, who are mostly the victims in these scenarios. These students are adversely affected when AI detectors are used. False negatives for these students may result in penalties that could lead to deportation. If enacted properly, the policy would involve these stakeholders in how AI detectors are used, involve them in the solution, keep them abreast of developments along the way, and decrease the number of false positives associated with AI detector use. 2) Instructors at institutions of higher education who need to be trained to identify and rectify these

---

[4] Power-Interest Stakeholder Matrix

false positives. Although they are currently overworked due to the number of AI detector false positives, this would not only ease that burden but keep them in the loop in determining how and when to use such tools. By helping to implement and use the proposed policy, instructors should encounter fewer false positives. 3) Companies that create algorithms to detect AI-generated text need to understand why these detectors are biased, to start to create a non-biased technical solution. The policy will help to identify inherent biases in current models. This will make it more likely that, once these biases are remedied, institutions will resume using AI detectors, though in a way that makes the tools fairer for all students. As part of this policy and to encourage vendors to become more involved, data collected during these trials will be anonymized and made available to them upon request. 4) Administrative personnel—Instructor supervisors are responsible for overseeing professors/instructors. They must understand the relevant processes to address any issues with these detectors. These supervisors need a thorough understanding of the potential impacts of AI detectors to efficiently explore possible solutions. These stakeholders are currently under increased scrutiny and duress due to the sheer number of false positives and the need to make determinations that could impact a student's entire life. The policy will alleviate the onus administrators currently feel by clearly placing ownership of these decisions in the hands of the individuals to whom they report and by reducing the number of false positives. 5) National, State, and local governments—these agencies are ultimately responsible for either providing or mandating training in the use of AI detectors. They need to have a clear understanding of the regulations and policies that govern these technologies. This includes understanding the effectiveness of and how to use "AI humanizers." Any policy instituted at the institutional level will make these stakeholders' jobs easier by clearly outlining what is expected.

Appendix I – AI usage

(NOTE: portions of this appendix have been taken from the previous submission)

1. **Grammarly** (https://app.grammarly.com/)
   a. Throughout the paper, both the spelling and grammar were checked using Grammarly.
   b. There were no "exchanges".
   c. Grammarly suggested changes to wording and corrected spelling. Changes were not always accepted.
   d. Grammarly was chosen to ensure consistent word choice throughout.
2. **Google summaries**
   a. Primarily used for finding relevant case studies and sources
   b. Some summaries generated by searches were used as a basis for some of the written material
3. **EndNote 2025-1** (https://endnote.com/)
   a. EndNote was used to create the bibliography
   b. EndNote was used to create in-text citations
   c. EndNote was chosen to ensure that APA style was used both in the in-text citations and the bibliography. The APA style, which EndNote provides, has been modified to include features such as line numbers.
4. **Lumo AI chatbot from Proton**
   a. Lumo was used to evaluate this writing and to suggest possible corrections (this was done prior to making changes)
      i. Link to conversation: https://lumo.proton.me/u/8/c/f143adb8-bce6-40fa-a484-13a6bbebf0d1
      ii. Prompt used: "Evaluate the uploaded paper using the uploaded OMDS rubric." Both the policy document and the rubric were uploaded to the Lumo chatbot.
5. **The Comet Assistant search engine** from Perplexity was used to search for sources

The various AI tools were used to identify different sources and case studies, including summaries and searches, to support the existing content and multiple ideas I presented through searches. These tools also helped in annotating this document with EndNote and checking spelling and grammar throughout using Grammarly. Although I could have completed this paper without AI, using AI made it easier to locate and identify the resources and relevant case studies I would have eventually found. *Grammarly*'s error checking also made writing much simpler. For annotation, EndNote 2025-1 was used; additional work was needed to import all citations into an EndNote library and then cite them in this document (due to technical difficulties). I still would use EndNote again as a citation tool.

Various AI-enabled features, listed above, played a significant role in this paper by highlighting its strengths and weaknesses and suggesting ways I could improve the final version. Lumo proposed including a Power interest stakeholder matrix and a rollout timeline as methods to enhance the paper.

Appendix II – Data

Stanford (Myers, 2023)

Sample Size	Source	Type	Generated by	Website
91 Essays	Chinese educational forum	TOEFL	Human	https://arxiv.org/pdf/2304.02819.pdf (full study)
88 Essays	Hewlett Foundation's Automated Student Assessment Prize (ASAP) dataset	US 8th grade essays	Human	https://www.kaggle.com/competitions/asap-aes

Further code and Data at:
https://github.com/Weixin-Liang/ChatGPT-Detector-Bias/tree/v1.0.0 (Liang et al., 2023a, 2023b)

Appendix III – Tables

Table 1. Studies on AI Detection and ESL Writers

Study / Report	Key Findings on ELL Writers	False Positive Rate / Metric
Center for Democracy & Technology brief (based on Stanford study) – "Disproportionate Effects of Generative AI Detectors on English Learners"	AI detectors that worked almost perfectly on native-speaker essays falsely flagged a majority of TOEFL essays.	61% of non-native essays flagged; 19% unanimously flagged; 97% flagged by at least one detector.
The Markup **(Mathewson, 2023)**	Replicated Stanford study. ELL writers disproportionately likely to be flagged; real disciplinary cases reported.	Same figures (~60% false positives).
Turnitin blog (internal research)	Acknowledges bias toward ELL writers. Reports <1% overall false positives when ≥20% of text is AI, but higher incidence with low-AI-content texts, especially for ELL writers.	No precise %, but confirms bias exists.

Table 2. Comparison of a few AI Detection/Humanizer Tools

Checker Name	Website	Offers Humanizer Tool	Comment
JustDone	https://app.justdone.ai	Yes	Requires subscription
GPTZero	https://gptzero.me/	No	Requires subscription
Undetectable	https://undetectable.ai/	Yes	Money back guarantee. If anything we produce is flagged as not human, we will refund the cost of humanization
Grammarly	https://app.grammarly.com/	Yes	Requires subscription. My subscription allowed me to run this through the AI-detector, which found 12% likely AI generated.
Copyleaks	https://copyleaks.com/	No	Over 99% accuracy*, verified through rigorous testing methodologies. Trusted globally to detect AI across 30+ languages and leading LLMs like ChatGPT, Gemini, DeepSeek, and Claude. *Accuracy rating is based on internal testing of the English language datasets.
HumanizeAI	https://www.humanizeai.pro/	Yes	Humanize AI stands out as the leading, cost-free online platform designed for transforming AI-generated text into human-like content. **Changed** "Effortlessly transform AI-generated text into human-like content with BypassGPT free online." **To** "Seamlessly convert AI-generated text into human-like content using BypassGPT free online."
BypassGPT	https://bypassgpt.co/	Yes - limited to 200 chr	Effortlessly transform AI-generated text into human-like content with BypassGPT free online. Bypass AI detection systems like GPTZero and ZeroGPT to ensure your content remains undetectable.
WinstonAI	https://gowinston.ai/	No – free trial with 2000 word limit	Claims a 99.98% accuracy rate. "Detects all bypassing strategies, including paraphrasing content with tools such as Quillbot, or even AI content humanizers."
WriteWell	https://brandwell.ai/		Yes – says 2500 word limit but used only first 356 words. Published this article https://brandwell.ai/blog/undetectable-ai-content/ on how to create fully human undetectable content even with AI.

Checker Name	Website	Offers Humanizer Tool	Comment
OriginalityAI	https://originality.ai/	No	We are 80% confident that the text scanned is Original (Human written), NOT to be interpreted as 80% of the text produced is Original (Human written).
Writefull	https://x.writefull.com/gpt-detector	Word Plugin	The word plugin will offer to rewrite your text for you, making suggestions. (free version is limited)

Table 3 - Stakeholders

Stakeholder	Power	Interest	Quadrant	Action Items
Provost / Academic Integrity Office	High	High	Manage Closely	• Sponsor steering committee • Approve budget
Legal / Compliance Office	High	Medium High	Manage Closely	• Review policy language for Title IV compliance
IT Services	High	Medium	Keep Satisfied	• Provide integration timeline • Allocate API support
Faculty Senate / Dept. Chairs	Medium	High	Manage Closely	• Draft department-level addenda • Host workshops
Individual Instructors	Low	High	Keep Informed	• Distribute quick-start guide • Collect usability feedback
ELL / International Students	Low	High	Keep Informed	• Conduct focus groups • Publish plain-language policy
AI-Detector Vendors	Medium	Low-Medium	Monitor	• Request bias-mitigation documentation
State Education Agency	High	Low-Medium	Keep Satisfied	• Submit compliance summary
Student Government	Low-Medium	High	Keep Informed	• Invite to review meetings
General Public / Media	Low	Low	Monitor	• Prepare press kit (on-demand)

Table 4 – Timeline

Phase	Duration	Core Goal	Key Milestones & Deliverables	Owner(s)	Success Metric
0 – Initiation & Governance	2 weeks	Secure sponsorship, define scope, assemble steering team.	▪ Charter & project charter signed ▪ Steering committee roster (Provost, Legal, IT, Faculty Senate rep, Student-advocate). ▪ Project-management workspace	Provost / Academic Integrity Office	Charter approved within 10 days; all roles confirmed.
1 – Discovery & Requirements	4 weeks	Gather evidence, map current workflows, capture stakeholder needs.	▪ Conduct 3 focus-groups (ELL students, faculty, admin). ▪ Survey ≥ 70 % of instructors on current detector use. ▪ Compile "Current State" process map (detector → decision) ▪ Draft "Requirements Document" (legal, technical, pedagogical).	Lead: Faculty Senate Chair; Support: Compliance & IT	≥ 80 % response rate; requirements sign-off by Legal.
2 – Policy Draft &	3 weeks	Write policy language, embed	▪ First-draft policy circulated to steering committee.	Policy Lead (Academic	Legal sign-off within 5 days of receipt.

Phase	Duration	Core Goal	Key Milestones & Deliverables	Owner(s)	Success Metric
Legal Review		bias-mitigation clauses, obtain legal clearance.	▪ Legal review memo (Title IV, AI-Bill of Rights compliance). ▪ Incorporate feedback → "Final Draft".	Integrity Office) + Legal Counsel	
3 – Technical Pilot (Semester 1)	6 weeks	Test detector + human-review workflow in a limited setting.	▪ Select 2-3 courses (high ELL enrollment). ▪ Deploy chosen detector (e.g., Pangram Labs) with API integration. ▪ Train 2-3 faculty champions on review protocol. ▪ Collect quantitative data: false-positive rate, review time, student satisfaction. ▪ Mid-pilot check-in meeting.	IT Services (integration) + Faculty Pilot Leads	≤ 10 % false-positive rate after human review; ≥ 75 % faculty satisfaction.
4 – Evaluation & Iteration	2 weeks (after pilot)	Analyse pilot data, refine policy & tech stack.	▪ Produce "Pilot Evaluation Report" (metrics, lessons learned). ▪ Update policy language (add any needed clarifications). ▪ Decide on detector vendor (cost-benefit).	Steering Committee (incl. Data Analyst)	Recommendation approved by majority vote; revised policy ready for full roll-out.
5 – Full-Scale Roll-out (Semester 2)	8 weeks	Institution-wide adoption, training, and support.	▪ Publish final policy on the university website & LMS. ▪ Mandatory 1-hour online training for all instructors (recorded webinar + quiz). ▪ Create "Help Desk" ticket queue for detector issues. ▪ Deploy detector across all courses (auto-enabled but with opt-out for exempted classes). ▪ Release "Student Guide" explaining rights, appeals, and error-rate transparency.	Academic Tech Team + Communications Office	100 % of instructors complete training; detector active in ≥ 95 % of courses.
6 – Monitoring & Continuous Improvement	Ongoing (quarterly reviews)	Track compliance, refine processes, and report to governance.	▪ Quarterly KPI dashboard (false-positive %, appeal volume, turnaround time). ▪ Annual stakeholder survey (students & faculty). ▪ Update policy/tech as needed (e.g., new detector releases).	Monitoring Sub-Committee (Legal, IT, Faculty, Student Rep)	KPI trends moving toward ≤ 5 % false-positive rate; ≤ 10 % appeal rate.
7 – Formal Close-out	2 weeks (end of academic year)	Archive documentation, hand-off to permanent governance.	▪ Final "Project Close-out Report" (budget, outcomes, lessons). ▪ Archive all artifacts in the Institutional Repository. ▪ Transition ongoing monitoring to the Academic Integrity Office.	Project Manager + Academic Integrity Office	Report signed off; repository entry complete.

The various costs (both political and monetary) may be initially high, but this policy will help create the kind of academic atmosphere that fosters students' potential, which is why most educators entered the field in the first place.

Appendix IV - AI Detector suggestions – things to look out for.

A few items to take into account when selecting an AI detection tool, that may save time: 1) **Cost** – How much does it cost to use this AI detector? If there is a free trial, how long is it, and is there a limit to the amount of text you can submit? 2) **Accuracy** – What is the reported accuracy of the detector (both from the company and third parties) and is this accuracy the same with ELL written work. 3) **What is included** - Does it include a plagiarism checker? Does it cost extra? Does the tool also include an AI humanizer? 4) **Integration** - Also instructive is whether there is a browser plug-in, access to an API, and integration with Large Language Models (LLMs) 5) **Batch** - check to see if there is a batch upload feature, which may save a lot of time.[5]

---

[5] Comparison of a few AI detectors/humanizers

Appendix V – Alternate Solutions

Technical Solutions

1. There is a need to develop a method for identifying AI-generated texts while minimizing the likelihood of false positives among ESL students. According to Jiang study (*Pangram Labs AI Detection*), "research investigating these detectors' [AI-detectors] fairness and potential bias is relatively rare…"(Jiang et al., 2024).

   Cost: The Pangram solution costs $5.00/student/year.

2. GPTZero also claims to have developed a technical solution to the problem (Tian, 2023) and even reran the Stanford study's algorithm on its new product, with a much better result.

   Cost: Educators can receive this product for free for life.

Assignment-design toolkits that are already published and used

- MIT Sloan's "Designing AI-Resistant Assignments" ("AI Detectors Don't Work. Here's What to Do Instead.", 2024) – This offers templates for prompts that require personal reflection, local data, or multiple drafts – making pure AI generation unfeasible.
- Carnegie Mellon "Generative AI FAQ for Instructors" (University, 2024) - Provides a decision tree on when to use a detector, when to rely on draft histories, and how to communicate expectations to students.

Appendix VI – Example Policies
(https://teaching.cornell.edu/generative-artificial-intelligence/ai-academic-integrity)

As you plan your syllabus and course policies with respect to generative AI, consider modifying the following language to communicate a general position in your syllabus. Please note that the following sample language reflects general, course-level perspectives on broadly permitting or prohibiting the use of generative AI tools. For sample statements at the assignment-level, see AI in Assignment Design.

1. Prohibiting Generative AI Use in Your Course (AI-FREE)
"To ensure development and mastery of the foundational concepts and skills in this course, the use of generative artificial intelligence (AI) tools is prohibited. This includes tools that help reorganize and edit your written work because the ability to self-assess, reflect on your writing process, and develop your own voice are essential in your growth as a writer. If you are unsure of any policy or any assignment-specific directions – including whether or not a tool is considered generative AI – please consult with me prior to using the technology or completing your assignment. You are responsible for verifying the accuracy of citations and references used in your writing. You will be asked to verbally explain your research and writing process during in-class discussions and in a one-on-one meeting with me.

In acknowledgment that AI is impacting our field, please know that appropriate and ethical use of generative AI tools will likely be a part of other courses in your academic program."
Permitting Generative AI Use with Attribution in Your Course (AS-UA)

"Mastering the essential, foundational concepts of this course takes effort and practice. Accordingly, the use of generative artificial intelligence (AI) tools is generally discouraged in this course, but will be permitted for select assignments. Whether or not generative AI assistance is permitted for each assignment will be explicitly communicated when that assignment is introduced.

If used in any capacity for an assignment, generative AI requires proper attribution for any and all generated work. As AI-generated materials are not retrievable by graders—and there is not a person to whom the work can be attributed—students should attribute directly quoted text to the creator of the generative AI tool used (e.g., cite OpenAI when directly quoting ChatGPT). This attribution should be used for both in-text citations and your reference list.

> **Example:** When prompted with, "Is it ethical to use generative AI without proper attribution?" ChatGPT indicated, "Using generative AI without proper attribution can be considered ethically problematic, as it raises issues related to intellectual property, transparency, and honesty" (OpenAI, 2023).
> **Reference**
> OpenAI. (2023). ChatGPT (Aug 10 GPT-3.5 version) [Large language model]. https://chat.openai.com

For full details on how to properly cite AI-generated work, please see the APA Style article, How to Cite ChatGPT.

Regardless of whether or not the use of generative AI is permitted or prohibited for an assignment, it is critical that you adhere to our communicated course policy (and Cornell's policy) on academic integrity. If you are unsure of any policy or any assignment-specific directions-–including whether or not a tool is considered generative AI-–please consult the instructor *prior to* using the technology or completing your assignment."

    2. <u>Encouraging Generative AI Use with Attribution in Your Course</u> (ANY-AI-UA)
"The use of generative artificial intelligence (AI) tools is encouraged on identified assignments. The directions for each assignment in this course will clearly indicate whether or not the use of generative AI is permitted for that assignment.

If used at all, generative AI requires proper attribution for any generated work. As AI-generated materials are not retrievable by graders—and there is not a person to whom the work can be attributed—students should attribute directly quoted text to the creator of the generative AI tool used (e.g., cite OpenAI when directly quoting ChatGPT). This attribution should be used for both in-text citations and your reference list**.**

> **Example:** When prompted with, "Is it ethical to use generative AI without proper attribution?" ChatGPT indicated, "Using generative AI without proper attribution can be considered ethically problematic, as it raises issues related to intellectual property, transparency, and honesty" (OpenAI, 2023).
> **Reference**
> OpenAI. (2023). ChatGPT (Aug 10 GPT-3.5 version) [Large language model]. https://chat.openai.com

For full details on how to properly cite AI-generated work, please see the APA Style article, How to Cite ChatGPT.

Regardless of whether or not the use of generative AI is permitted or prohibited for an assignment, it is critical that you adhere to our communicated course policy (and Cornell's policy) on academic integrity. If you are unsure of any policy or any assignment-specific directions-–including whether or not a tool is considered generative AI-–please consult the instructor prior to using the technology or completing your assignment."

    3. <u>Permitting Generative AI Use on an Assignment-by-Assignment Basis</u> (AS)
"Policies concerning the use of generative artificial intelligence (AI) tools will be decided on an assignment-by-assignment basis. These policies will be clearly communicated alongside other details of each specific assignment.

As generative AI technologies evolve, class policies with respect toward the use of generative AI tools may be shaped by in-class discussions regarding the fair use of AI and its implications on careers in this field.

Regardless of whether or not the use of generative AI is permitted or prohibited for an assignment, it is critical that you adhere to our communicated course policy (and Cornell's policy) on academic integrity. If you are unsure of any policy or any assignment-specific directions -–including whether or not a tool is considered AI-– please consult the instructor *prior to* using the technology or completing your assignment."(Cornell University, 2023)

Bibliography

1   *The 14th Amendment and the Evolution of Title IX*. (2025).
    https://www.uscourts.gov/about-federal-courts/educational-resources/educational-activities/14th-amendment-and-evolution-title-ix

2   *34 CFR Part 99 -- Family Educational Rights and Privacy*. (2025).
    https://www.ecfr.gov/current/title-34/subtitle-A/part-99?toc=1

3   *AB 2013- CHAPTERED*. (2024).
    https://leginfo.legislature.ca.gov/faces/billTextClient.xhtml?bill_id=202320240AB2013

4   Academic Integrity. (2025). *Office of the Dean of Faculty*.
    https://deanoffaculty.cornell.edu/faculty-and-academic-affairs/academic-integrity/

5   *Access to Equal Education Opportunity | U.S. Department of Education*. (2024).
    https://www.ed.gov/laws-and-policy/civil-rights-laws/access-equal-education-opportunity

6   Alsharefeen, R. (2025). Frontiers | Faculty as street-level bureaucrats: discretionary
    decision-making in the era of generative AI. https://doi.org/10.3389/feduc.2025.1662657

7   Assessing Student Writing. (2025). *The National Children's Book and Literacy Alliance*.
    https://thencbla.org/assessing-student-writing/

8   Berman, L. (2025). *Professors shift many essay-based assessments to in-person exams
    amid AI concerns*. https://www.tuftsdaily.com/article/2025/10/professors-shift-many-essay-based-assessments-to-in-person-exams-amid-ai-concerns

9   Bohannon, M. (2023). *Lawyer Used ChatGPT In Court—And Cited Fake Cases. A Judge
    Is Considering Sanctions*.
    https://www.forbes.com/sites/mollybohannon/2023/06/08/lawyer-used-chatgpt-in-court-and-cited-fake-cases-a-judge-is-considering-sanctions/

10 BU. (2023). Generative AI Assistance Policy. https://www.bu.edu/files/2023/02/GAIA-Final-2023.pdf

11 *California Bill to Require Disclosures on Data Used to Train AI Models*. (2025).
    https://www.jonesday.com/en/insights/2025/09/california-bill-will-require-disclosures-about-data-used-in-training-ai-models

12 *California's New AI Laws Focus on Training Data, Content Transparency // Cooley //
    Global Law Firm*. (2024). Cooley. https://www.cooley.com/news/insight/2024/2024-10-16-californias-new-ai-laws-focus-on-training-data-content-transparency

13 Cambridge. (2025). *What is a Baseline Assessment?*
    https://www.cambridge.org/insight/baseline-assessments

14 Castellanos-Canales, A. (2025). Explainer: Revocation of Student Visas and Termination
    of SEVIS Records. *National Immigration Forum*.
    https://forumtogether.org/article/explainer-revocation-of-student-visas-and-termination-of-sevis-records/

15 CELT. (2025). *AI Course Policy Examples | CELT*. https://celt.uky.edu/ai-course-policy-examples

16 Check, E. (2025). *False Positives in AI Detection: Are We Wrongly Accusing Students?*
    https://essaycheck.ai/blog/false-positives-in-ai-detection

17 *Civil Rights Division | Types Of Educational Opportunities Discrimination*. (2015, 2015–08–06T09:30:07–04:00). https://www.justice.gov/crt/types-educational-opportunities-discrimination

18   Code of Academic Integrity. (2025). *Office of the Dean of Faculty*. https://deanoffaculty.cornell.edu/faculty-and-academic-affairs/academic-integrity/code-of-academic-integrity/

19   Coley, M. (2023). *Guidance on AI Detection and Why We're Disabling Turnitin's AI Detector*. Retrieved 09/09 from https://www.vanderbilt.edu/brightspace/2023/08/16/guidance-on-ai-detection-and-why-were-disabling-turnitins-ai-detector/

20   Cornell University, C. f. T. I. (2023). *AI & Academic Integrity | Center for Teaching Innovation*. https://teaching.cornell.edu/generative-artificial-intelligence/ai-academic-integrity

21   D'Agostino, S. (2023). *Turnitin's AI Detector: Higher-Than-Expected False Positives*. https://www.insidehighered.com/news/quick-takes/2023/06/01/turnitins-ai-detector-higher-expected-false-positives

22   Dang, A., & Wang, H. (2024). Ethical use of generative AI for writing practices: Addressing linguistically diverse students in U.S. Universities' AI statements. *Journal of Second Language Writing*, *66*, 101157. https://doi.org/10.1016/j.jslw.2024.101157

23   Employability-Survey-Report. (2024). https://cengage.widen.net/s/bmjxxjx9mm/cg-2024-employability-survey-report

24   Equal_Treatment_of_Students_at_Universities_Act. https://www.equalrightstrust.org/ertdocumentbank/Microsoft%20Word%20-%20Equal_Treatment_of_Students_at_Universities_Act.pdf

25   *FERPA | Protecting Student Privacy*. (2011). https://studentprivacy.ed.gov/ferpa

26   Furze, L. (2024). AI Detection in Education is a Dead End. *Leon Furze*. https://leonfurze.com/2024/04/09/ai-detection-in-education-is-a-dead-end/

27   Giray, L. (2024). The Problem with False Positives: AI Detection Unfairly Accuses Scholars of AI Plagiarism. *The Serials Librarian*, *85*(5-6), 181–189. https://doi.org/10.1080/0361526X.2024.2433256

28   Hirsch, A. (2024). AI detectors: An ethical minefield - Center for Innovative Teaching and Learning. https://citl.news.niu.edu/2024/12/12/ai-detectors-an-ethical-minefield/

29   Horn, T. J., & Polce, J. (2024). *Lies, Damned Lies, and AI: Lie-Detecting AI May Expose Employers to Liability | Insights | Venable LLP*. https://www.venable.com/insights/publications/2024/04/lies-damned-lies-and-ai-lie-detecting-ai-may

30   Jones, K. (2025). *Databases & Subject Guides: Introduction to AI Literacy : For Instructors*. https://guides.libs.uga.edu/AI/instructors

31   Kratzer, A. (2020). The Why and How of a Baseline Writing Assignment. *Angie Kratzer*. https://angiekratzer.com/the-why-and-how-of-a-baseline-writing-assignment/

32   Liang, W., Yuksekgonul, M., Mao, Y., Wu, E., & Zou, J. (2023a). *AI-Detectors Biased Against Non-Native English Writers*. https://hai.stanford.edu/news/ai-detectors-biased-against-non-native-english-writers

33   Liang, W., Yuksekgonul, M., Mao, Y., Wu, E., & Zou, J. (2023b). GPT detectors are biased against non-native English writers. https://doi.org/10.48550/arXiv.2304.02819

34   Mahmood, S. H. (2023, 2023–02–25T16:30:17Z). *The 8 Most Accurate AI Text Detectors You Can Try*. https://www.makeuseof.com/accurate-ai-text-detectors/

35   Marcus, G. (2022). *AI Platforms like ChatGPT Are Easy to Use but Also Potentially Dangerous*. https://www.scientificamerican.com/article/ai-platforms-like-chatgpt-are-easy-to-use-but-also-potentially-dangerous/

36   Markey, E. *Senator Markey Introduces AI Civil Rights Act to Eliminate AI Bias, Enact Guardrails on Use of Algorithms in Decisions Impacting People's Rights, Civil Liberties, Livelihoods | U.S. Senator Ed Markey of Massachusetts*. https://www.markey.senate.gov/news/press-releases/senator-markey-introduces-ai-civil-rights-act-to-eliminate-ai-bias-enact-guardrails-on-use-of-algorithms-in-decisions-impacting-peoples-rights-civil-liberties-livelihoods

37   Mathewson, T. G. (2023, 2023–08–14). *AI Detection Tools Falsely Accuse International Students of Cheating – The Markup*. https://themarkup.org/machine-learning/2023/08/14/ai-detection-tools-falsely-accuse-international-students-of-cheating

38   McLogan, J. (2025, 2025–10–10). *Adelphi University facing lawsuit after AI-assisted plagiarism accusation against student*. https://www.cbsnews.com/newyork/news/adelphi-university-artificial-intelligence-plagiarism-lawsuit/

39   MIT_Sloan. (2023). AI Detectors Don't Work. Here's What to Do Instead. *MIT Sloan Teaching & Learning Technologies*. https://mitsloanedtech.mit.edu/ai/teach/ai-detectors-dont-work/

40   Myers, A. (2023). *AI-Detectors Biased Against Non-Native English Writers | Stanford HAI*. Stanford University. Retrieved 09/09 from https://hai.stanford.edu/news/ai-detectors-biased-against-non-native-english-writers

41   Najarro, I. (2023, 2023–10–06T20:31:40.743). What Teachers Should Know Before Using AI With English Learners. *Education Week*. https://www.edweek.org/teaching-learning/what-teachers-should-know-before-using-ai-with-english-learners/2023/10

42   Orr, D. (2022). The Unintended Consequences of Algorithmic Bias. *Congressional Black Caucus Foundation*. https://www.cbcfinc.org/wp-content/uploads/2022/04/2022_CBCF_CPAR_TheUnintendedConsequencesofAlgorithmicBias_Final.pdf

43   Palmer-Clarke, Y. (2025). *Research Guides: English as an Additional Language: Using Artificial Intelligence*. https://libguides.usask.ca/c.php?g=109857&p=5362687

44   *Pangram Labs AI Detection*. (2025). https://www.pangram.com/

45   Prothero, A. (2024). More Teachers Are Using AI-Detection Tools. Here's Why That Might Be a Problem. *Education Week*. https://www.edweek.org/technology/more-teachers-are-using-ai-detection-tools-heres-why-that-might-be-a-problem/2024/04

46   Qanbar, Y. (2025). *Emerging Theories of Liability in the Age of AI: A New Frontier for Litigation | Rain Intelligence Research*. https://www.rainintelligence.com/blog/emerging-theories-of-liability-in-the-age-of-ai-a-new-frontier-for-litigation

47   Student rights in U.S. higher education. (2025). In *Wikipedia*.

48   *Supreme Court Holds That Considering Race in College Admissions Violates the Equal Protection Clause and Title VI of the Civil Rights Act*. (2025). https://www.crowell.com/en/insights/client-alerts/supreme-court-holds-that-considering-race-in-college-admissions-violates-the-equal-protection-clause-and-title-vi-of-the-civil-rights-act

49   Tian, E. (2023, 2023–10–25T01:43:00.000Z). *ESL Bias in AI Detection is an Outdated Narrative*. Retrieved 09/09 from https://gptzero.me/news/esl-and-ai-detection/

50   Univ_Pittsburgh. (2023). *Teaching Center doesn't endorse any generative AI detection tools*. https://www.utimes.pitt.edu/news/teaching-center-doesn-t

51   University_of_Rochester. (2025). *International Students*. https://www.rochester.edu/college/honesty/students/international.html

52   UTEXAS. (2025). AI Detection Software Guidance. *Office of the Executive Vice President and Provost*. https://provost.utexas.edu/the-office/academic-affairs/office-of-academic-technology/ai-detection-software-guidance/

53   Vanderbilt_Univ. (2023). Academic-Affairs-Guidance-for-Artificial-Intelligence. https://cdn.vanderbilt.edu/vu-URL/wp-content/uploads/sites/439/2023/11/29200845/Academic-Affairs-Guidance-for-Artificial-Intelligence-Nov-29-2023.pdf

54   Weaver, A. (2023a). AI hallucinations: How to reduce inaccurate content outputs. *WRITER*. https://writer.com/blog/ai-hallucinations/

55   Weaver, A. (2023b). Why you can't replace human insight: The importance of editorial skills in the AI era. *WRITER*. https://writer.com/blog/editorial-skills-and-ai/

56   Weaver, A. (2025). Generative AI: risks and countermeasures for businesses. *WRITER*. https://writer.com/blog/generative-ai-risks-and-countermeasures/

57   Why AI Detection Tools are Ineffective | Center for Academic Innovation. (2025). https://facultyhub.chemeketa.edu/technology/generative-ai/why-ai-detection-tools-are-ineffective/

58   Woelfel, K. (2023). Brief – Late Applications: Disproportionate Effects of Generative AI-Detectors on English Learners. *Center for Democracy and Technology*. https://cdt.org/insights/brief-late-applications-disproportionate-effects-of-generative-ai-detectors-on-english-learners/

59   Woelfel, K., Aboulafia, A., & Laird, E. (2023). Report - Late Applications: Protecting Students' Civil Rights in the Digital Age. *Center for Democracy and Technology*. https://cdt.org/insights/report-late-applications-protecting-students-civil-rights-in-the-digital-age/

60   Wood, C. (2024). AI detectors are easily fooled, researchers find. *EdScoop*. https://edscoop.com/ai-detectors-are-easily-fooled-researchers-find/