



Doi: <https://doi.org/10.70577/asce.v5i2.861>

Recibido: 2026-04-28

Aceptado: 2026-05-12

Publicado: 2026-05-26

Falsos positivos en detectores de Inteligencia Artificial en trabajos académicos: análisis comparativo de plataformas digitales en contextos educativos

False Positives in Artificial Intelligence Detectors in Academic Assignments: A Comparative Analysis of Digital Platforms in Educational Contexts

Autor(s)

Paúl Alexander Yanangómez Suárez¹
p.alexander0181@outlook.com
<https://orcid.org/0009-0003-3919-3022>
Unidad Educativa Orianga
Loja – Ecuador

Enma Regina Díaz Díaz²
enmadiadz06@gmail.com
<https://orcid.org/0009-0005-1242-5254>
Unidad Educativa Orianga
Loja – Ecuador

Lilia Margarita Ojeda Cabrera³
liliamarg89j@gmail.com
<https://orcid.org/0009-0008-9356-605X>
Unidad Educativa Orianga
Loja – Ecuador

Carmen Beneranda Yaguachi Quichimbo⁴
carmenyaguachi@yahoo.es
<https://orcid.org/0009-0009-8360-3408>
Unidad Educativa Orianga
Loja - Ecuador

María Cristina Carrillo Díaz⁵
maricris114@yahoo.es
<https://orcid.org/0009-0006-8860-5910>
Unidad Educativa Libertador Bolívar
Azuay – Ecuador

Como Citar

Yanangómez Suárez. P. A. & Díaz Díaz. E. R. & Ojeda Cabrera. L. M. & Yaguachi Quichimbo. C. B. & Carrillo Díaz. M. C. (2026) Falsos positivos en detectores de Inteligencia Artificial en trabajos académicos: análisis comparativo de plataformas digitales en contextos educativos ASCE MAGAZINE 5(2) 1976-



Resumen

La presente investigación tuvo como objetivo analizar la confiabilidad de diferentes detectores de Inteligencia Artificial aplicados a trabajos académicos en contextos educativos. El estudio se desarrolló mediante un enfoque cuantitativo, descriptivo y comparativo, utilizando como muestra a cinco docentes que evaluaron un mismo contenido textual correspondiente a un poema antiguo de Mario Benedetti a través de las plataformas JustDone, Reprism, Grammarly, Sidekicker y ZeroGPT. Los resultados evidenciaron diferencias significativas entre las herramientas analizadas. JustDone, Reprism y Sidekicker clasificaron el contenido humano con porcentajes elevados de detección de IA, alcanzando valores de hasta 99%, mientras que ZeroGPT y Grammarly identificaron el texto como completamente humano, registrando 0% de contenido generado por Inteligencia Artificial. Asimismo, se aplicó una encuesta a los docentes participantes, cuyos resultados reflejaron desconfianza frente a la precisión de estas herramientas y preocupación por el posible impacto negativo en estudiantes de Educación Básica y Bachillerato. La investigación concluye que algunos detectores de IA presentan falsos positivos y limitaciones importantes en su precisión, por lo que no deben utilizarse como prueba definitiva dentro de procesos de evaluación académica. Se recomienda complementar estas herramientas con revisión humana, análisis contextual y criterios pedagógicos antes de emitir juicios sobre la autenticidad de un trabajo estudiantil.

Palabras clave: Inteligencia Artificial, falsos positivos, detectores de IA, evaluación académica, educación, trabajos académicos, ética educativa, estudiantes.



Abstract

This research aimed to analyze the reliability of different Artificial Intelligence detection tools applied to academic assignments in educational contexts. The study was conducted using a quantitative, descriptive, and comparative approach, involving five teachers who evaluated the same textual content corresponding to an old poem by Mario Benedetti through the platforms JustDone, Reprism, Grammarly, Sidekicker, and ZeroGPT.

The results revealed significant differences among the analyzed tools. JustDone, Reprism, and Sidekicker classified the human-written content with high AI detection percentages, reaching values of up to 99%, whereas ZeroGPT and Grammarly identified the text as entirely human, reporting 0% AI-generated content. Additionally, a survey was conducted with participating teachers, whose responses reflected distrust regarding the accuracy of these tools and concern about their potential negative impact on Basic Education and High School students. The study concludes that some AI detectors present false positives and important limitations in accuracy; therefore, they should not be used as definitive evidence in academic evaluation processes. It is recommended that these tools be complemented with human review, contextual analysis, and pedagogical criteria before making judgments about the authenticity of student academic work.

Keywords: Artificial Intelligence, false positives, AI detectors, academic evaluation, education, academic assignments, educational ethics, students



Introducción

La incorporación de la inteligencia artificial generativa en la educación ha transformado las formas de producción, revisión y evaluación de textos académicos. Frente a este escenario, diversas instituciones han recurrido a aplicaciones detectoras de IA con el propósito de identificar posibles trabajos elaborados mediante herramientas automatizadas. Sin embargo, la literatura reciente advierte que estos sistemas no ofrecen certeza absoluta, sino resultados probabilísticos que pueden variar según el tipo de texto, la plataforma utilizada y las condiciones de análisis.

Uno de los principales problemas asociados a estas herramientas es la aparición de falsos positivos, es decir, casos en los que un texto escrito por una persona es clasificado erróneamente como generado por inteligencia artificial. OpenAI retiró en 2023 su propio clasificador de textos por su “baja tasa de precisión”, reconociendo además que podía identificar textos humanos como generados por IA. De manera similar, estudios recientes han señalado que los detectores pueden producir errores relevantes y que los falsos positivos representan un riesgo serio en contextos académicos, especialmente cuando se emplean como prueba única para acusar a estudiantes de deshonestidad académica.

En el ámbito educativo, esta problemática adquiere mayor gravedad porque una clasificación errónea puede afectar la evaluación, la reputación y el bienestar emocional de los estudiantes. Investigaciones recientes sostienen que los detectores de IA pueden generar resultados inconsistentes, carecen de suficiente transparencia y deben utilizarse con cautela dentro de los procesos de evaluación académica. Incluso plataformas ampliamente utilizadas, como Turnitin, han establecido umbrales para reducir la incidencia de falsos positivos y recomiendan no interpretar los resultados como evidencia definitiva.

En este contexto, la presente investigación analiza el comportamiento de cinco aplicaciones detectoras de inteligencia artificial: JustDone, Reprism, Grammarly, Sidekicker y ZeroGPT. El estudio se desarrolló con la participación de cinco docentes, quienes evaluaron un mismo contenido textual desde distintos dispositivos y navegadores, con el fin de identificar variaciones en los porcentajes de detección y posibles inconsistencias entre plataformas.



Los resultados preliminares permitieron comprobar que tres de las herramientas analizadas generaron falsos positivos al clasificar como contenido producido por IA un poema de origen humano. En contraste, otras aplicaciones mostraron resultados más coherentes con la naturaleza del texto evaluado. Estos hallazgos evidencian la necesidad de cuestionar el uso automático de los detectores de IA en tareas escolares y trabajos académicos, especialmente en Educación Básica y Bachillerato, donde una acusación infundada puede provocar consecuencias pedagógicas, emocionales y éticas significativas.

Por tanto, este estudio busca aportar evidencia empírica sobre la confiabilidad de los detectores de IA en contextos educativos y promover una reflexión crítica sobre su uso responsable. La investigación sostiene que estas herramientas pueden ser utilizadas como apoyo orientativo, pero no deben constituirse en mecanismos definitivos de sanción o rechazo académico sin revisión humana, diálogo pedagógico y análisis contextual del proceso de escritura del estudiante.

Materiales y métodos

Enfoque de investigación

La presente investigación se desarrolló bajo un enfoque cuantitativo, de tipo descriptivo y comparativo, orientado al análisis del comportamiento de diferentes aplicaciones detectoras de Inteligencia Artificial frente a un mismo contenido textual.

Diseño de investigación

El estudio se enmarcó en un diseño no experimental de corte transversal, debido a que los datos fueron recolectados en un único momento temporal sin manipulación de variables independientes. Asimismo, se aplicó una estrategia comparativa entre plataformas digitales detectoras de IA.



Participantes

La muestra estuvo conformada por cinco docentes pertenecientes al ámbito educativo, quienes participaron voluntariamente en el experimento. Las pruebas fueron realizadas desde computadores y dispositivos móviles con el propósito de identificar posibles inconsistencias en los resultados emitidos por las aplicaciones analizadas.

Aplicaciones evaluadas

Las plataformas detectoras de Inteligencia Artificial utilizadas en el estudio fueron:

- JustDone
- Reprism
- Grammarly
- Sidekicker
- ZeroGPT

Instrumento y procedimiento

Como instrumento de análisis se utilizó un mismo contenido textual correspondiente a un poema inédito atribuido a Mario Benedetti, seleccionado con fines comparativos para evaluar la consistencia de los detectores de IA.

El procedimiento se desarrolló en las siguientes fases:

Cada docente ingresó el mismo contenido textual en las cinco plataformas detectoras de IA.

Las pruebas fueron ejecutadas desde computadores y dispositivos móviles.

Se registraron mediante capturas de pantalla los porcentajes de detección emitidos por cada aplicación.

Posteriormente, los participantes respondieron una encuesta final orientada a identificar:

- percepción de confiabilidad,
- experiencia durante el experimento,

-
- nivel de sorpresa frente a los resultados,
 - y opinión sobre el uso de detectores de IA en contextos educativos.

Finalmente, los resultados fueron organizados y comparados para identificar:

- coincidencias,
- variaciones,
- inconsistencias,
- y presencia de falsos positivos.

Técnica de recolección de datos

La recolección de información se realizó mediante:

- observación directa de resultados,
- registro fotográfico mediante capturas de pantalla,
- y aplicación de una encuesta estructurada a los docentes participantes.

Técnica de análisis de datos

Los datos obtenidos fueron analizados mediante estadística descriptiva simple, utilizando tablas comparativas y análisis porcentual para identificar diferencias entre plataformas detectoras de IA.

Asimismo, se realizó un análisis interpretativo de las respuestas obtenidas en la encuesta final para comprender la percepción de los participantes respecto a la confiabilidad de estas herramientas en el ámbito educativo.

Consideraciones éticas

La participación de los docentes fue voluntaria y confidencial. La información recopilada se utilizó exclusivamente con fines académicos y científicos, garantizando el respeto a la privacidad de los participantes y el uso responsable de los datos obtenidos.

Resultados

Primera fase: presentación de resultados

Con el propósito de analizar la confiabilidad de diferentes aplicaciones detectoras de Inteligencia Artificial, se evaluó un mismo contenido textual correspondiente a un poema antiguo de Mario Benedetti mediante cinco plataformas digitales. Las pruebas fueron realizadas por seis docentes desde distintos computadores y dispositivos móviles.

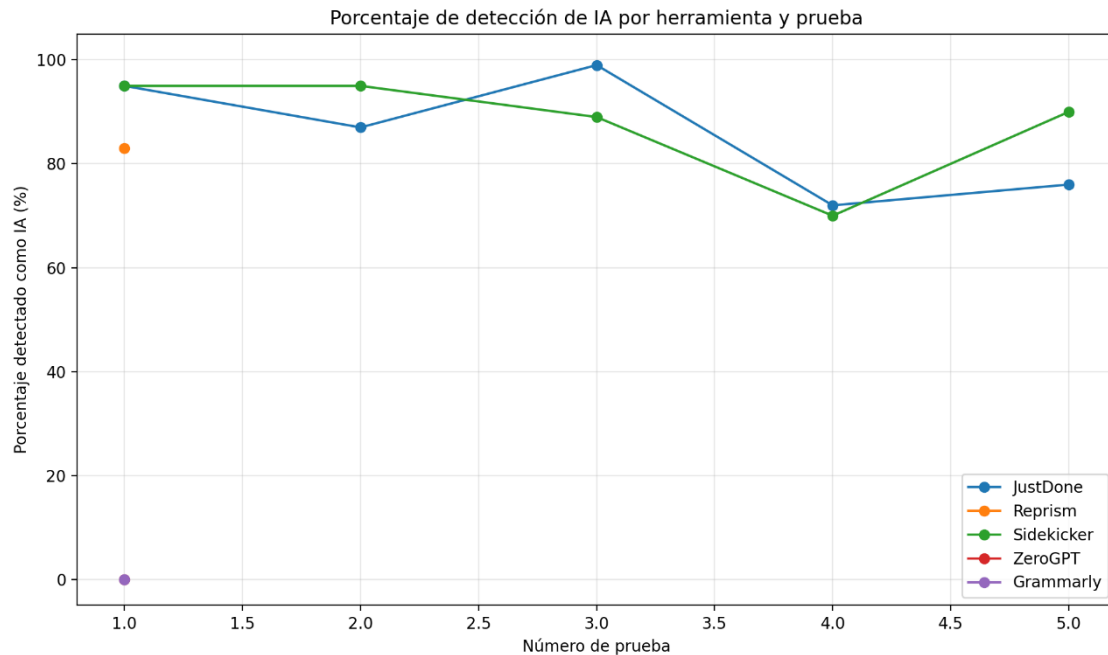
Los resultados obtenidos evidenciaron diferencias significativas entre las herramientas analizadas, particularmente en los porcentajes de detección de IA emitidos para un mismo texto de origen humano.

Tabla 1. Resultados obtenidos en detectores de IA

Herramienta	Resultados obtenidos	Promedio de IA
JustDone	95%, 87%, 99%, 72%, 76%	85,8%
Reprism	83%	83%
Sidekicker	95%, 95%, 89%, 70%, 90%	87,8%
ZeroGPT	0%	0%
Grammarly	0%	0%

Figura 1. Resultados individuales por herramienta

Análisis de la Figura 1



La Figura 1 evidencia diferencias importantes entre las plataformas detectoras de Inteligencia Artificial evaluadas. JustDone y Sidekicker registraron porcentajes elevados de detección de IA, alcanzando valores superiores al 90% en varias pruebas, pese a tratarse de un poema humano previamente publicado.

Por su parte, Reprism clasificó el contenido con un 83% de generación artificial, manteniendo igualmente una tendencia alta de detección.

En contraste, ZeroGPT y Grammarly identificaron el texto como completamente humano, registrando un 0% de contenido generado por IA.

Estos resultados permiten observar inconsistencias significativas entre herramientas que analizan exactamente el mismo contenido textual.

Interpretación

Los hallazgos obtenidos en esta primera fase sugieren la presencia de falsos positivos en algunas aplicaciones detectoras de Inteligencia Artificial. Mientras ciertas plataformas clasifican el texto como altamente generado por IA, otras reconocen su origen humano sin detectar contenido artificial.

Esta falta de uniformidad evidencia limitaciones en la precisión y confiabilidad de los detectores de IA utilizados en contextos académicos. Asimismo, los resultados muestran que dichas herramientas no deberían constituirse como mecanismos definitivos de evaluación o sanción académica, especialmente en estudiantes de Educación Básica y Bachillerato.

Figura 2. Promedio de detección de IA por herramienta

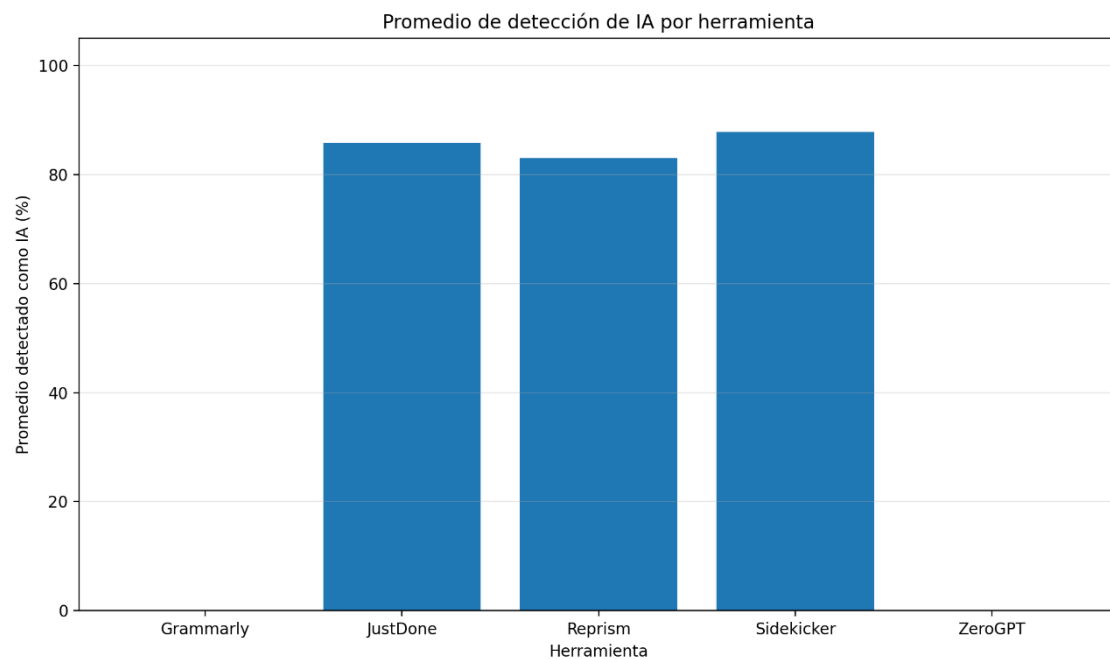


Figura 2. Promedio de detección de IA por herramienta

Fuente: Datos obtenidos de las pruebas realizadas con detectores de Inteligencia Artificial sobre un mismo poema de Mario Benedetti.

Análisis de la Figura 2

La Figura 2 presenta el promedio de detección de IA registrado por cada una de las plataformas evaluadas. Los resultados muestran que **Sidekicker** obtuvo el promedio más alto con **87,8%**, seguido de **JustDone** con **85,8%** y **Reprism** con **83%**.

En contraste, **ZeroGPT** y **Grammarly** identificaron el contenido como completamente humano, registrando un promedio de **0%** de detección de Inteligencia Artificial.

Los resultados evidencian diferencias considerables entre herramientas que analizaron exactamente el mismo contenido textual, lo cual pone en evidencia inconsistencias en los sistemas de detección automática.

Interpretación

La variabilidad observada entre plataformas permite inferir que ciertos detectores de IA presentan una alta probabilidad de generar falsos positivos, especialmente cuando se enfrentan a textos literarios humanos con estructuras lingüísticas complejas o estilísticamente elaboradas.

Asimismo, los resultados sugieren que los detectores de IA aún presentan limitaciones técnicas importantes para diferenciar de manera precisa entre escritura humana y contenido generado artificialmente. En consecuencia, estas herramientas no deberían utilizarse como criterio único para validar o rechazar trabajos académicos dentro del contexto educativo.

Figura 3. Variabilidad de detección de IA por herramienta

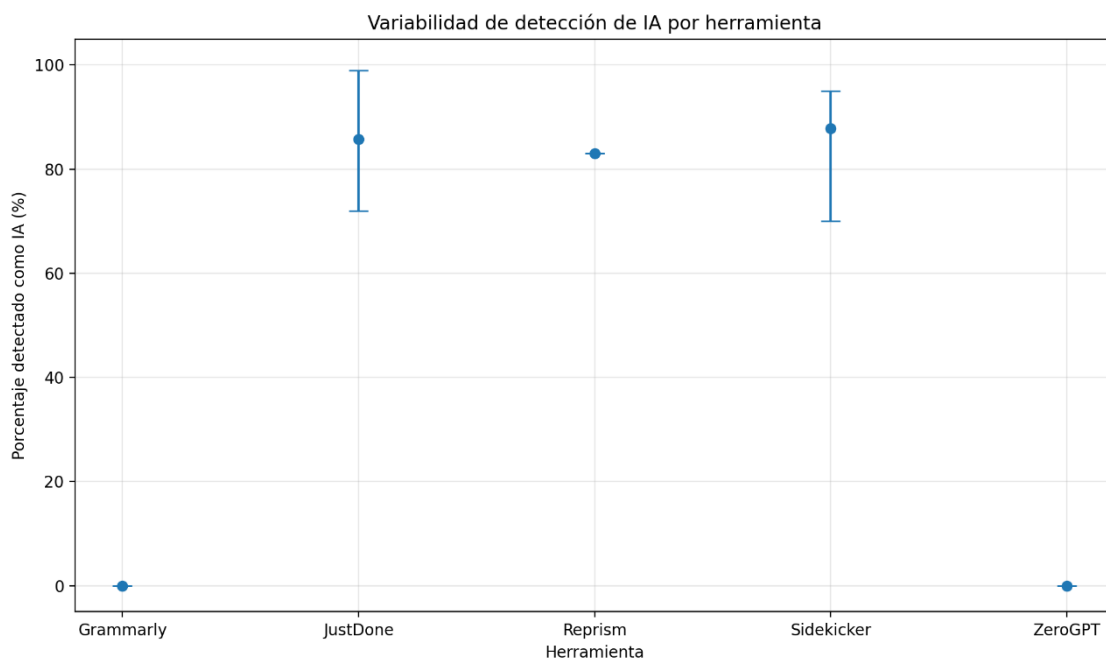


Figura 3. Variabilidad de detección de IA por herramienta

Fuente: Datos obtenidos de las pruebas realizadas con detectores de Inteligencia Artificial sobre un mismo poema de Mario Benedetti.

Análisis de la Figura 3

La Figura 3 muestra la variabilidad existente en los porcentajes de detección de IA emitidos por cada herramienta. Se observa que **JustDone** presentó resultados entre **72%** y **99%**, mientras que **Sidekicker** registró valores entre **70%** y **95%**.

Por otra parte, **Reprism** mantuvo un único resultado de **83%**, mientras que **ZeroGPT** y **Grammarly** se mantuvieron en **0%**.

Interpretación

La variabilidad observada en JustDone y Sidekicker evidencia falta de estabilidad en los resultados emitidos por estas plataformas. Esto resulta relevante porque el mismo contenido textual fue evaluado por diferentes docentes desde computadores y dispositivos móviles.

Estos hallazgos refuerzan la presencia de **falsos positivos** en determinadas herramientas de detección de IA y demuestran que los resultados no deben ser asumidos como evidencia definitiva dentro de procesos de evaluación académica.

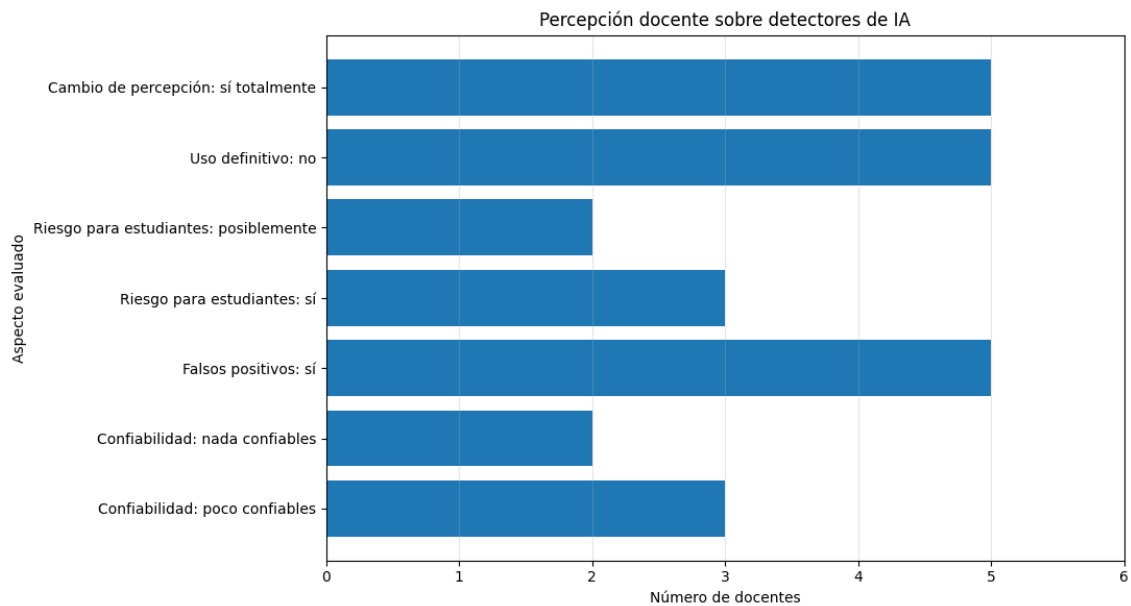
Segunda fase: resultados de la encuesta aplicada a docentes

Con el propósito de complementar los resultados técnicos obtenidos mediante las aplicaciones detectoras de IA, se aplicó una encuesta final a los docentes participantes. Esta permitió conocer su percepción sobre la confiabilidad de las herramientas, la presencia de falsos positivos, el posible riesgo para los estudiantes y el uso de estos detectores como prueba definitiva en procesos académicos.

Tabla 2. Percepción docente sobre los detectores de IA

Aspecto evaluado	Resultados obtenidos
Confiabilidad de las herramientas	Poco confiables: 3 / Nada confiables: 2
Presencia de falsos positivos	Sí: 5
Riesgo para estudiantes	Sí: 3 / Posiblemente: 2
Uso como prueba definitiva	No: 5
Cambio de percepción tras el experimento	Sí, totalmente: 5

Figura 4. Percepción docente sobre detectores de IA



Fuente: Encuesta aplicada a docentes participantes.

Análisis de la Figura 4

Los resultados evidencian una percepción crítica por parte de los docentes frente al uso de detectores de Inteligencia Artificial en contextos educativos. En relación con la

confiabilidad de las herramientas, **3 docentes las consideraron poco confiables** y **2 docentes las calificaron como nada confiables**, lo que demuestra una valoración negativa generalizada.

Asimismo, **los 5 docentes encuestados afirmaron que las aplicaciones pueden generar falsos positivos**, confirmando que los resultados obtenidos en la primera fase provocaron dudas sobre la precisión de estas plataformas.

Respecto al posible impacto en los estudiantes, **3 docentes señalaron que estas herramientas sí pueden afectar injustamente a los estudiantes**, mientras que **2 indicaron que posiblemente podrían hacerlo**. Esto evidencia una preocupación ética y pedagógica frente al uso automático de estos sistemas en la evaluación académica.

Además, **los 5 docentes manifestaron que los detectores de IA no deben utilizarse como prueba definitiva** para rechazar o sancionar trabajos académicos. Finalmente, **los 5 participantes indicaron que su percepción cambió totalmente después del experimento**, lo que demuestra que la experiencia práctica permitió reconocer limitaciones importantes en estas aplicaciones.

Interpretación

La segunda fase confirma que el problema de los detectores de IA no es únicamente técnico, sino también educativo, ético y evaluativo. Los docentes participantes reconocieron que estas herramientas pueden generar falsos positivos y, por tanto, no deben emplearse como único criterio para juzgar la autoría de trabajos académicos.

Estos resultados refuerzan la necesidad de que las instituciones educativas promuevan una evaluación más justa, basada en la revisión humana, el diálogo con el estudiante, la sustentación oral del trabajo y el análisis del proceso de elaboración académica.

En consecuencia, los detectores de IA pueden ser utilizados como herramientas de apoyo, pero no como mecanismos definitivos de sanción o rechazo, especialmente en estudiantes de Educación Básica y Bachillerato.

Discusión

Los resultados obtenidos en la presente investigación evidencian inconsistencias importantes entre las plataformas detectoras de Inteligencia Artificial analizadas. A pesar de utilizar un mismo contenido textual correspondiente a un poema humano antiguo de Mario Benedetti, algunas herramientas registraron porcentajes extremadamente altos de detección de IA, mientras otras identificaron el texto como completamente humano.

Las plataformas **JustDone**, **Reprism** y **Sidekicker** presentaron porcentajes superiores al 70%, llegando incluso al 99% de detección de IA. Estos resultados permiten inferir la presencia de falsos positivos, debido a que el contenido analizado corresponde a un texto humano previo al desarrollo de los actuales modelos de Inteligencia Artificial generativa. En contraste, **ZeroGPT** y **Grammarly** identificaron correctamente el contenido como humano, registrando 0% de detección artificial.

Los hallazgos coinciden con lo planteado por OpenAI (2023), quienes reconocieron limitaciones en los detectores automáticos de IA y retiraron su clasificador experimental debido a su baja precisión y posibilidad de clasificar erróneamente textos humanos como generados artificialmente. De igual manera, Dalalah y Dalalah (2023) sostienen que los detectores de IA presentan problemas de confiabilidad y pueden producir falsos positivos dentro de contextos académicos.

Asimismo, los resultados obtenidos reflejan que los detectores de IA operan mediante cálculos probabilísticos y patrones lingüísticos, mas no mediante mecanismos capaces de determinar con certeza absoluta el origen humano o artificial de un texto. Esta situación se vuelve especialmente preocupante en el ámbito educativo, donde una clasificación errónea puede derivar en sanciones injustas, afectaciones emocionales y vulneración de derechos académicos de los estudiantes.

En la segunda fase del estudio, la encuesta aplicada a docentes confirmó esta preocupación. La mayoría de los participantes consideró que las herramientas analizadas son poco confiables y coincidieron en que los detectores de IA no deberían utilizarse como prueba definitiva para rechazar tareas académicas. Además, los docentes señalaron que estas aplicaciones podrían afectar injustamente a estudiantes de Educación Básica y Bachillerato.



Estos resultados también guardan relación con lo señalado por Turnitin (2024), quienes recomiendan interpretar los resultados de detección de IA únicamente como indicadores orientativos y no como evidencia concluyente. Del mismo modo, investigaciones recientes advierten que la utilización indiscriminada de estas plataformas puede generar conflictos éticos y pedagógicos dentro de las instituciones educativas.

En consecuencia, la presente investigación demuestra que existen diferencias significativas entre detectores de IA al analizar un mismo contenido humano, evidenciando limitaciones técnicas y alta probabilidad de falsos positivos en determinadas plataformas. Por ello, se considera necesario que las instituciones educativas promuevan procesos de evaluación más integrales, incorporando revisión humana, análisis contextual y sustentación académica antes de emitir juicios sobre la autenticidad de un trabajo estudiantil.

El estudio aporta evidencia empírica relevante para el debate actual sobre Inteligencia Artificial y educación, reafirmando la necesidad de utilizar estas tecnologías de manera ética, responsable y complementaria, evitando que algoritmos automatizados sustituyan el criterio pedagógico y humano dentro de los procesos educativos.

Conclusiones

La presente investigación permitió evidenciar que algunos detectores de Inteligencia Artificial presentan inconsistencias significativas al analizar un mismo contenido textual humano. Los resultados obtenidos demostraron que plataformas como **JustDone**, **Reprism** y **Sidekicker** clasificaron un poema antiguo de Mario Benedetti como contenido generado por IA, alcanzando porcentajes elevados de detección artificial, pese a tratarse de un texto humano previamente publicado.

En contraste, herramientas como **ZeroGPT** y **Grammarly** identificaron el contenido como completamente humano, registrando 0% de detección de IA. Esta diferencia entre plataformas confirma la existencia de falsos positivos y evidencia limitaciones importantes en la precisión de ciertos detectores automáticos.



Asimismo, los resultados muestran que los detectores de IA no operan mediante mecanismos de verificación absoluta, sino a partir de cálculos probabilísticos y patrones lingüísticos, lo que incrementa el riesgo de errores en textos académicos, literarios o estructuralmente complejos.

La encuesta aplicada a los docentes participantes reforzó esta problemática, ya que la mayoría manifestó desconfianza frente a las herramientas analizadas y coincidió en que los detectores de IA no deben utilizarse como prueba definitiva para sancionar o rechazar trabajos académicos. Los participantes también señalaron que estas aplicaciones podrían afectar injustamente a estudiantes de Educación Básica y Bachillerato.

En consecuencia, se concluye que los detectores de Inteligencia Artificial deben emplearse únicamente como herramientas complementarias de apoyo y no como mecanismos automáticos de validación académica. La evaluación educativa requiere necesariamente del criterio humano, la revisión pedagógica y el análisis contextual del proceso de aprendizaje del estudiante.

Finalmente, esta investigación aporta evidencia científica relevante sobre la necesidad de desarrollar políticas educativas responsables frente al uso de Inteligencia Artificial en la educación, promoviendo prácticas éticas, justas y basadas en evidencia antes de emitir juicios sobre la autenticidad de los trabajos académicos.

Referencias Bibliográficas

OpenAI. (2023). *New AI classifier for indicating AI-written text*.
<https://openai.com/index/new-ai-classifier-for-indicating-ai-written-text/>

Dalalah, D., & Dalalah, O. M. A. (2023). *The false positives and false negatives of generative AI detection tools in education and academic research: The case of ChatGPT*. *The International Journal of Management Education*, 21(2), 100822.
<https://doi.org/10.1016/j.ijme.2023.100822>



Turnitin. (2024). *Using the AI Writing Report*.
<https://guides.turnitin.com/hc/en-us/articles/22774058814093-Using-the-AI-Writing-Report>

Deep, P. D. (2025). *Evaluating the effectiveness and ethical implications of AI detection tools in academic settings*. *Information*, 16(10), 905.
<https://www.mdpi.com/2078-2489/16/10/905>

Northern Illinois University. (2024). *AI detectors: An ethical minefield*.
<https://citl.news.niu.edu/2024/12/12/ai-detectors-an-ethical-minefield/>

Conflicto de intereses:

Los autores declaran que no existe conflicto de interés posible.

Financiamiento:

No existió asistencia financiera de partes externas al presente artículo.

Agradecimiento:

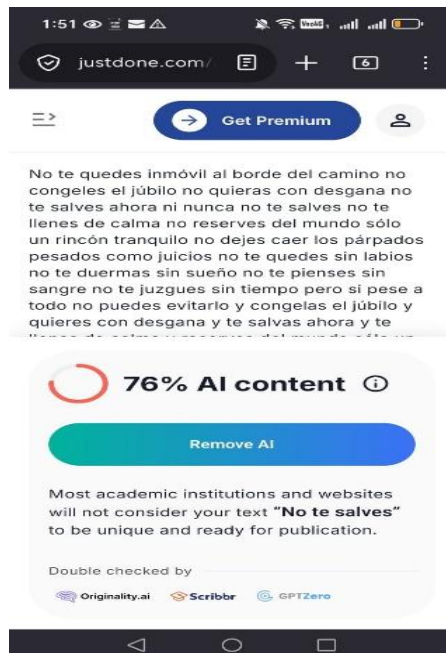
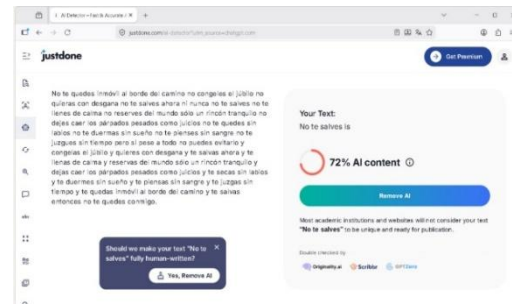
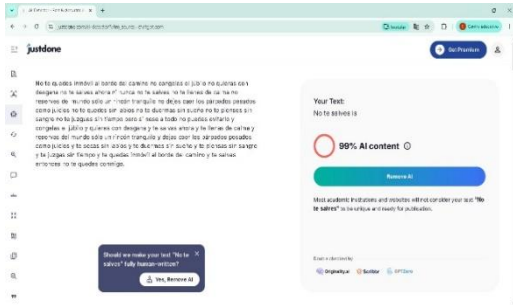
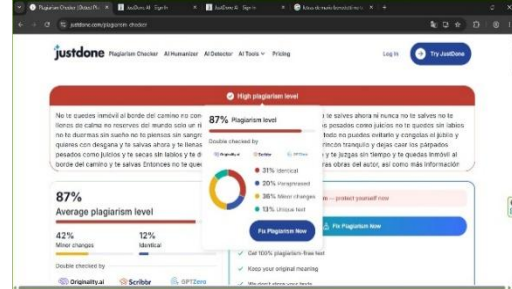
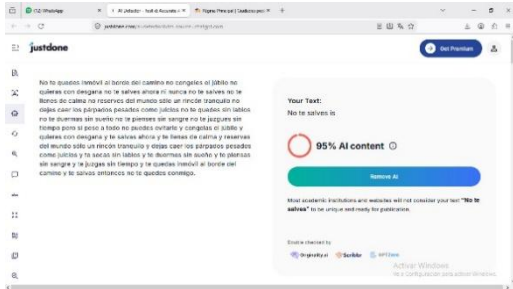
N/A

Nota:

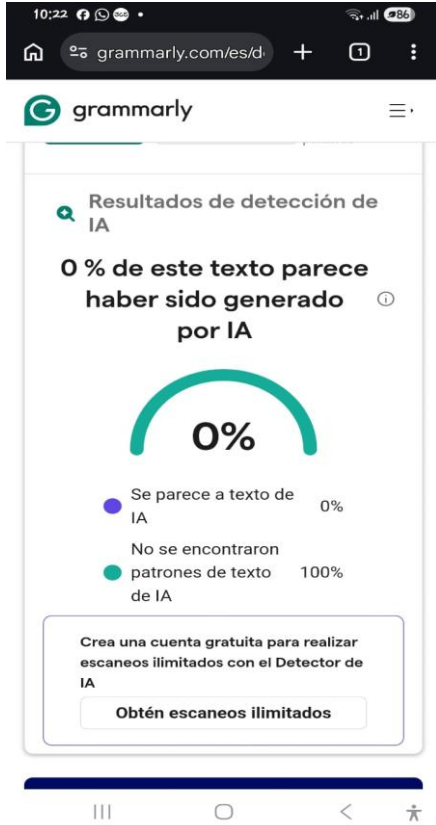
El artículo no es producto de una publicación anterior.

Anexos

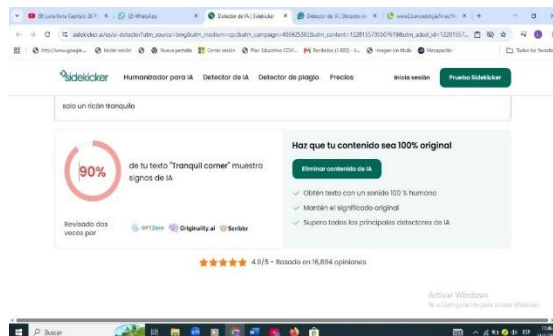
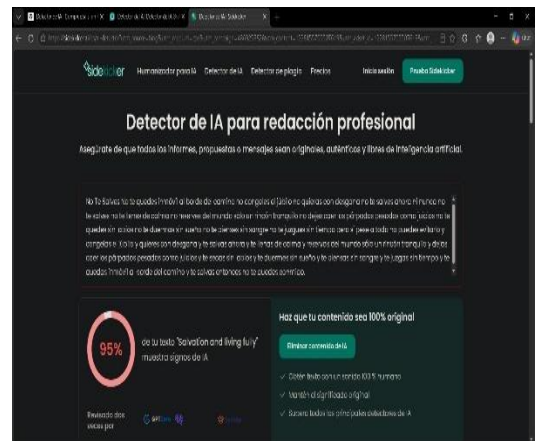
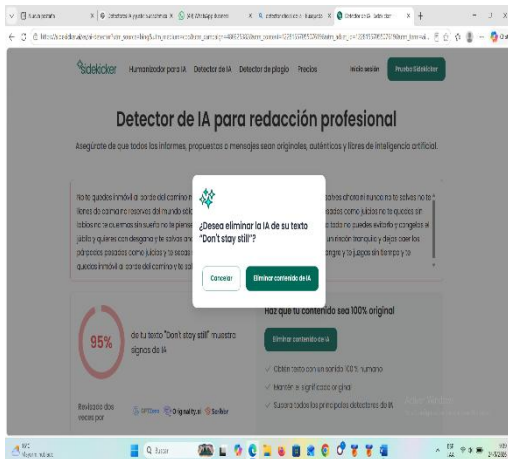
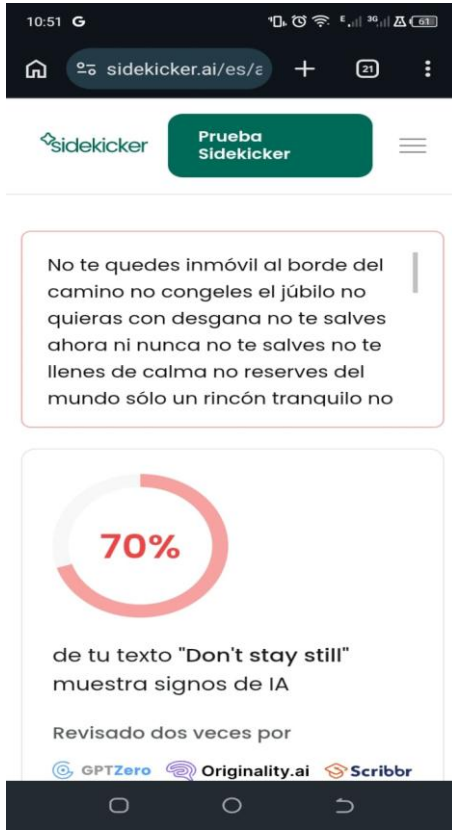
Just Done



Grammarly



Sidekicker



Zerogpt

