



Sandbox Series



Ian Linkletter

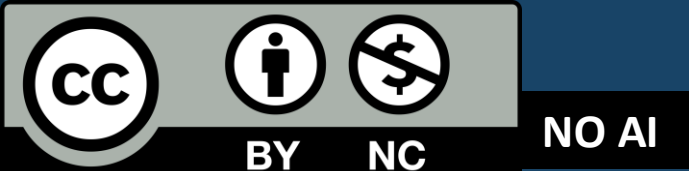
*British Columbia
Institute of Technology*



Beyond Surveillance: The Case Against AI Detection and AI Proctoring

September 18, 2024

10 a.m. - 12 p.m.



I would like to acknowledge that I live and work on the traditional, ancestral, and unceded territory of the Coast Salish Nations of Sk̓wx̓wú7mesh (Squamish), sə'liiwətaʔt̓ (Tseil-Waututh), and xwməθkwəy̓əm (Musqueam).



Session Overview

- The Case Against AI Surveillance (40-60 minutes)
- Facilitated discussions (40-60 minutes):
 - Questions and Reflections
 - Experiences and Anecdotes
 - Talking to Students about AI
 - Working Against Surveillance
 - Moving Beyond Surveillance
 - Fairness and Integrity
- Algorithmic Impact Assessment Hands-On (20 minutes)
- Reminder



My background

- The Evergreen State College, BA
- The University of Western Ontario, MLIS
- Educational technologist for 15 years
- Worked in UBC's Faculty of Education for 10 years
- Sued by Proctorio in 2020
- Now an Emerging Technology and Open Education Librarian at BCIT
- Recently created the Canadian Privacy Library
- "Do no harm" is my fundamental principle





The Case Against:

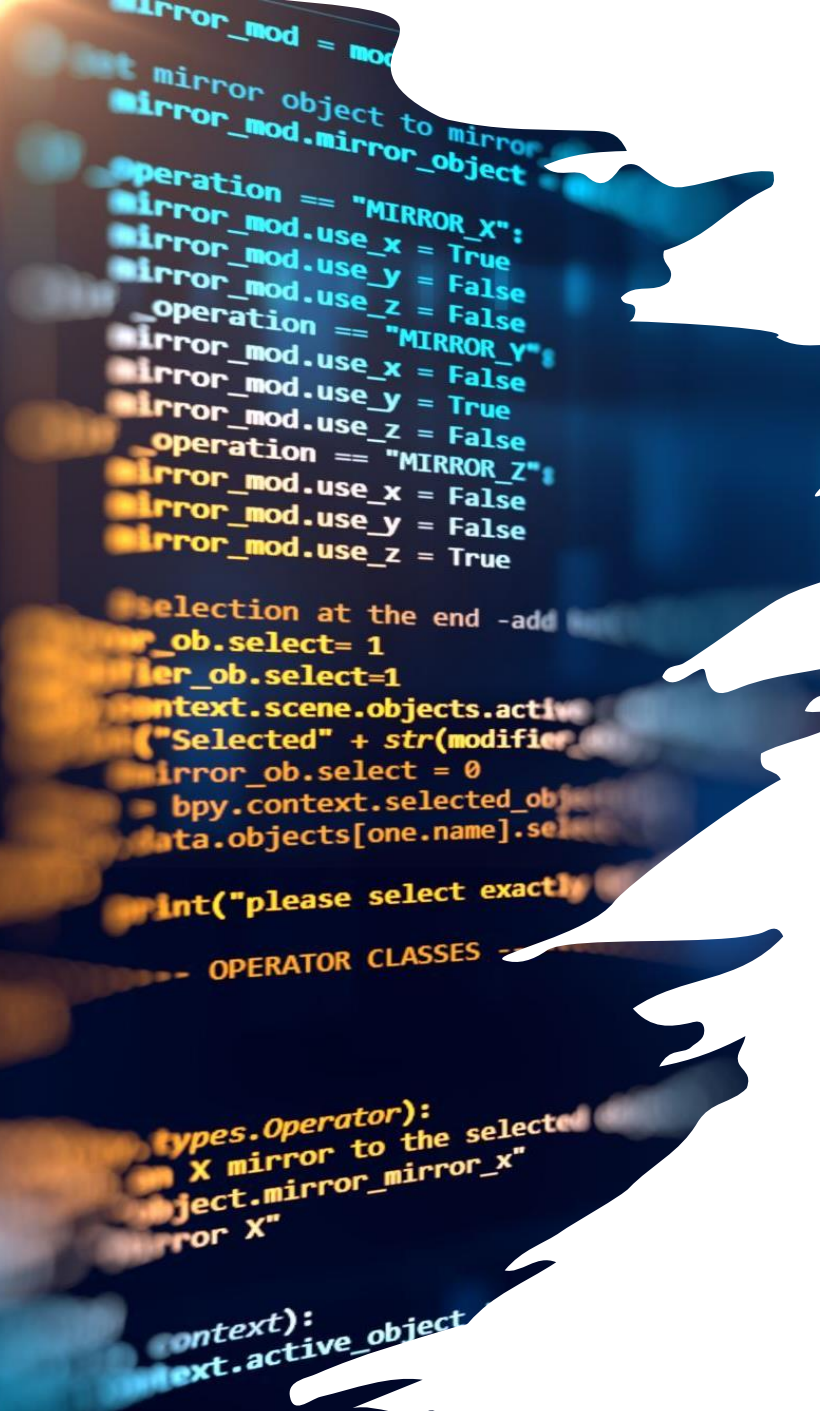
AI Proctoring and AI Detection

Links: <https://bit.ly/beyondsurveillancelinks>



Academic surveillance software

- Academic surveillance software monitors student behaviour
- Educational technology serves a pedagogical purpose
- Surveillance technology has no place in education
- We can work together against surveillance
- 2024 can be the year we move **beyond surveillance**





Surveillance
demonstrates values

Power and control over
teaching and learning

A classroom without
mutual trust is not a good
place to be



Surveillance sends a message

- You are not trusted.
- We are watching you.
- You can be removed.

Ethical criticisms of GenAI apply to AI surveillance

1. Model Bias
2. Environmental Impact
3. Privacy Concerns
4. Copyright



Two useful libraries by Charles Logan

- Against School Surveillance Technologies Library:
<https://docs.google.com/spreadsheets/d/1jdulo4LzswN6xYCI4GvnpCdBdpu6zIPfi01yDOBwDU/edit?gid=572288548#gid=572288548>
- Against AI and its Environmental Harms Library:
https://pad.riseup.net/p/Against_AI_and_Its_Environmental_Harms-keep

Academic Surveillance Software

AI Proctoring

- Mandated spyware used during remote exams
- Exploded in use during the pandemic
- Monitors bodies and behaviour
- Surveils webcam, microphone, web browser
- Room scans were found [unconstitutional](#)
- A spectrum from 100% AI to human-assisted AI

Source: <https://www.npr.org/2022/08/25/1119337956/test-proctoring-room-scans-unconstitutional-cleveland-state-university>

Synthetic image.



Abnormalities

- Compares each test taker's behaviour
- Outliers get a higher suspicion score
- Outliers are flagged



Abnormalities

"One system, Proctorio, uses gaze-detection, face-detection and computer-monitoring software to flag students for any “abnormal” head movement, mouse movement, eye wandering, computer window resizing, tab opening, scrolling, clicking, typing, and copies and pastes. A student can be flagged for finishing the test too quickly, or too slowly, clicking too much, or not enough."

Source (Gift Link): <https://wapo.st/4esCN33>

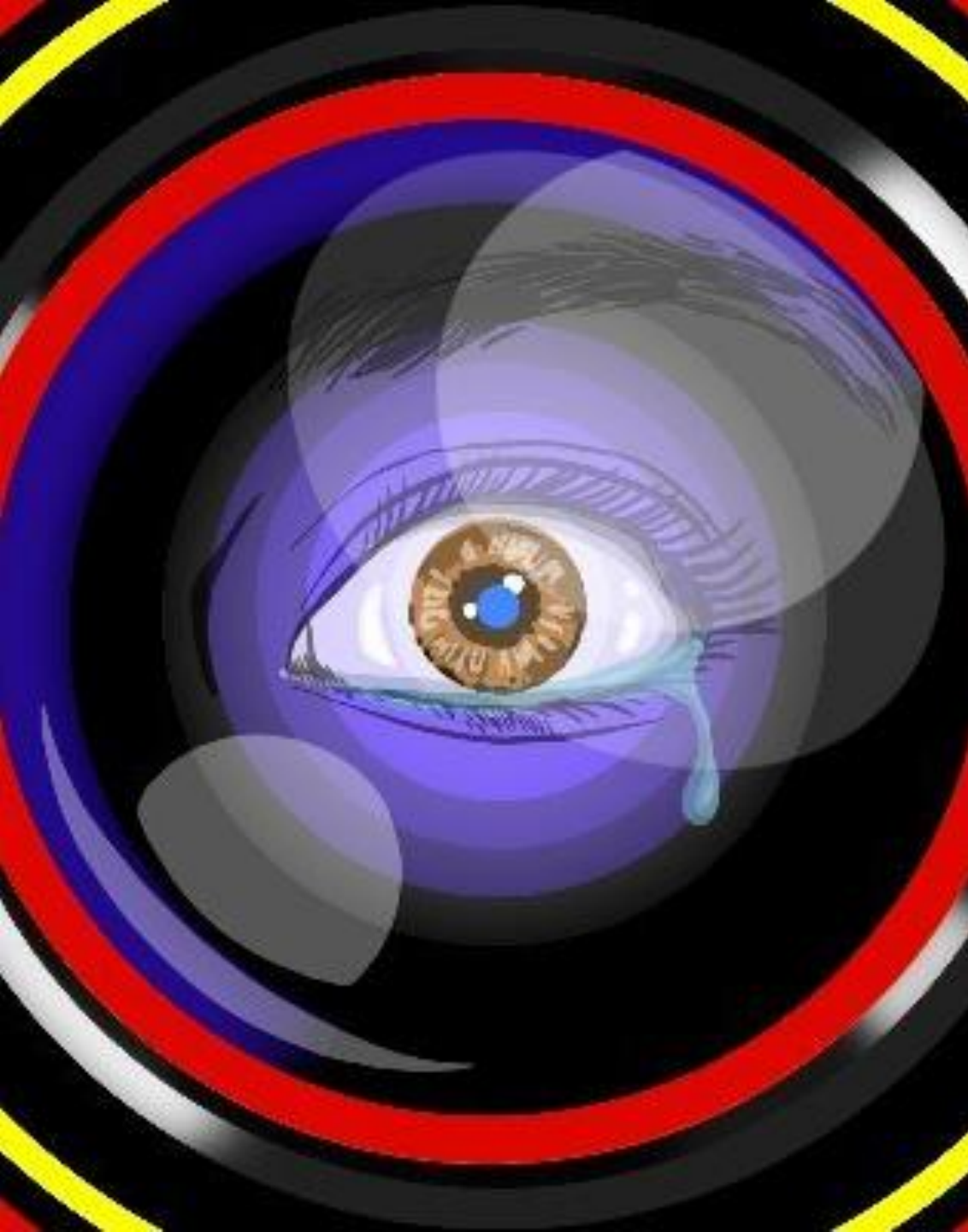


Abnormalities

- Changes in audio levels
- Clicking
- Copy and paste
- Exam duration
- Eye movement
- Head movement
- Number of faces
- Mouse movement
- Keystrokes
- Reloading
- Switching networks

Source: <https://twitter.com/ejohnson99/status/1303142450828054529>





AI makes proctoring especially harmful

- Discrimination against disabled students.
- Discrimination against students of colour.
- Unwarranted invasion of privacy.
- Causes emotional harm.

Source: <https://www.twitter.com/Procteario>

Discriminatory algorithms

Inaccessible to blind students (room scan and screen readers)

Kicks you out if you are away from your computer (i.e. bathroom break)

Eye and head movement tracking targets neurodiverse students

Minimal movement causes stress

Kicks you out if it detects another face



Facial detection software is biased

Face not found: Students of colour have for years reported problems being recognized by the facial detection software and being blocked, flagged, or removed from exams as a result.

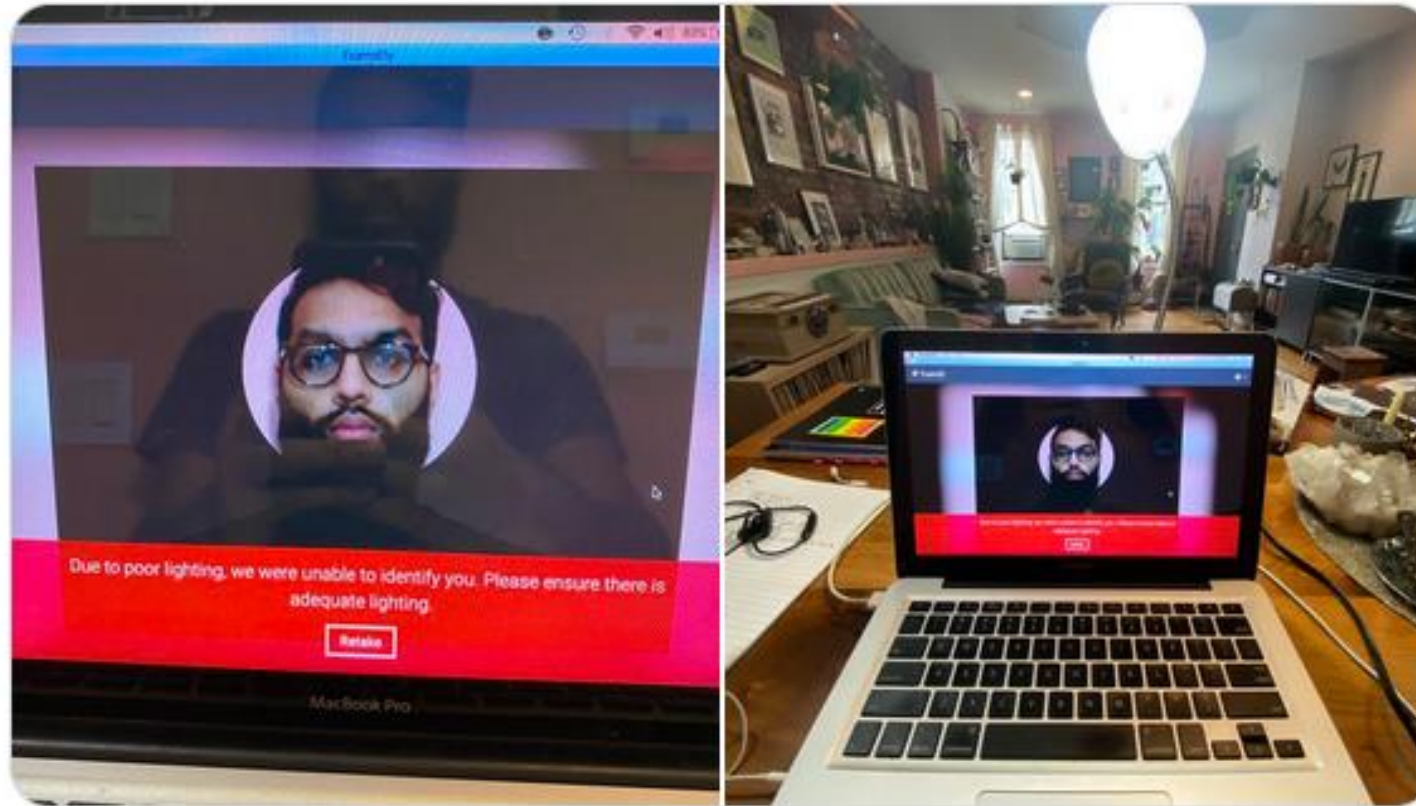
"Imagine all you want to do is take a test, and the system your institution uses as a gateway to testing doesn't recognize you as a human being." - Dr. Chris Gilliard

Find Chris on Bluesky: <https://bsky.app/profile/hypervisible.bsky.social>



@uhreeb

Ok @ExamSoft support told me to “sit directly in front of a lighting source such as a lamp.” I’m receiving the same issue preventing me from completing the NY UBE mock exam. Facial recognition technology is racist. @DiplomaPriv4All do y’all think I have “adequate lighting”?



9:05 AM · Sep 11, 2020

Source: <https://x.com/uhreeb/status/1304451031066083331?lang=en>

Robin Pocornie's story



Error 404: Human Face Not Found | Robin Aisha Pocornie | TEDxAmsterdam



TEDx Talks ✓
41.5M subscribers

Subscribe

44



Share



Video: <https://www.youtube.com/watch?v=pVfvYYUkIcY>



Amaya Ross's story



Video: <https://www.youtube.com/watch?v=iELUtvhbukE>





glazed donut 🍩

@WHO_isKenny



This happens to me using @proctorio 😡



Janice Wyatt-Ross, EdD @JaniceWyattRoss · Feb 22, 2021

Daughter 1 was taking an exam today being proctored by some type of software that apparently was not tested on dark skin. She had to open her window, turn on the lights, and then shine a flashlight over her head to be detectable. 😡😡😡

8:38 AM · Feb 23, 2021



sudo stallion ❤️

@cham_omot



THIS! There's no reason I should have to collect all the light God has to offer, just for Proctorio to pretend my face is still undetectable 🥲



Janice Wyatt-Ross, EdD @JaniceWyattRoss · Feb 22, 2021

Daughter 1 was taking an exam today being proctored by some type of software that apparently was not tested on dark skin. She had to open her window, turn on the lights, and then shine a flashlight over her head to be detectable. 😡😡😡

4:46 PM · Feb 23, 2021

Gavin Gordon had the same experience

"During the pre-test tech check, the software was unable to recognize my face. Being fairly experienced with technology, I tried the usual troubleshooting procedures: using a solid background, making sure I wasn't backlit, increasing the lighting on my face, trying different angles, etc. Yet it still wasn't recognizing my face. To ensure that this was not a problem with my webcam, I had my Caucasian roommate try the tech check which instantly recognized his face without any issue... I know that if it happened to me, it's likely happening to other students."

Femi Yemi-Ese also experienced face not found

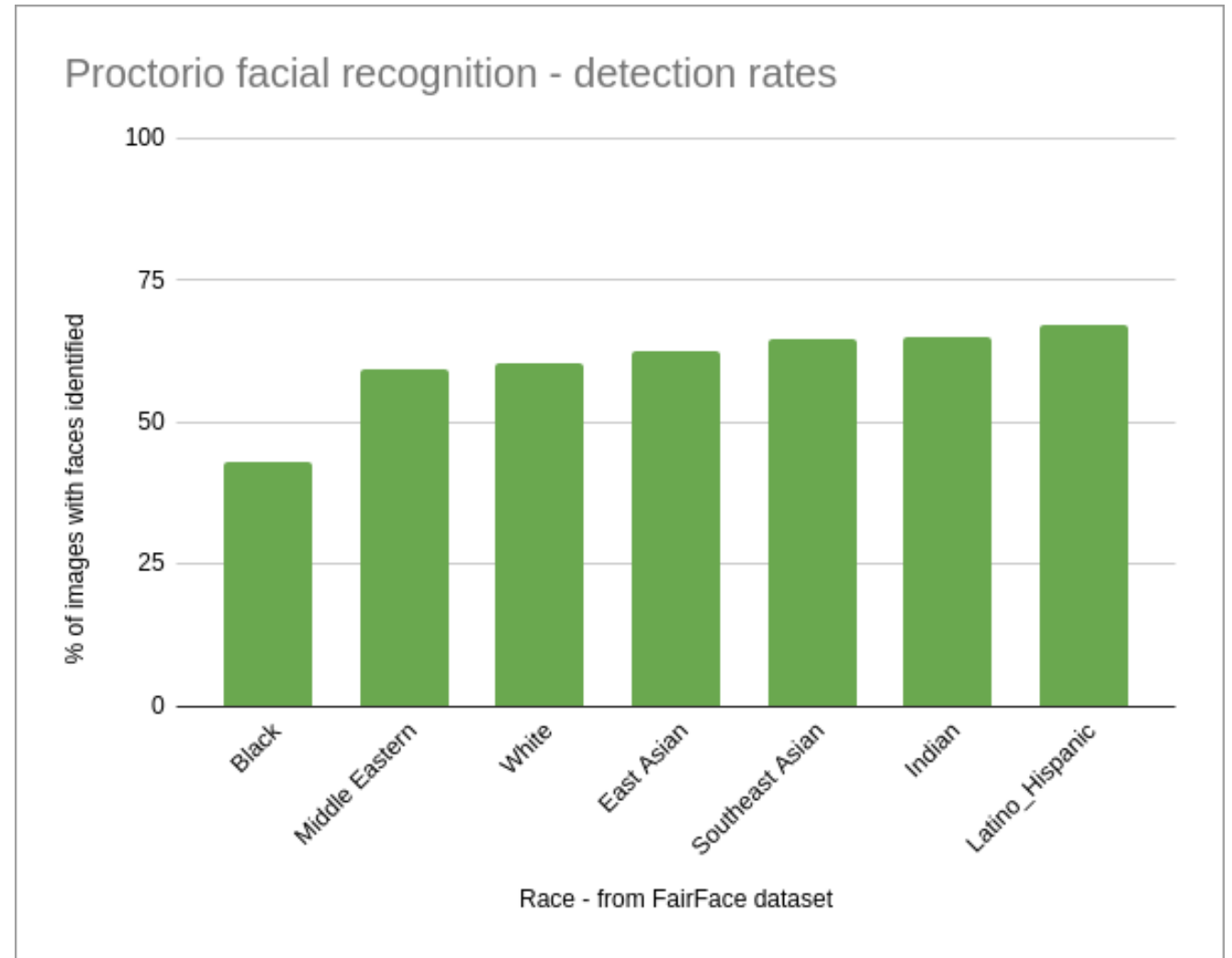
"Like many test-takers of color, Yemi-Ese, who is Black, has spent the past three semesters using software that reliably struggles to locate his face. Now, whenever he sits down to take an exam using Proctorio, he turns on every light in his bedroom, and positions a ring light behind his computer so that it shines directly into his eyes. Despite these preparations, "I know that I'm going to have to try a couple times before the camera recognizes me," he said... Adding sources of light seems to help, but it comes with consequences. "I have a light beaming into my eyes for the entire exam," he said. "That's hard when you're actively trying not to look away, which could make it look like you're cheating.""

Source: <https://www.newyorker.com/tech/annals-of-technology/is-online-test-monitoring-here-to-stay>

The root of the problem: biased AI

Lucy Satheesan discovered that Proctorio was using open source facial recognition software called Open CV. She tested Open CV against an open source database of faces called FairFace and found that Proctorio could detect Black faces less than half of the time.

Source: <https://proctor.ninja/proctorios-facial-recognition-is-racist>



TECH / POLICY

Students of color are getting flagged to their teachers because testing software can't see them / Proctorio reportedly uses facial detection software that failed to recognize black faces over half the time


By [Mitchell Clark](#)






Apr 8, 2021, 5:34 PM PDT

Link: <https://www.theverge.com/2021/4/8/22374386/proctorio-racial-bias-issues-opencv-facial-detection-schools-tests-remote-learning>

Proctorio Is Using Racist Algorithms to Detect Faces




A student researcher has reverse-engineered the controversial exam software—and discovered a tool infamous for failing to recognize non-white faces.

 Listen to this article now 5 min listen

00:00      -04:37

Powered by [Trinity Audio](#)

 By [Todd Feathers](#)

April 8, 2021, 9:49am  [Share](#)  [Tweet](#)  [Snap](#)

Link: <https://www.vice.com/en/article/g5gwg3/proctorio-is-using-racist-algorithms-to-detect-faces>

This problem is known within the industry

- "Respondus, a vendor of online proctoring software, has been granted a patent for their “systems and methods for assessing data collected by automated proctoring.” The patent shows that their example method for calculating a risk score is adjusted on the basis of people’s skin colour.
- The “Adjustment to Final Tally” for dark skin is described as follows in the granted patent:

For dark complexion persons, a racial detection feature may be provided so that a downward adjustment can be made to the final risk tally."

Questions for our group conversation

Can an algorithm be racist?

How many "face not found" errors are acceptable?

What does it mean to say "AI proctoring is racist?"

How are our institutions protecting students from discriminatory AI?

What does accountability look like?

Proctorio Spend at Ohio State University

Purchase orders:

03/19/2018: \$220,000

08/07/2018: \$9,000

03/05/2019: \$220,000

09/09/2020: \$170,000

07/01/2022: \$170,000

04/19/2023: \$125,000

Overall spend: \$465,531.25

Source: Personal Freedom of Information request



Proctorio Spend at University of Colorado Boulder

Purchase orders:

2015: \$50,000
2018: \$4,290
2019: \$4,896
2020: \$180,448
2021: \$180,448
2022: \$114,448
2023: \$66,732.52
2024: \$114,448

Total: \$722,491.50

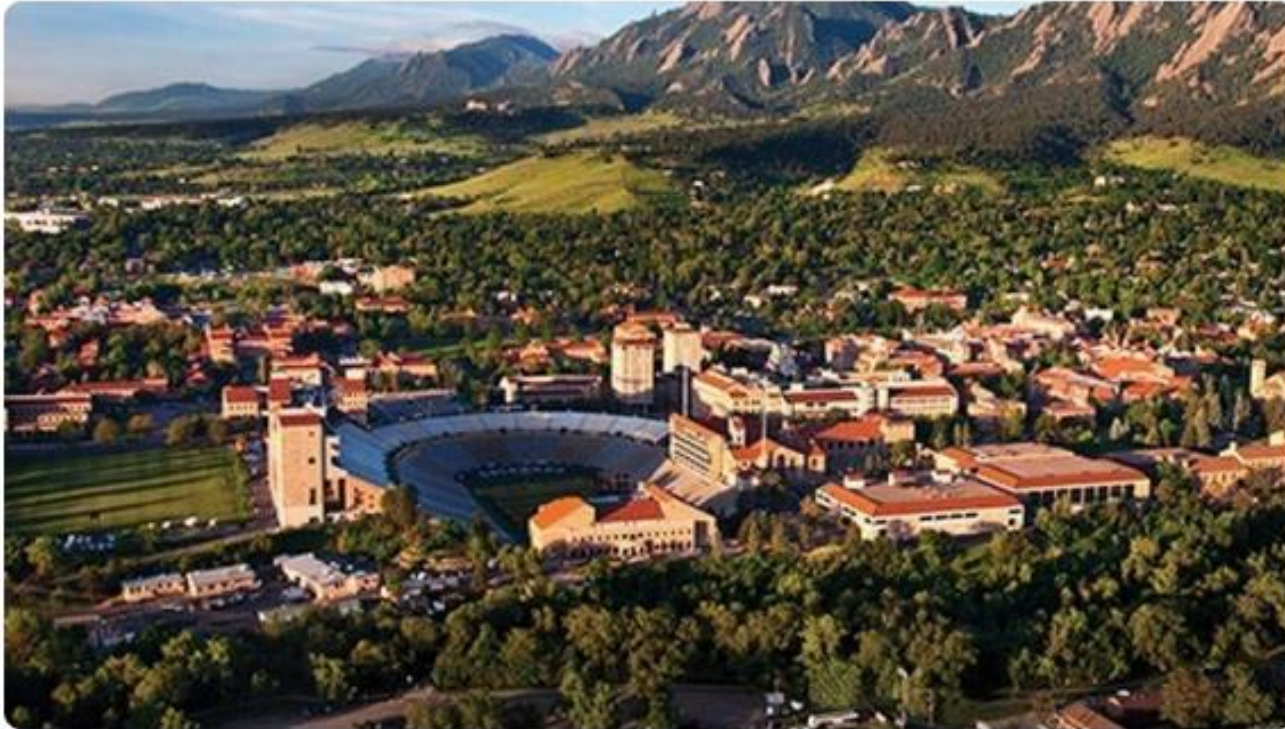
Pricing: \$3.68 per user based on the institution's current full-time equivalent enrollment number.

5% of instructors used Proctorio when the institutional contract was discontinued.

Source: Personal Freedom of Information request and <https://oit.colorado.edu/services/consulting-professional-services/academic-technology-initiatives-team/needs-assessments/online-proctoring>



Stop Use of Proctorio at CU Boulder over Privacy Concerns



Started
Petition to

September 6, 2020
[University of Colorado Boulder](#)

1,403

Signatures

1,500

Next Goal

 Support now

Sign this petition

First name

Last name

Email



Source: <https://www.change.org/p/university-of-colorado-boulder-stop-use-of-proctorio-at-cu-boulder-over-privacy-concerns>



Office of Information Technology

Proctorio

Snipping Tool

Services / Teaching & Learning Applications / Proctorio

FEATURES

GET STARTED

COST

SUPPORT

Proctorio to be retired August 2024

Proctorio online proctoring software will be retired on **August 15, 2024**. Proctorio was made available as a measure to help ensure academic integrity during the COVID-19 pandemic. Due to Proctorio's overall low usage (5% of all courses), CU Boulder leadership has made the decision to retire the product as a supported campuswide service. OIT is currently evaluating online proctoring needs to inform our long term online proctoring strategy.

Instructors will still be able to utilize Proctorio through individual or departmental arrangements.

Source: <https://oit.colorado.edu/services/teaching-learning-applications/proctorio> (Retrieved July 2024)



Proctorio Spend at UBC

Overall spend:

FY20 (ended March 31, 2020): \$40,077

FY21 (ended 2021): \$286,140

FY22 (ended 2022): Free

Source:

<https://finance.ubc.ca/sites/finserv.ubc.ca/files/FY20%20UBC%20Statement%20of%20Financial%20Information.pdf> and

<https://finance.ubc.ca/sites/finserv.ubc.ca/files/FY21%20UBC%20Statement%20of%20Financial%20Information.pdf>



Message to the community regarding Senate vote on use of remote proctoring software

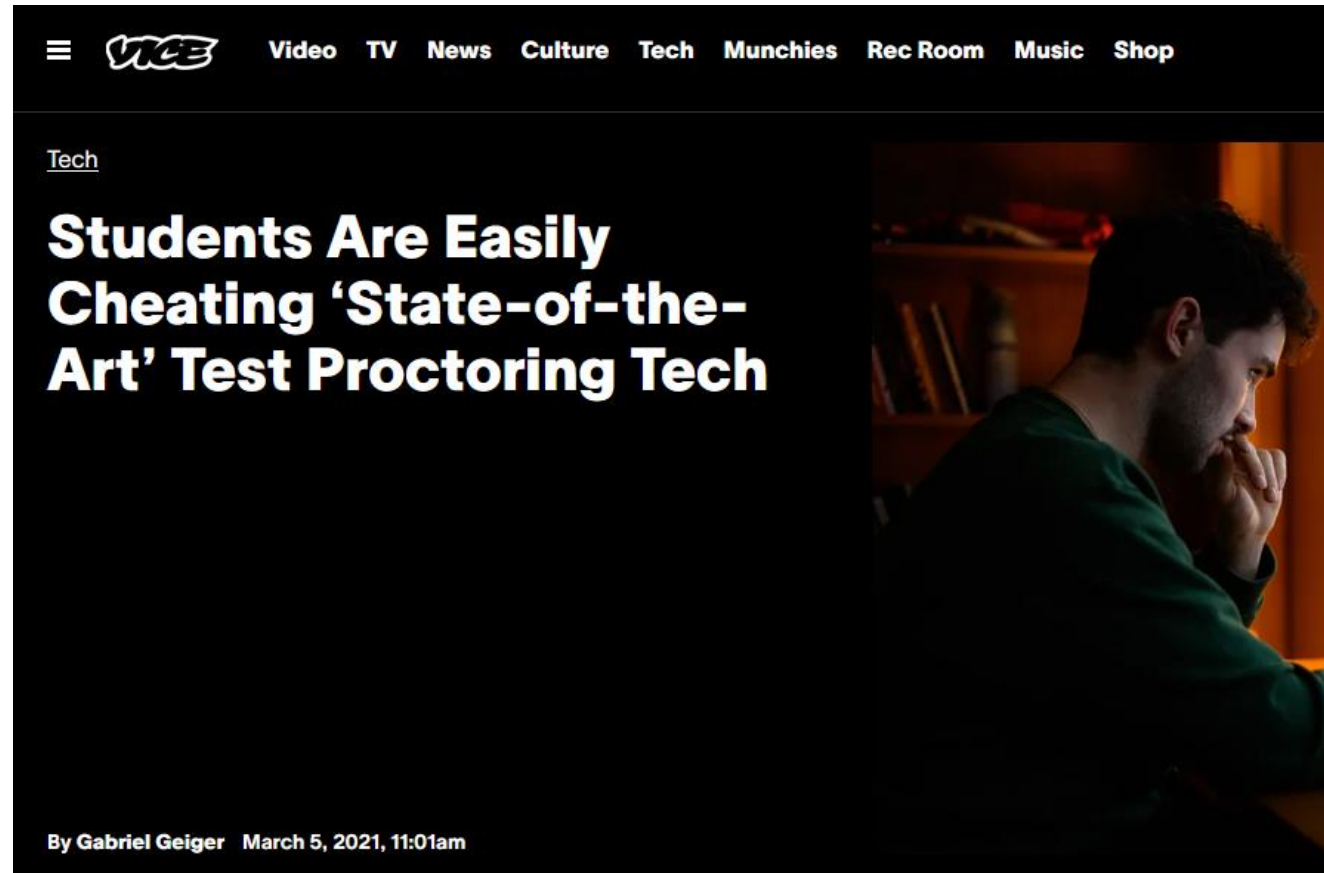
March 18, 2021

Last evening, the UBC Vancouver Senate voted in favour of a motion to direct UBCV Faculties to stop using remote invigilation tools that involve automated recording and algorithmic analysis of data captured during invigilation, except in cases where the use of such 'remote proctoring software' is required by external accreditation bodies. This motion, which is effective immediately, comes after much discussion and consultation with stakeholders across campus, including students and faculty.

Source: <https://academic.ubc.ca/academic-community/news-announcements/news/senate-vote-remote-proctoring-software>



Does this million dollar software actually stop academic dishonesty?



Source: <https://www.vice.com/en/article/students-are-easily-cheating-state-of-the-art-test-proctoring-tech/>



Does this million dollar software actually stop academic dishonesty?

On the Efficacy of Online Proctoring using Proctorio

Laura Bergmans, Nacir Bouali, Marloes Luttikhuis and Arend Rensink
University of Twente, Enschede, The Netherlands

- "The experiment involved 30 students, 6 of which were asked to cheat in various ways, while 5 others were asked to behave nervously but make the test honestly..."
- The most important findings were that none of the cheating students were flagged by Proctorio, whereas only one (out of 6) was caught out by an independent check by a human agent. The sensitivity of Proctorio, based on this experience, should therefore be put at very close to zero...
- The use of online proctoring is therefore best compared to taking a placebo: it has some positive influence, not because it works but because people believe that it works, or that it might work."

Source:

<https://ris.utwente.nl/ws/portalfiles/portal/275927505/3e2a9e5b2fad237a3d35f36fa2c5f44552f2.pdf>



Listen to students

Read what they have to say:

- <https://www.twitter.com/Procteario>
- <https://www.twitter.com/ProcterrorU>

Read thousands of student reviews:

- <https://googlereviews.linkletter.org>

Share these voices with people who may not otherwise know what students are experiencing. They contain evidence of harm.



What about lockdown browsers?

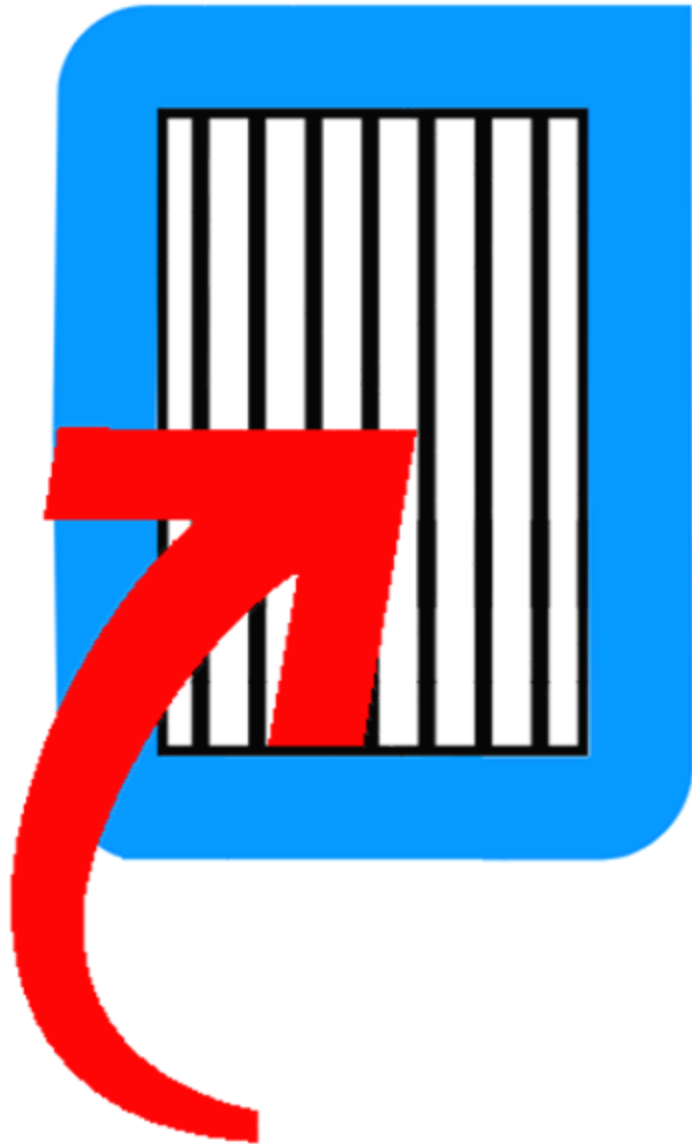
Technical issues abound

Compatibility issues with accessibility software and extensions

Use of second devices not prevented

Potential for data breaches





Academic Surveillance Software

Plagiarism Detection

- Assignments are scanned for similarity
- Student work becomes part of the database
- Internet content is also searched for similarity
- Business model in danger because of ChatGPT
- Turnitin added AI to the product without warning
- Over 200 million papers have been scanned since release

How do AI detectors work?

- Use natural language processing to evaluate text perplexity.
- Perplexity means how surprising the word choice is in an essay.
- The more common your words, the lower the perplexity.
- LLMs like ChatGPT 3 and 3.5 generate low perplexity text by default, but you can instruct them not to. Perplexity is increasing with more recent models.



Turnitin's AI detector

- Released in April 2023 (but many opted out)
- The promise was a 1% rate of false positives
- The truth is that false positives are higher in the real world
- Sentence positive rate is actually 4%
- We tested a new ChatGPT-detector for teachers. It flagged an innocent student. - Washington Post testing found it to be 1 in 16

Source: <https://www.turnitin.com/blog/ai-writing-detection-update-from-turnitins-chief-product-officer> and <https://wapo.st/3Tz8zTY>



According to Turnitin: how does it work?

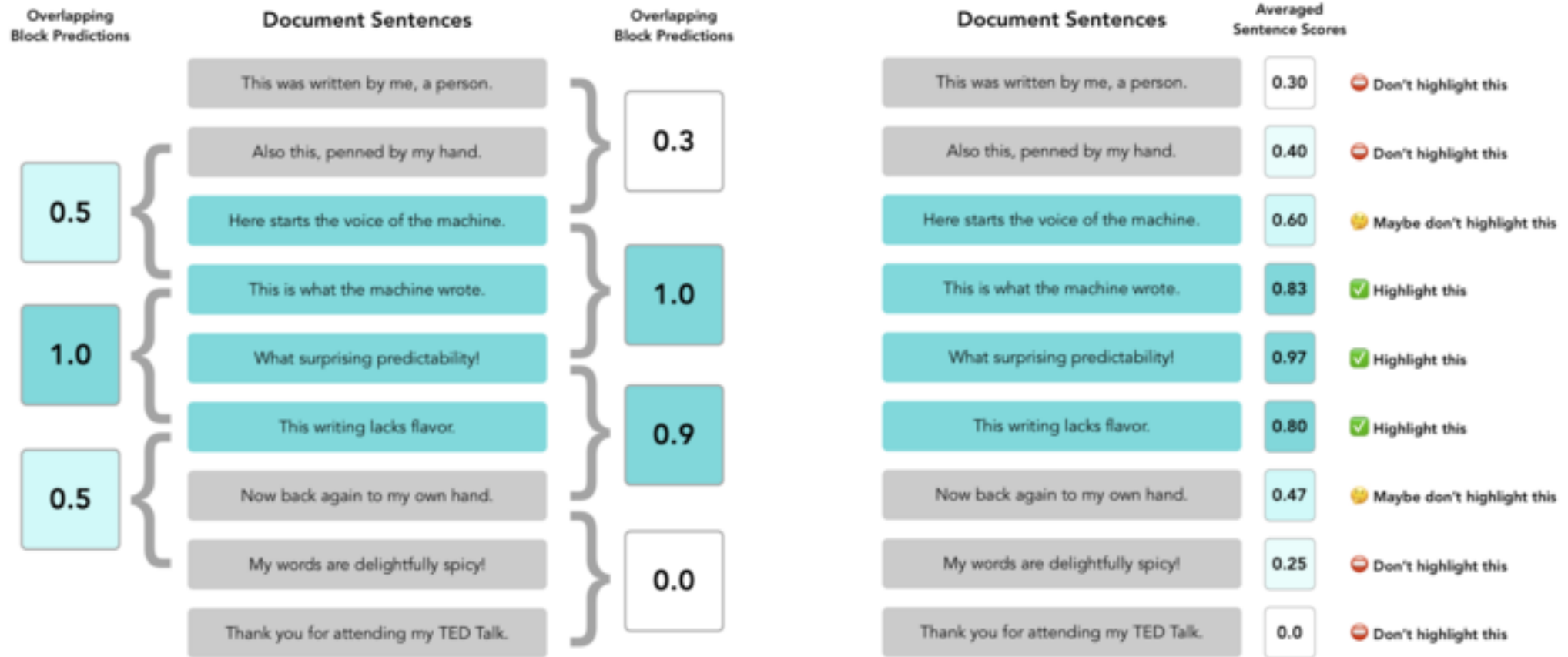
“When a paper is submitted to Turnitin, the submission is first broken into segments of text that are roughly a few hundred words (about five to ten sentences). Those segments are then overlapped with each other to capture each sentence in context.

The segments are run against our AI detection model and we give each sentence a score between 0 and 1 to determine whether it is written by a human or by AI. If our model determines that a sentence was not generated by AI, it will receive a score of 0. If it determines the entirety of the sentence was generated by AI it will receive a score of 1.

Source: https://guides.turnitin.com/hc/en-us/articles/28477544839821-Turnitin-s-AI-writing-detection-capabilities-FAQs#h_01J2CQS37K893R6BYYQ1QTGCMK



According to Turnitin: how does it work?



Source: https://guides.turnitin.com/hc/en-us/articles/28477544839821-Turnitin-s-AI-writing-detection-capabilities-FAQs#h_01J2CQS37K893R6BYYQ1QTGCMK

Turnitin demo

turnitin

1 of 17: Lucas Anderson

Navigating the Complex Land...

Download Settings Details Help

Similarity 23% AI Writing 56% Flags 0

The pervasive presence of social media platforms in contemporary society has fundamentally altered the dynamics of human interaction and communication. From the proliferation of Facebook to the rise of Instagram influencers, social media has become an integral aspect of daily life for billions worldwide. However, amid the convenience and connectivity afforded by these platforms lies a nuanced relationship with mental health that warrants deeper exploration. This essay embarks on a comprehensive journey to dissect the multifaceted impact of social media on mental well-being, delving into its intricate layers and divergent effects.

Despite the myriad challenges posed by social media, it is essential to acknowledge its potential for positive impact on mental health when utilized mindfully. Online support groups and communities offer invaluable resources and solidarity for individuals facing mental health challenges, providing a sense of belonging and understanding in times of need. Additionally, social media platforms serve as powerful tools for raising awareness about mental health issues and reducing stigma through advocacy efforts and storytelling.

To mitigate the negative effects of social media on mental health, individuals can adopt proactive strategies to cultivate a healthy relationship with these platforms. Setting boundaries around social media usage, such as limiting screen time and prioritizing offline interactions, can help restore

addition. Moreover, promoting digital literacy and

critical thinking skills can empower individuals to navigate the digital landscape with greater

56% detected as AI

The percentage indicates the combined amount of likely AI-generated text as well as likely AI-generated text that was also likely AI-paraphrased.

Submission Breakdown

page 1 page 4

- 1 AI-generated only 24%
Likely AI-generated text from a large-language model.
- 2 AI-generated text that was AI-paraphrased 32%
Likely AI-generated text that was likely revised using an AI-paraphrase tool or word spinner.

FAQs Resources Guides

View FAQs View guides

Hide Disclaimer

Our AI writing assessment is designed to help educators identify text that might be prepared by a generative AI tool. Our AI writing assessment may not always be accurate (it may misidentify writing that is likely AI-generated as AI-generated and AI-paraphrased or likely AI-generated and AI-paraphrased writing as only AI-generated) so it should not be used as the sole basis for adverse actions against a student. It takes further scrutiny and human judgment in conjunction with an organization's application of its specific academic policies to determine whether any academic misconduct has occurred.

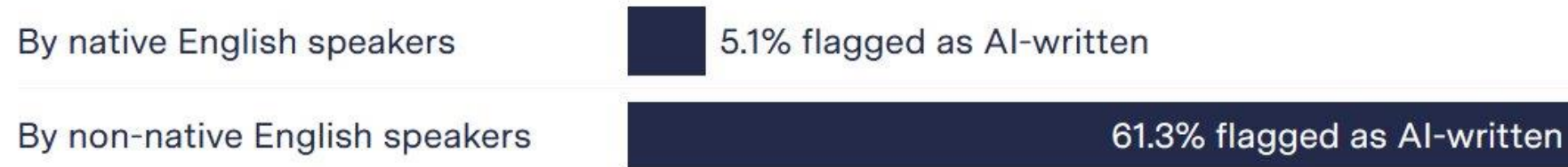
Page 1 of 4 738 words 100%

<https://go.turnitin.com/new-similarity-report>

Writing by non-native English speakers more frequently confused with AI

Seven AI detectors frequently misclassified writing by non-native English speakers. Changing the complexity of the vocabulary affected the AI's error rate.

Human-written text



Modified human-written text



Note: Average misclassification of seven AI detectors.

Chart: Joel Eastwood • Source: [W. Liang et al.](#)

Link: <https://themarkup.org/machine-learning/2023/08/14/ai-detection-tools-falsely-accuse-international-students-of-cheating>

How does Turnitin mitigate bias?

“Our model is trained on a representative sample of data spread over a period of time, that includes both AI generated and authentic academic writing across geographies and subject areas. While creating our sample dataset, we also took into account statistically under-represented groups like second-language learners, English users from non-English speaking countries, students at colleges and universities with diverse enrollments, and less common subject areas such as anthropology, geology, sociology, and others to minimize bias when training our model.”

Source: https://guides.turnitin.com/hc/en-us/articles/28477544839821-Turnitin-s-AI-writing-detection-capabilities-FAQs#h_01J2CQS37K893R6BYYQ1QTGCMK



Turnitin on decision-making

"We're not quite sure. We won't say much about the parts where it may be transitioning between human writing and AI writing. It's a fuzzy boundary, **you don't want to do any harm by saying the wrong thing.**"

- David Adamson, Turnitin AI Scientist

<https://www.youtube.com/watch?t=96&v=g85aB8qaSGc&feature=youtu.be>

Students know AI detectors don't work

- "complain to the dean. your professor is using an unreliable tool to avoid fairly grading your work. "
- "Run some of your professor's works through an AI detector :) "
- "make sure you complain in writing, and use terms like "hostile environment" and "creating an unnecessarily adversarial relationship between student and professor" and of course, "I no longer have an expectation of my work being evaluated fairly". "
- "This is the blog post in which OpenAI announced its own detection tool, complete with the update that they shut it down due to low accuracy:
<https://openai.com/blog/new-ai-classifier-for-indicating-ai-written-text>"

Source:

https://www.reddit.com/r/ChatGPT/comments/17gj5ky/my_professor_said_turnitin_rated_my_essay_62_ai/

UBC hasn't enabled AI-detection feature

Rationale

- "UBC has not been able to review and validate the feature
- Testing for accuracy in the feature is in early stages
- Instructors cannot double-check the feature results
- Testing for potential bias in the feature is also in early stages
- Results from the feature are not available to students
- Ability of the feature to keep up with rapidly evolving AI is unknown"

Source: <https://lthub.ubc.ca/2023/08/28/ubc-affirms-decision-to-not-enable-turnitin-ai-detection/>

Students can easily circumvent AI detectors... if they have the resources

- Humanizers, word spinners, and rewriters abound
- Students get access to their own AI detector to check and recheck their work until AI is not detected
- Cutting edge models are less likely to be detected
- Students with money and digital literacy have an advantage



AI detectors have no place in education

- Even a 1% rate of false positives is unacceptable
- False negatives exacerbate inequity
- Student assignments are confidential
- Students hold copyright on their work
- In 2023, Turnitin's AI detector was banned at BCIT
- LLM development far exceeds Turnitin development
- Don't buy into the arms race

Still going to use AI detectors? 🙄

Ethical Principles for Detecting AI-Generated Text in Student Work

Sarah Elaine Eaton, PhD

I am not a fan of using detection tools to try to identify text generated by artificial intelligence apps in student work, but if you are going to use them, here are some guidelines to do so ethically and transparently.

Covert or deceptive use of any kind of detection tools is antithetical to assessing student work with integrity.

If we expect students to act with integrity, then educators must model ethical behaviour themselves.

Policy Check

Before you use any detection apps check to ensure their use is permitted by your institution.

Departmental Check

Check with your department head or supervisor to ensure that they have no objections to the use of detection tools for student assessment.

Course Syllabus

Declare the use of any detection tools in writing in your course outline. Know that having a statement in your course outline does not replace the need for open conversations with students.

Talk to Students

Have open and honest conversations with students about how you plan to use detection tools. Do this before using the tools, so students are not surprised. Be upfront and clear.



May 5, 2023

<https://drsaraheaton.com/2023/05/06/the-use-of-ai-detection-tools-in-the-assessment-of-student-work/>

Thank you!

This concludes the presentation. Do you have any questions for me? We have some time now.

Please feel free to email me at Ian_Linkletter@bcit.ca or connect with me on LinkedIn, Bluesky, or Mastodon.

Now let's talk about privacy, AI, integrity, and more.

