# Late Applications

## Disproportionate Effects of Generative AI-Detectors on English Learners

**C**DT recently released [legal research](#) on the application of civil rights laws to uses of education data and technology, including AI. As the use of generative AI increases both inside and outside the classroom, one group of students at particular risk of unequal treatment are those who are not yet able to communicate fluently or learn effectively in English – that is, English Learner (EL) students. Research indicates that so-called AI detectors are disproportionately likely to falsely flag the writing of non-native English speakers as AI-generated, putting them at greater risk for being disciplined for cheating in school. Schools need to be aware of this potential disparity and take steps to ensure it does not result in violating the civil rights of EL students.

## Who Are EL Students?

Nationally, English learners (ELs) are the [fastest growing student population](#), accounting for 10 percent of the overall student population in 2019, with [81 percent](#) of public schools serving at least one EL student. While some EL students are immigrants themselves, most are actually the [U.S.-born children of immigrants](#). Both face unique challenges in school. For example, non-U.S. born ELs who enter the K-12 system as high schoolers are under immense pressure to graduate on time while also reaching English language proficiency; they may also have entered the U.S. without their family, meaning that they [bear significant burdens](#) such as unstable housing and the obligation to work to support themselves.

The goal for all ELs is to reach English proficiency – once they achieve this, they are reclassified and no longer considered ELs. This reclassification process makes ELs a dynamic student group who are more difficult than other vulnerable student populations to properly track. By 12th grade, [ELs make up only 4 percent](#) of the total population of students, down from 16 percent in kindergarten. Even after reclassification, however, [studies have historically suggested](#) that EL students still struggle – "sizable proportions of the reclassified students, while able to keep pace in mainstream classrooms in the early elementary school years, later encountered difficulties in middle and high school," with some ending up having to repeat a grade. [Data out of California](#) shows ELs lagging behind their peers academically, from test scores to grades to graduation rates. However, some advocates are optimistic that ELs, with the right support and tracking, are [closing this gap](#).

# Generative AI, EL Students, and the Risk of Disproportionate Discipline

EL students already are at higher risk for school discipline. The risk of suspension for a student with EL status is 20 percent higher than a non-EL student. Moreover, approximately three quarters of EL students are native Spanish speakers, and Hispanic students are overrepresented in alternative schools, where students are typically placed due to disciplinary issues and where they tend to have less access to support staff like counselors and social workers. CDT research also found that Hispanic students are more likely than non-minority students to use school-issued devices, and thus more likely to be subject to continuous monitoring by student activity monitoring software, which can lead to even higher rates of discipline.

The increased use of chatbots such as ChatGPT threatens to exacerbate the discipline disparity for EL students. Generative AI has become a contentious topic in the education sector. Concerns about academic dishonesty are high, with 90 percent of teachers reporting that they think their students have used generative AI to complete assignments. As CDT has previously reported, student accounts suggest that generative AI is actually primarily used for personal reasons rather than to cheat, and that certain populations, such as students with disabilities, are more likely to use the technology and more likely to have legitimate accessibility reasons for doing so. Still, disciplinary policies are cropping up across the country to penalize student use of generative AI and are sometimes accompanied by newly acquired programs that purport to detect the use of generative AI in student work.

For EL students, this could be uniquely problematic. A recent study out of Stanford University shows that AI-detectors are very likely to falsely flag the writing of non-native English speakers as AI-generated, and that there is significant disparity in false flags for non-native English speakers versus native speakers. The study was conducted using the test of English as a foreign language (TOEFL) done by eighth graders. Detectors were "near perfect" in evaluating essays written by U.S. born writers, but falsely flagged 61.22 percent of TOEFL essays written by non-native English speakers as AI-generated (particularly troubling as this is a test that would, by its nature, not ever be administered to native English speakers in the first place). All seven AI detectors that the study tested unanimously but falsely identified 18 of the 91 TOEFL student essays (19 percent) as AI-generated and a remarkable 89 of the 91 TOEFL essays (97 percent) were flagged by at least one of the detectors. James Zou, who conducted the study, said of its results: "These numbers pose serious questions about the objectivity of AI detectors and raise the potential that foreign-born students and workers might be unfairly accused of or, worse, penalized for cheating."

Like students with disabilities, there might be legitimate uses of generative AI that could benefit EL students in ways that might make them more likely users, and thus even more likely to be disciplined under new school policies. According to some EL educators, generative AI "can potentially address some of the pressing needs of second language writers, including timely and adaptive feedback, a platform for practice writing, and a readily available and dependable writing assistant tool." Some say that generative AI could benefit both students and teachers in the classroom, by providing students with engaging and personalized language learning experiences, while allowing teachers to "help students improve their language skills in a fun and interactive way, while also exposing them to natural-sounding English conversations."

# Civil Rights Considerations

These concerns about disproportionate flagging and discipline are not just a matter of bad policy. Where students belonging to a protected class are being treated differently from others because of their protected characteristics, civil rights alarm bells sound. The Civil Rights Act of 1964 (the Act) generally prohibits state-sponsored segregation and inequality in crucial arenas of public life, including education. Title VI of the Act protects students from discrimination on the basis of, among other attributes, race, color, and national origin, and was enacted to prevent (and in some cases, mandate action to actively reverse) historical racial segregation in schools. ELs are protected from discrimination under Title VI on the basis of both race and national origin, and are entitled to receive language services and specialized instruction from their school in the "least segregated" manner possible. Under the circumstances described above, EL students arguably experience unlawful discrimination under the theories of disparate treatment, disparate impact, or hostile learning environment as a result of false flagging.

i. **Disparate impact and disparate treatment.** Disparate impact occurs where a neutral policy is applied to everyone, but primarily members of a protected class experience an adverse effect. Disparate impact does not require intentional discrimination. Disparate treatment requires a showing of intent to treat a student differently (at least in part because of their protected characteristics) and can occur either where a neutral policy is selectively enforced against students belonging to a protected class, or where the policy explicitly targets that protected group. Here, an education agency's generative AI and discipline policy might be over-enforced against EL students, due to the sheer disproportionality of false flags for non-native English speakers suggested by the Stanford study. Where an education agency is aware of these high error rates and consequent adverse effects for a protected group of students but nonetheless chooses to deploy the technology, it arguably meets requirements for a disparate impact or even a disparate treatment claim.

ii. **Hostile learning environment.** A hostile learning environment occurs where a student — or group of students — experiences severe, pervasive, or persistent treatment that interferes with the student's ability to participate in or benefit from services or activities provided by the school. For EL students, having their work frequently flagged for cheating by AI detectors and dealing with the accusation, investigation, and discipline that results, might create such an environment. Education agencies are tasked with the general obligation of ensuring a nondiscriminatory learning environment for all students. This obligation extends to responsibility for the conduct of third parties, such as vendors or contractors, with which the agency contracts, even if the conduct was not solely its own.

# Recommendations

Given the known inadequacies of AI detectors and the clear potential for disproportionate adverse effects on marginalized groups of students such as EL learners, education agencies should at minimum consider taking the following steps.

*Contemplate necessity of use*
Assess whether the use of this technology will be helpful in accomplishing the stated goal and should be used at all. As a starting point, the goal of deploying these technologies is to prevent academic dishonesty.

Educators are skilled professionals who are tasked with understanding their students' skills and challenges. More traditional mechanisms for cheating, such as purchasing essays online or having them written by a friend or family member, are often easy to identify for an educator familiar with that student's work and skill level. Given the known error rates of AI detectors, there is nothing to suggest that these technologies could or should be used to supplant a teacher's professional judgment in determining whether a piece of writing was actually the student's own work.

*Provide training regarding reliability*
Ensure educators understand: (i) the success and error rates of AI detectors, and the disproportionate error rate for non-native English speakers; (ii) that AI detectors should not supplant an educator's professional judgment; and (iii) that AI detector flags are not reliable as concrete proof of academic dishonesty. At most, if they use AI detectors at all, educators should recognize they can only be one piece of a broader inquiry for identifying potential academic dishonesty.

*Provide students an appeal process to challenge flags*
To the extent that schools use AI detectors, they must put in place significant procedural protections especially given the known error rates. Among the checks and balances that should be in place following a flag by an AI detector is the opportunity for implicated students to respond and advocate for themselves. Understand, however, that there are likely to be equity concerns with this process as well, as some students may not be as equipped as others (depending on grade level, English proficiency, etc.) to even understand the allegations or refute them.

## Conclusion

As schools grapple with rapidly emerging technologies, it is understandable that the response may include adopting innovative technologies of their own to combat undesired uses. However, it remains vital to stay vigilant of the potential pitfalls of these technologies and ensure that the protection of civil rights for all students in the classroom is a key priority.

**For more information from this research, read CDT's report, *Late Applications: Protecting Students' Civil Rights in the Digital Age*.**