Bias of AI-Detectors towards English Language Learners

Author: The author of the study/article is Tara García Mathewson. The author wrote this article to highlight the issues involved in several cases. Ms. Mathewson works as a technology writer for The Markup, an online publication.

Topic: This case study examines how AI detection tools show significant bias against English Language Learners (ELL), producing false plagiarism/AI accusations at disproportionate rates.

Summary: AI detectors disproportionately return false positives for English language learners (ELL). A primary reason is how these algorithms are trained. This is borne out by a Stanford University study, which found that the false positive rate of essays written by ELLs could be higher than 61% (Myers, 2023). A separate study compared AI detectors' false-positive rates. The study found that essays written by ELLs were more than twice as likely to be identified as AI-generated when compared with those of native speakers. (Blog, 2024)

An international student at Johns Hopkins University was wrongly accused of using AI because of the limitations of the Turnitin AI detector. When Professor Taylor Hahn contacted the student, she was already prepared to demonstrate that her writing was her own, since this issue often affects ELL students.

Dr. Hahn assisted another student with both the outline and drafts of their essay, only to find that most of the submitted paper was incorrectly flagged as AI-generated. Both students were English Language Learners, specifically international students.

It is perhaps most indicative of the problem that many companies have removed their "AI detectors" until further research can confirm their accuracy, most notably among them is OpenAI (Carter, 2023; Coldewey, 2023; Epstein-Gross, 2023; Forlini, 2023; Mathewson, 2023; Northwestern_University, 2025; Quill.org, 2025; Rifalixa, 2023). Turnitin has introduced a feature that enables institutions to disable the AI detector it uses. As of July 2023, only 2% of institutions using Turnitin had utilized this feature. (Mathewson, 2023). In May 2024, the top 20 universities advised against using AI detectors (Rumi, 2024). Some educators, including one at the University of Pittsburgh, decided that AI detectors "could do more harm than good."(Furze, 2024; Mathewson, 2023, 2025)

This study illustrates how biased detection tools can jeopardize students' legal status and erode trust in academic institutions.

Stakeholders:

1. International students and English language learners are mostly victims in these situations. The consequences can be severe enough to lead to deportation from the USA. (Castellanos-Canales, 2025; University_of_Rochester, 2025). Such reactions are not uncommon and are even more unjustified if the accusation turns out to be a false positive. The specific students remain unnamed.

Bias of AI-Detectors towards English Language Learners

2. Taylor Hahn, a professor at Johns Hopkins University, observed that Turnitin's tool was much more likely to flag international students' writing as AI-generated. The professor identified the potential problem with this technology.
3. Turnitin—the company Johns Hopkins University was using at the time to detect AI-generated text. The Turnitin website references an article that states the Turnitin AI detector is accurate but does not explicitly address ELL students. (Walters, 2023). The company itself feels that its AI detector should not be the sole determinant of writing originality:

   > *"Crucially, Turnitin Originality does not make a determination of misconduct through our AI writing detection (or similarity checking) technology. ... we provide data for educators to make an informed decision based on their academic and institutional policies."*

4. Administrative personnel at Johns Hopkins—Instructor supervisors are responsible for overseeing professors, such as Dr. Hahn. 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 explore possible solutions efficiently.
5. National, State, and local governments—these agencies are ultimately responsible for providing 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 knowing the effectiveness of, and how to use, "AI humanizers."

Broader Issues: This case illustrates the tension between algorithmic efficiency and equity. The false positives generated when AI detectors are used with ELL highlight broader issues, including civil rights, increased risks of discipline or deportation, and legitimate uses of AI. (Corrigan, 2024).

In this case, the AI detection algorithms were not trained on a representative sample; instead, they primarily relied on the writing of native English speakers. This resulted in false positives when applied to ELL writing.

Strategies to address these potential issues should be implemented in a way that benefits the widest range of learners fairly. The questions of why false positives occur in the first place (SkylineAcademic, 2025) and why these instructors believe students rely so heavily on AI that they feel the need for AI detectors need to be addressed. Furthermore, if students are indeed using AI, why do they feel compelled to do so?

With the advent of AI chatbots, professors are, in some cases, changing how they assess student performance (Berman, 2025; Stanford, 2025; Supiano, 2023).

The study addresses civil rights issues because if a false positive disproportionately affects ELL, then Title IV of the Civil Rights Act concerning "disparate treatment," "disparate impact,"

Bias of AI-Detectors towards English Language Learners

or “hostile learning environment” could be invoked. This could have more serious implications than initially expected. (Markey; Orr, 2022). The Office of Civil Rights stated in November 2024, not only directly addressing AI detector use and ELL, but also cautioning that inappropriate use of AI detectors could lead to an OCR investigation. (Riess, 2025)

Other technology, such as automated scoring, has significantly less bias against ELL. (Guo et al., 2025) Facial recognition, though, has been repeatedly shown to be “least reliable for people of color, women, and nonbinary individuals.” (Fergus, 2024)

Importance: Any bias against ELL may arise from issues in the training data or the algorithms used. Fixing these issues for this group is more manageable and will help foster a broader discussion on the best ways to train such algorithms(Williams, 2025). The larger concern of AI use and privacy is so critical that the White House issued an AI Bill of Rights in 2022(White_House, 2022). This specific issue is addressed in the Human Alternatives, Consideration, and Fallback section of that document. To paraphrase the European Union document titled “*Ethics guidelines for trustworthy AI*,”

> “…marginalization of vulnerable groups, [will] exacerbate prejudice and discrimination(European_Union, 2019).”

Investigating the use and training of AI detectors can only facilitate the broader issue of ethics required when creating any model and/or neural network. As with computer security, ethical considerations must be taken into account from the outset.

It is significant that a false positive could lead to a student’s visa being revoked and the student being deported. (Castellanos-Canales, 2025; University_North_Carolina_Charlotte, 2025; University_of_Minnesota, 2025; University_of_Rochester, 2025; VnExpress, 2025) Given the possibility of such an action, AI detectors should never serve as the sole proof that AI has been used in a student’s writing.

Bias of AI-Detectors towards English Language Learners

Appendix I – AI usage

(*Note: this appendix was copied, in part, from my previous policy paper in this course*)

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 the writing of this case study and to suggest possible corrections (this was done prior to making changes, other than including the link, this comment, the prompt, and the full text of the conversation)
        i. Link to conversation: https://lumo.proton.me/u/0/c/2d757bf6-a2df-457a-a72d-fd07976bd148
        ii. Prompt used: "Evaluate the case study using the Rubric provided" while uploading the rubric and the case study
        iii. Full conversation – Appendix 1A
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 aided in annotating this document (EndNote) and checking spelling and grammar throughout the document (Grammarly). Although this paper could have been completed without the use of AI, utilizing AI made it easier to locate and identify the resources and relevant case studies that I would have eventually found. The error checking by Grammarly also made writing much simpler.

In terms of annotation, EndNote 2025-1 was used; additional work was needed to get 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.

Bias of AI-Detectors towards English Language Learners

Appendix II – Data

Copyleaks (Blog, 2024)

FCE v2.1 - This dataset contains written answers from the First Certificate in English (FCE) exam

ELLs - This dataset includes essays from English Language Learners (ELLs) in grades 8 through 12.

COREFL - This dataset features written texts from learners of English as a second or foreign language.

Summary:

Dataset	Texts	False Positives	Accuracy
FCE v2.1	2,116	4	0.9981
ELLs	3,911	0	1
COREFL	1,455	8	0.9945
Total	7,482	12	0.9984

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., 2023; Weixin Liang, 2023)

Bias of AI-Detectors towards English Language Learners

Appendix 1A – Complete Lumo conversation text

I've read both the **OMDS M5 Case-Study Rubric** (PDF) and the **Thomas Zucker-Scharff case study** (PDF). Below is a point-by-point assessment against each rubric criterion, the resulting score out of 35, and concrete suggestions for moving every section from "Proficient" toward "Exemplary".

---

1. Case-Study Summary (250-300 words) – **4 / 5 (Proficient)**

What the rubric expects: a comprehensive, clear, concise overview that organizes the key events and gives the reader a solid grasp of the issue.

How the paper performs:

- Covers the bias problem, cites the Stanford study, describes the Johns Hopkins incident, and notes institutional responses.
- Word count is within the required range.
- Flow is mostly linear, but occasional jumps between statistics and anecdote interrupt readability.

Improvement tip: tighten transitions (e.g., link the Stanford findings directly to the Johns Hopkins case) and trim any redundant phrasing. A smoother narrative would earn the "Exemplary" rating.

---

2. Stakeholders & Impact (250-300 words) – **5 / 5 (Exemplary)**

- Identifies five relevant groups (ELL students, Prof. Hahn, Turnitin, university administrators, government regulators).
- Provides thoughtful, evidence-backed analysis of impact on each (academic penalties, visa risk, policy implications, reputational damage).
- No gaps—meets the rubric perfectly.

---

3. Broader Issues in Technology & Ethics (250-300 words) – **4 / 5 (Proficient)**

- Connects the case to algorithmic bias, civil-rights implications (Title IV), the AI Bill of Rights, and EU ethics guidelines.
- Mentions other biased technologies (facial-recognition) only in passing and does not give a concrete comparative example.

Improvement tip: add a brief parallel (e.g., facial-recognition misclassification of darker-skinned faces) or another AI-detector bias study. That extra depth will push the score to "Exemplary".

---

4. Argument for Importance (150-200 words) – **5 / 5 (Exemplary)**

- Clear, well-reasoned claim that false positives threaten visas, erode trust, and expose policy gaps.
- Each assertion is supported by citations.
- No deficiencies.

Bias of AI-Detectors towards English Language Learners

5. Writing Quality – **4 / 5 (Proficient)**

- Generally polished, professional tone.
- Minor typographical errors ("Ell" vs. "ELL") and a few long sentences.

Improvement tip: run a final proofread (or a tool like Grammarly) to catch remaining slips and break up overly long sentences.

---

6. Formatting & Submission – **4 / 5 (Proficient)**

- All required sections are present; author name is included.
- Visual inconsistencies: spacing, bullet alignment, missing page numbers, and the rubric is not embedded.

Improvement tip: apply the exact course template (header/footer, uniform heading styles, page numbers) and attach/reference the rubric as required.

---

7. Appendix on AI Use – **5 / 5 (Exemplary)**

- Lists every AI tool used (Grammarly, Google, EndNote, Lumo, Perplexity).
- Provides URLs, purpose, and reflective commentary.
- Fully satisfies all sub-criteria.

Overall Score

Criterion	**Points**
Summary	4
Stakeholders & Impact	5
Broader Issues	4
Argument for Importance	5
Writing Quality	4
Formatting	4
AI-Use Appendix	5
Total	**31 / 35** (≈ 89 %)

Targeted Action Plan to Reach 35/35

1. **Summary** – Rewrite a few transition sentences; keep the word count unchanged.
2. **Broader Issues** – Insert a concise comparative example of another biased AI system (≈ 40 words).
3. **Writing Quality** – Run a spell-check and split any sentences > 30 words.
4. **Formatting** – Adopt the official rubric template: consistent headings, page numbers, and embed the PDF rubric (or a link) as the final annex.

Implementing these four quick edits should elevate every "Proficient" rating to "Exemplary", giving you a perfect score.

Bias of AI-Detectors towards English Language Learners

Bibliography

1 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

2 Blog. (2024). Evaluating AI Detection Accuracy for Non-Native English Writers. *Copyleaks*. https://copyleaks.com/blog/accuracy-of-ai-detection-models-for-non-native-english-speakers

3 Carter, T. (2023). *Some universities are ditching AI detection software amid fears students could be falsely accused of cheating by using ChatGPT*. https://www.businessinsider.com/universities-ditch-ai-detectors-over-fears-students-falsely-accused-cheating-2023-9

4 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/

5 Coldewey, D. (2023). OpenAI scuttles AI-written text detector over 'low rate of accuracy'. *TechCrunch*. https://techcrunch.com/2023/07/25/openai-scuttles-ai-written-text-detector-over-low-rate-of-accuracy/

6 Corrigan, S. (2024). Top 7 AI Tools for Language Teachers and Language Learners. *ATC Language Schools*. https://atclanguageschools.com/top-7-ai-tools-for-language-teachers-and-language-learners/

7 Epstein-Gross, C. (2023). OpenAI Abruptly Shuts Down ChatGPT Plagiarism Detector—And Educators Are Worried. *Observer*. https://observer.com/2023/07/openai-shut-ai-classifier/

8 European_Union. (2019). *Ethics guidelines for trustworthy AI | Shaping Europe's digital future*. https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai

9 Fergus, R. (2024, 2024–02–29T11:51:29–05:00). *Biased Technology: The Automated Discrimination of Facial Recognition | ACLU of Minnesota*. https://www.aclu-mn.org/en/news/biased-technology-automated-discrimination-facial-recognition

10 Forlini, E. (2023, 2023–07–25T15:57:32+00:00). *OpenAI Quietly Shuts Down AI Text-Detection Tool Over Inaccuracies*. https://www.pcmag.com/news/openai-quietly-shuts-down-ai-text-detection-tool-over-inaccuracies

11 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/

12 Guo, S., Wang, Y., Yu, J., Wu, X., Ayik, B., Watts, F. M., Latif, E., Liu, N., Liu, L., & Zhai, X. (2025). Artificial Intelligence Bias on English Language Learners in Automatic Scoring. https://doi.org/10.48550/arXiv.2505.10643

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

14 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

15 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

16 Mathewson, T. G. (2025, 2025–06–26T12:31:00+00:00). Turnitin helps colleges detect AI writing — but its cost is more than just a price tag. *CalMatters*. https://laist.com/news/education/california-colleges-spend-millions-to-catch-plagiarism-and-ai-is-the-faulty-tech-worth-it

17 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

18 Northwestern_University. (2025). *Use of Generative Artificial Intelligence in Courses: Artificial Intelligence at Northwestern - Northwestern University*. https://ai.northwestern.edu/education/use-of-generative-artificial-intelligence-in-courses.html

19 Orr, D. (2022). The Unintended Consequences of Algorithmic Bias. *Congressional Black Caucus Foundation*.

20 Quill.org. (2025). *AI Writing Check*.

21 Riess, K. (2025). OCR Weighs In On Discriminatory AI Use. *King, Spry, Herman, Freund & Faul, LLC.* https://kingspry.com/ocr-weighs-in-on-discriminatory-ai-use/

22 Rifalixa. (2023, 2023–07–25T17:16:37.000Z). *OpenAI quietly shuts down its AI detection tool due to poor accuracy*. https://www.reddit.com/r/GPT3/comments/159eoqq/openai_quietly_shuts_down_its_ai_detection_tool/

23 Rumi. (2024). *Top 20 US Schools Recommend Against AI Detectors*. https://www.rumidocs.com/newsroom/top-20-us-schools-recommend-against-ai-detectors

24 SkylineAcademic. (2025). Are AI Detectors Accurate? What My Tests On 1,000 Essays Revealed - Skyline. *Skyline Academic*. https://skylineacademic.com/are-ai-detectors-accurate-what-my-tests-on-1000-essays-revealed/

25 Stanford, L. (2025, 2025–09–15T18:20:18.167). *Connecticut Professors Fear Dependence, Cognitive Decline Over AI Use*. Courant, Hartford. https://www.govtech.com/education/higher-ed/connecticut-professors-fear-dependence-cognitive-decline-over-ai-use

26 Supiano, B. (2023, 2023–04–05T21:41:38.835). *Will ChatGPT Change How Professors Assess Learning?* https://www.chronicle.com/article/will-chatgpt-change-how-professors-assess-learning?utm_source=Iterable&utm_medium=email&utm_campaign=campaign_6572262_nl_Academe-Today_date_20230410&cid=at&source=&sourceid=&cid2=gen_login_refresh

27 University_North_Carolina_Charlotte. (2025). Dismissal, Academic Integrity, & Visa Consequences. *International Student and Scholar Office*.

28 University_of_Minnesota. (2025). *Accruing Unlawful Presence for F-1 Students | ISSS*.

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

30 VnExpress. (2025). *How an alleged AI mistake nearly cost a Japanese student his US visa - VnExpress International*. https://e.vnexpress.net/news/news/education/how-an-alleged-ai-mistake-nearly-cost-a-japanese-student-his-us-visa-4878219.html

31 Walters, W. (2023). *The Effectiveness of Software Designed to Detect AI-Generated Writing: A Comparison of 16 AI Text Detectors*. https://www.degruyterbrill.com/document/doi/10.1515/opis-2022-0158/html

32 Weixin Liang, M. Y., Yining Mao, Eric Wu, James Zou. (2023). *AI-Detectors Biased Against Non-Native English Writers*.

33 White_House. (2022). Blueprint for an AI Bill of Rights | OSTP. *The White House*.

34 Williams, R. (2025). *Algorithmic Justice League - Unmasking AI harms and biases*.