NOVARA
CONSULTING GROUP

This document was translated from English by machine and may contain mistakes. The English original is at https://www.novaracg.com/2025/05/22/accused-by-perplexity-ai-detection-as-an-unvalidated-automated-decision-system-in-graduate-education/.

Perplexity Tarafından Suçlanmak: Lisansüstü Eğitimde Doğrulanmamış Otomatik Karar Sistemi Olarak Yapay Zeka Tespiti

Heather Grizzle · May 22, 2025

AI dedektörleri, yazının kim tarafından yazıldığını değil, ne kadar alışılmadık olduğunu ölçer. Bu durum en ağır şekilde, ikinci dilde yazanları ve onaylı destekleyici teknoloji kullanan öğrencileri etkiler.

İçindekiler

1. Tespit aracının ölçtüğü şey
2. Hata oranları ve bunları kimlerin taşıdığı
3. Bu durumun ulaştığı kesim ve bunun neden yanlış kesim olduğu
4. Kanıt standardı, karşılaştırılabilir herhangi bir şeyden daha düşüktür
5. Bir yüksek lisans öğrencisinin gerçekte yapması gereken şey
6. Bir kurumun belirlemesi gereken şey
7. Sonuç
8. Kaynakça

Herkesin verdiği tavsiye mekanizma konusunda yanlış

Yapay zeka konusunda yüksek lisans öğrencilerine verilen standart öğüt aşağı yukarı şöyledir: onu bir yerine koyma aracı değil bir araç olarak kullanın, kendi düşüncenizi kullanın ve öğretim üyelerinin keskin gözlemciler olduğunu, derinlikten yoksun çalışmaları er ya da geç fark edeceklerini bilin. İlk iki önerme sağlamdır. Üçüncüsü yanlıştır ve bu yanlışlık önemlidir, çünkü büyük ölçüde yerini başka bir şeye bırakmış olan bir insan muhakemesi sürecini tarif etmektedir.

Çoğu kurum yapay zeka kullanımını deneyimli okuyucuların ayırt etme yeteneğiyle tespit etmiyor. Bunu, ya da tespit ettiğine inanıyor, teslim platformlarına gömülü otomatik sınıflandırıcılar aracılığıyla yapıyor; bu sınıflandırıcılar bir belgeye olasılık puanı atayıp onu bir akademik dürüstlük sürecine aktarıyor. Dolayısıyla bir yüksek lisans öğrencisinin karşı karşıya olduğu soru, bir profesörün sığ düşünceyi fark edip etmeyeceği değildir. Soru, istatistiksel bir sınıflandırıcının yazısını işaretleyip işaretlemeyeceği ve işaretlerse başına ne geleceğidir.

Novara Consulting Group's position is that AI detection in higher education is an automated decision system producing consequential decisions about individuals, that it is deployed with less validation than institutions require of almost any other instrument affecting student standing, and that its error distribution falls hardest on multilingual and disabled writers. The governance problem is institutional. The consequences are individual.

Tespit aracının ölçtüğü şey

Bir yapay zeka yazı tespit aracı yazarlığı belirlemez. Her kelimenin önceki metne göre ne kadar öngörülebilir olduğunu ölçen şaşkınlık (perplexity) ve bir belge boyunca cümle yapısı ile uzunluğundaki değişimi ölçen patlayıcılık (burstiness) aracılığıyla, bir metnin istatistiksel olarak ne kadar olağandışı olduğunu tahmin eder. Yüksek derecede öngörülebilir ve yapısal olarak tekdüze olan metin, makine tarafından üretilme olasılığı daha yüksek olarak puanlanır.

That is the entire mechanism, and stating it plainly makes the problem visible. The detector is not measuring whether a person wrote something. It is measuring whether the writing is unusual, and treating unusualness as evidence of humanity. Writing that is careful, plain, structurally consistent, and lexically conservative is scored as machine-like, because that is what the model's proxy for machine-likeness is.

Bu durum, belirli ve öngörülebilir bir haksız suçlama nüfusu üretir ve bu rastgele bir nüfus değildir.

Hata oranları ve bunları kimlerin taşıdığı

Stanford'dan Liang ve meslektaşları tarafından *Patterns* dergisinde Temmuz 2023'te yayımlanan araştırma, test edilen yedi tespit aracı genelinde, anadili İngilizce olmayan kişilerin yazdığı TOEFL denemelerinin yaklaşık yüzde altmış birinin yapay

zeka tarafından üretilmiş olarak işaretlendiğini, buna karşılık anadili İngilizce olan kişilerin benzer yazılarında yanlış pozitif oranının neredeyse sıfır olduğunu ortaya koydu. Bunlar, İngilizce yeterliliğini kanıtlamak için standart bir sınava giren insanlar tarafından yazılmış denemelerdi. Yazarlar mekanizmayı doğrudan tespit etti: tespit araçları metin şaşkınlığına (perplexity) dayanıyordu ve ikinci bir dilde yazmak daha düşük şaşkınlık değeri üretiyordu.

Bağımsız testler, satıcı doğruluk iddialarını desteklemekte tekrar tekrar başarısız oldu. Weber-Wulff ve meslektaşları, Aralık 2023'te *International Journal for Educational Integrity* dergisinde yayımladıkları çalışmada, çeşitli belge türleri genelinde bir dizi tespit aracını değerlendirdi ve bunların ne doğru ne de güvenilir olduğu sonucuna vardı. Satıcılar ise genellikle ürünleri için bir yanlış pozitif oranı yayımlamayı reddediyor; bu da bir aracı benimseyen kurumun, disiplin işlemlerinde kullandığı bir aracın hata oranını belirleyemeyeceği anlamına geliyor.

Institutions have begun to act on this. Vanderbilt University disabled Turnitin's AI detection feature in August 2023, publishing the arithmetic that made the decision straightforward: a one percent false positive rate applied to seventy-five thousand submitted papers produces roughly seven hundred and fifty wrongful accusations a year. Other institutions have followed, and the direction of published university guidance since has been toward treating a detector score as a prompt for inquiry rather than as a finding.

Herhangi bir kurumsal değerlendirmede yer alması gereken bir piyasa gerçeği daha var. Tespit araçları pazarlayan bazı sağlayıcılar, kendi araçları da dahil olmak üzere tespiti atlatmak üzere tasarlanmış "insanlaştırıcı" ürünler de pazarlıyor. Başka herhangi bir güvence bağlamında, hem testi hem de onu geçmenin yolunu satan bir tedarikçi, ikisinden birini sağlamaktan men edilirdi.

Bu durumun ulaştığı kesim ve bunun neden yanlış kesim olduğu

Belgelenen eşitsizlik, anadili İngilizce olmayan yazarları ilgilendiriyor ve bunun açıklaması esrarengiz değil, mekaniktir. Şaşkınlık temelli tespit, daha basit kelime

Başka herhangi bir güvence bağlamında, hem testi hem de onu geçmenin yolunu satan bir tedarikçi, ikisinden

dağarcığını ve daha düzenli dilbilgisi yapısını cezalandırır. İkinci bir dilde çalışan bir yazar tam olarak bu profili üretir; yazı makine tarafından üretildiği için değil, ikinci bir dilde olağandışı ifade kurmak doğru ifade kurmaktan daha zor olduğu için.

> birini sağlamaktan men edilirdi.

Aynı mekanizma, neredeyse hiç dikkat çekmemiş bir kesime de ulaşır ve bu şirketin var olma nedeni de tam olarak bu kesimi ele almaktır.

A Deaf writer whose first language is American Sign Language is writing English as a second language, and the syntactic conservatism that follows is precisely what the detector scores as machine-like. A student with dyslexia using text prediction, grammar assistance, or structured writing support, all of which are ordinary and often formally approved accommodations, produces text that has been regularized by the assistive tool. A student with a motor impairment composing by speech-to-text produces text shaped by dictation software's own language model. A student using translation assistance produces output carrying the statistical signature of the translation system.

Her durumda, uyarlama işaretlemeye yol açar. Kurumunun onayladığı desteği kullanan öğrenci, buna ihtiyaç duymayan öğrenciye kıyasla suçlanma olasılığı daha yüksektir ve suçlama, neyi ölçtüğüne dair hiçbir açıklama içermeyen bir sayıyla birlikte gelir. Ardından öğrencinin kendi çalışmasını kendisinin yazdığını kanıtlaması istenir ki bu, bunu karşılamaya en az yetkin olan kişiden bir olumsuzluğu kanıtlamasını talep etmek demektir.

An institution that has approved an assistive technology accommodation and simultaneously runs a detector penalizing its output has built a contradiction into its own disciplinary process. It is unlikely to have noticed, because the disability services office and the academic integrity office do not typically review each other's instruments. Postsecondary institutions carry obligations under the Americans with Disabilities Act and Section 504 of the Rehabilitation Act to provide academic adjustments and to avoid practices that screen out students with disabilities. A detection practice that systematically flags the products of approved accommodations sits uncomfortably against both.

Kanıt standardı, karşılaştırılabilir herhangi bir şeyden daha düşüktür

Consider what an institution would require before acting on any other automated instrument affecting a student′s standing.

Bir kabul sınavı, yayımlanmış geçerlilik kanıtları, kanıtlanmış güvenilirlik, demografik gruplar genelinde belgelenmiş performans ve savunulabilir bir kesme puanı gerektirir. Bir dürüstlük ihbarını tetikleyen bir gözetim aracı, en azından kaydın incelenebilir olmasını gerektirir. Bir engellilik belirlemesi, belgeleme ve adı belirtilen bir karar vericiyi gerektirir. Satıcısının hata oranını açıklamayacağı bir araç tarafından üretilen, açıklanmamış sayısal bir puana dayanarak sonuç doğuran bir karar veren bir kurum, normalde keyfi hareket etmiş olarak anlaşılırdı.

Yapay zeka tespit puanları farklı ele alınmaktadır ve bunun nedeni ilke değil, tür gibi görünmektedir. Bu puanlar, kurumların zaten güvendiği bir intihal aracının içinde gelirler, yanında bulundukları benzerlik endeksine benzerler ve akademik yönetişim yoluyla değil, satın alma süreci yoluyla benimsenmişlerdir. Benzerlik endeksi en azından bir öğrencinin inceleyip itiraz edebileceği bir kaynak belgeye işaret eder. Yapay zeka puanı hiçbir şeye işaret etmez. Kaynak yoktur, pasaj yoktur, karşılaştırma yoktur ve öğrencinin iddianın temelini inceleyebileceği hiçbir mekanizma yoktur.

Better institutional practice treats a detector output as triage rather than evidence: a prompt to open a conversation in which the student explains her argument, walks through her drafting, and discusses her sources. That is the correct posture, and it is worth noting what it concedes. If the detector′s output is only a reason to look more closely, and the looking is what actually determines the outcome, then the detector is a mechanism for deciding which students get investigated. Given the error distribution above, it is a mechanism for deciding that multilingual and disabled students get investigated more often.

Bir yüksek lisans öğrencisinin gerçekte yapması gereken şey

Yararlı tavsiye, insan gibi görünmek için daha çok çalışmak değildir; bu hem onur kırıcı hem de ulaşılamaz bir hedeftir. Yararlı tavsiye, açıklanmamış bir otomatik karar ile karşı karşıya kalan herhangi bir kurumun kayıt tutması gereken aynı nedenle,

kaydı tutmaktır: suçlama süreç kanıtı olmadan çürütülemez ve süreç kanıtının ihtiyaç duyulmadan önce oluşturulmuş olması gerekir.

Draft in an environment that retains version history, and do not disable it. A document with a two-month revision trail is dispositive in a way no argument about writing style will ever be. Keep research notes, annotated sources, and outlines, and keep them dated. Where an assistive or language support tool is used, disclose it in advance under the course's stated policy rather than after a flag, because prior disclosure converts a defense into a record. Where an accommodation involves writing support, ensure the disability services office has documented it, and understand that this documentation is now also integrity documentation. And be able to explain the argument aloud, not because a professor's perception is the safeguard, but because the structured conversation is where these matters are actually resolved.

Bunların hiçbiri aldatmakla ilgili değildir. Bu, incelenemeyen bir sistemin kendisi hakkında kendisinin çürütmesi istenecek bir iddiada bulunabileceği herhangi bir kişinin yapması gereken şeydir.

Bir kurumun belirlemesi gereken şey

Bir tespit aracı çıktısı bir öğrenciye ulaşmadan önce, bir akademik dürüstlük ofisinin yanıtlayabilmesi gereken dört soru.

1. What is the instrument's false positive rate, on what population, established by whom. If the vendor will not state it, that is the finding, and it should be recorded as one rather than treated as an absence of information.
2. Çok dilli öğrenciler, uluslararası öğrenciler ve onaylı yardımcı teknoloji kullanan öğrenciler dahil olmak üzere, kurumun hizmet verdiği nüfuslar genelinde ayrıştırılmış performans nedir. Bu nüfuslara hizmet veren ve yanıt veremeyen bir kurum, aracı kendi bağlamı için değerlendirmemiş demektir.
3. Çıktıya yazılı olarak hangi kanıt ağırlığı atanmaktadır ve herhangi bir olumsuz bulgu buna mı dayanmaktadır. Yalnızca bir puan bir süreci başlatabiliyorsa, kurum doğrulanmamış bir aracı disiplin tetikleyicisi olarak benimsemiş demektir.
4. Onaylanmış uyarlamalar ile tespit çıktıları arasındaki etkileşimi kim gözden geçiriyor. Eğer kimse geçirmiyorsa, kurum bu etkileşimin olumsuz olduğunu varsaymalıdır, çünkü mekanizma bunu garanti eder.

Sonuç

Yüksek lisans öğrencilerinin kendi başlarına düşünmesi, kendi fikirlerini savunması ve kestirme yollar yerine uzmanlık geliştirmesi gerektiği tavsiyesi doğrudur ve her çağda doğru olurdu. Ancak bu, öğrencilerin gerçekte karşı karşıya olduğu riske yanıt vermemektedir; bu risk, sığ bulunacakları değil, istatistiksel olağandışılığı ölçen bir sınıflandırıcı tarafından işaretlenip bir olumsuzluğu kanıtlamalarının istenmesidir.

Bu risk eşit olmayan biçimde dağılmıştır ve dağılımı, dürüstlük politikasının üretmesi gereken dağılımın tam tersidir. Haksız yere suçlanma olasılığı en yüksek olan öğrenciler, ikinci bir dilde yazanlar, ilk dilde işaret edenler ve kurumlarının resmi olarak onayladığı uyarlamaları kullananlardır. Bunlar daha sık soruşturulmaktadır, çünkü kopya çekme olasılıkları daha yüksek olduğu için değil, makine yazısı için bir vekil ölçütün aslında kısıtlama altında yazmanın bir vekil ölçütü olduğu ortaya çıktığı için.

Kurumlar bu sonucu seçmediler. Bunu, doğrulama kanıtı olmadan alınan bir satın alma kararıyla edindiler ve devreye sokulmadan önce sorulması gereken soruları birisi sorana kadar bunu üretmeye devam edecekler.

Kaynakça

Liang, Weixin, Mert Yuksekgonul, Yining Mao, Eric Wu, and James Zou. "GPT Detectors Are Biased Against Non-Native English Writers." *Patterns* 4, no. 7 (July 2023).

Weber-Wulff, Debora, et al. "Testing of Detection Tools for AI-Generated Text." *International Journal for Educational Integrity* 19 (December 2023).

Vanderbilt University Center for Teaching, statement on disabling Turnitin's AI detection feature, August 2023.

1990 tarihli Engelli Amerikalılar Yasası ve 1973 tarihli Rehabilitasyon Yasası'nın 504. Bölümü'nün, yükseköğretim kurumlarına ve akademik düzenlemelere uygulanışı.

TİPS

Know something about a sign language AI system that the public should know? Send it to tips@novaracg.com. Every tip is read by me and kept confidential. If you work in the UK, writing to us is not by itself a protected disclosure under the Public Interest Disclosure Act 1998, so before you share employer information, check your options with Protect (protect-advice.org.uk), the free whistleblowing advice line.

Nasıl alıntılanır

Grizzle, H. M. (2025, May 22). Perplexity Tarafından Suçlanmak: Lisansüstü Eğitimde Doğrulanmamış Otomatik Karar Sistemi Olarak Yapay Zeka Tespiti. Novara Consulting Group. https://www.novaracg.com/tr/2025/05/22/accused-by-perplexity-ai-detection-as-an-unvalidated-automated-decision-system-in-graduate-education/