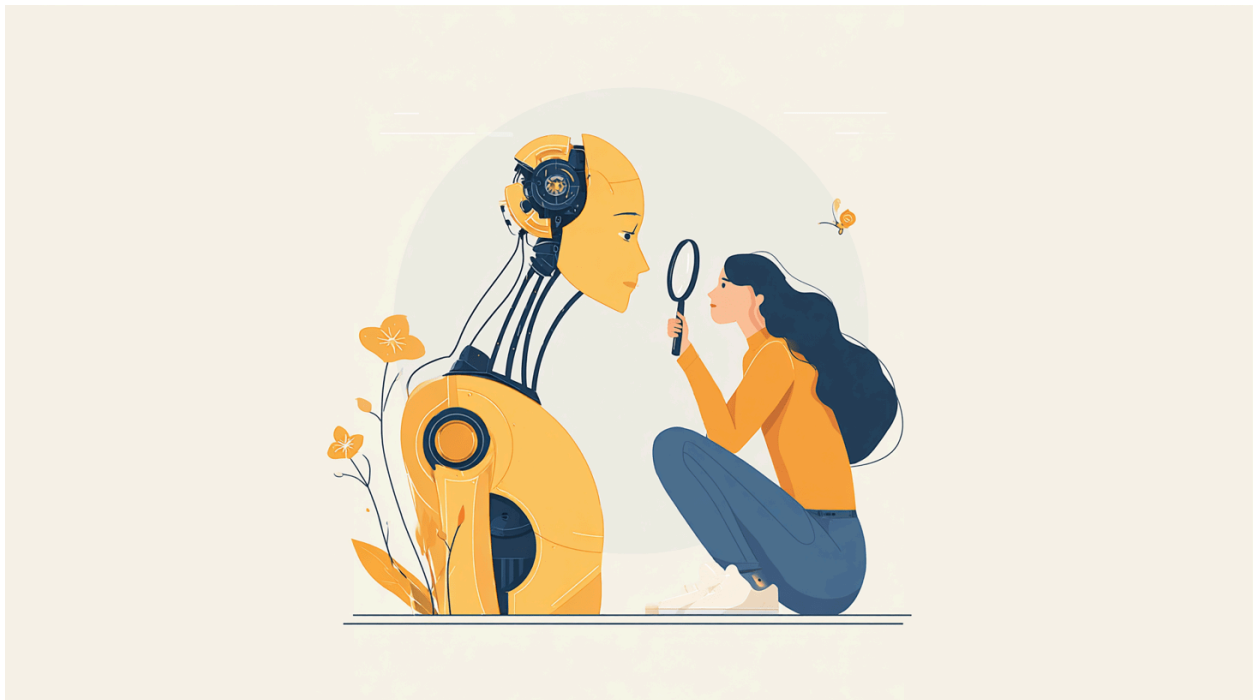


Precisión de los detectores de IA [Actualizado en marzo de 2026]



Conclusiones principales:

Las investigaciones académicas sobre la precisión de la detección mediante IA muestran que el rendimiento registrado suele oscilar entre el 83 % y el 100 %, y la mayoría de los estudios indican niveles de precisión entre el 96 % y el 99 %, lo que sugiere que los detectores de IA modernos pueden alcanzar un alto rendimiento en condiciones de prueba controladas.

Las comparaciones entre herramientas individuales muestran una variación notable en la precisión de la detección de IA, con valores que van desde el 77 % de Copyleaks hasta el 98 % de Originality.ai, mientras que Turnitin AI reporta una precisión del 92 % y GPTZero muestra una precisión del 86 % en el conjunto de datos evaluado.

El análisis de la precisión de detección de Turnitin AI muestra que el sistema detectó texto generado por IA con una precisión del 29 %, en comparación con el 83 % de Originality.ai en el mismo conjunto de datos, mientras que ambas herramientas mostraron una precisión relativamente alta a la hora de identificar texto escrito por humanos (93 % y 96 %, respectivamente).

La precisión de detección varía en función del modelo de IA que generó el texto, ya que los sistemas identificaron el contenido de ChatGPT, Grok y Gemini con una precisión del 100 %, mientras que la precisión de detección fue ligeramente inferior para GPT-3.5 (99,7 %) y GPT-4 (98,7 %).

Los resultados también muestran que el rendimiento de la detección difiere según el tipo de texto, con una precisión del 98 % para el texto generado íntegramente por IA, del 96 % para el texto escrito por humanos, del 90 % para el contenido humano editado por IA () y del 87 % para la escritura híbrida entre IA y humanos, lo que indica que la autoría mixta es más difícil de clasificar para los detectores.

El análisis de la tasa de error demuestra que los sistemas de detección de IA pueden producir errores de clasificación, con tasas de falsos positivos que oscilan entre el 2 % y el 38 % y tasas de falsos negativos que oscilan entre el 2 % y el 20 %, mientras que los revisores humanos mostraron una tasa de falsos negativos del 15 % al identificar contenido generado por IA.

Las comparaciones entre múltiples plataformas de detección indican que la precisión declarada de las herramientas líderes oscila entre el 77 % y el 99 %, con GPTZero, Smodin y Hive declarando una precisión del 99 %, Turnitin del 98 % y Quillbot AI Detector del 80 %, lo que ilustra una variación significativa en el rendimiento entre los distintos sistemas de detección de IA.

En general, los resultados muestran que la precisión de la detección de IA depende en gran medida de los conjuntos de datos, la metodología de evaluación, la fuente del modelo de IA y el tipo de texto, lo que significa que los valores de precisión declarados deben interpretarse como resultados de investigación dependientes del contexto, más que como garantías universales de rendimiento.

El rápido crecimiento de las herramientas de IA generativa ha planteado una cuestión importante en los ámbitos de la educación, la edición y los contenidos digitales: ¿son precisos los detectores de IA a la hora de identificar textos generados por IA? A medida que los sistemas de escritura con IA se vuelven más avanzados, muchas organizaciones recurren a herramientas de detección para distinguir entre contenidos escritos por humanos y generados por máquinas.

En la práctica, evaluar la precisión de la detección de IA es más complejo que un simple valor porcentual. Los investigadores suelen evaluar los detectores utilizando múltiples métricas, conjuntos de datos y entornos de prueba. Debido a esto, los resultados comunicados pueden variar significativamente en función de la metodología utilizada. Esta variación explica por qué los debates sobre la precisión de los detectores de IA suelen arrojar conclusiones diferentes.

Las tasas de precisión de los detectores de IA varían significativamente en función del conjunto de datos, el tipo de texto y la metodología de evaluación. El siguiente análisis agrega los resultados publicados en diferentes estudios académicos y pruebas comparativas.

Los valores de precisión que se analizan en este artículo representan los resultados comunicados de diferentes evaluaciones, más que una única prueba de referencia estandarizada.

En las secciones siguientes, examinamos la evidencia empírica sobre la precisión de la detección de IA, incluyendo resultados de investigaciones académicas, comparaciones entre las principales herramientas de detección y tasas de error, como los falsos positivos y los falsos negativos. Este análisis ayuda a aclarar si los detectores de IA son precisos en situaciones reales y cómo varía el rendimiento entre los distintos detectores, conjuntos de datos y tipos de texto.

Tras abordar el concepto general de la precisión de la detección de IA, resulta útil examinar la evidencia empírica de los estudios académicos que han medido la eficacia con la que los detectores de IA identifican el texto generado por IA.

Precisión de la detección de IA en estudios académicos

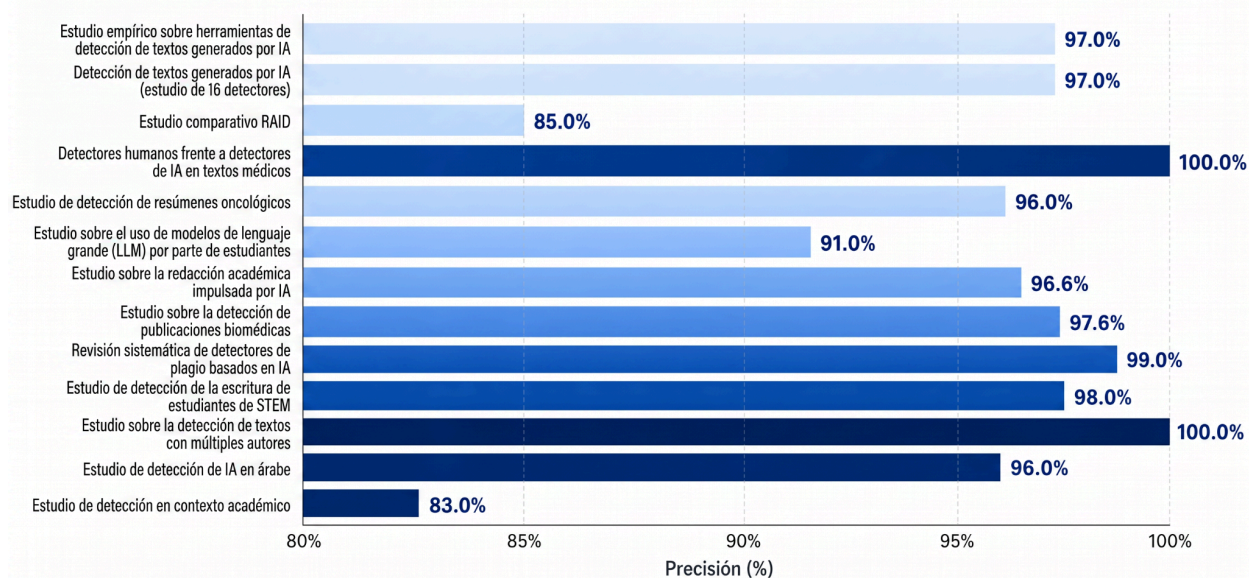
El gráfico siguiente resume los resultados publicados en múltiples estudios revisados por pares y de referencia que evalúan la precisión de la detección de IA en diferentes conjuntos de datos y contextos de investigación. Estos estudios abordan la pregunta habitual de si los detectores de IA son precisos midiendo la fiabilidad con la que las herramientas de detección distinguen entre el texto generado por IA y el escrito por humanos.

Las tasas de precisión de los detectores de IA varían significativamente en función del conjunto de datos, el tipo de texto y la metodología de evaluación. La siguiente comparación agrupa los resultados publicados en diferentes estudios académicos y pruebas comparativas.

En general, los resultados ofrecen a los lectores una visión general basada en datos, que aborda la precisión de los detectores de IA y si los sistemas actuales pueden identificar de forma consistente el contenido generado por IA.

Precisión de Detectores de IA

Análisis comparativo de herramientas de detección de textos generados por IA



- La mayor precisión de detección de IA registrada alcanzó el 100 % tanto en el estudio “Humans vs AI Detectors in Medical Writing” como en el “Multi-Authored Text Detection Study”.
- La precisión más baja en el conjunto de datos fue del 83 % en el “Estudio de detección en contexto académico”, lo que demuestra que los resultados pueden variar significativamente en función del entorno de prueba.
- La mayoría de los estudios registraron una precisión de entre el 96 % y el 99 %, incluyendo el 97,6 % en el “Estudio de detección en publicaciones biomédicas” y el 98 % en el “Estudio de detección en escritos de estudiantes de STEM”.

¿Qué grado de precisión tienen los detectores de IA?
Evidencia procedente de la investigación académica

Estudio	Precisión, %
Estudio empírico sobre herramientas de detección de textos generados por IA	97,0
Detección de textos generados por IA (estudio de 16 detectores)	97,0
Estudio comparativo RAID	85,0
Detectores humanos frente a detectores de IA en textos médicos	100,0
Estudio de detección de resúmenes oncológicos	96,0
Estudio sobre el uso de modelos de lenguaje grande (LLM) por parte de estudiantes	91,0
Estudio sobre la redacción académica impulsada por IA	96,6
Estudio sobre la detección de publicaciones biomédicas	97,6

Revisión sistemática de detectores de plagio basados en IA	99,0
Estudio de detección de la escritura de estudiantes de STEM	98,0
Estudio sobre la detección de textos con múltiples autores	100,0
Estudio de detección de IA en árabe	96,0
Estudio de detección en contexto académico	83,0 %

En los estudios analizados, la mayoría de los valores de precisión reportados se sitúan en el rango del 90-100 %, lo que indica que las herramientas modernas de detección de IA pueden alcanzar un rendimiento relativamente alto en entornos de investigación controlados. Al mismo tiempo, la variación entre el 83 % y el 100 % de precisión demuestra que la eficacia de estos sistemas depende en gran medida del conjunto de datos, el ámbito del texto y la metodología de evaluación.

Estos hallazgos ayudan a responder preguntas como qué grado de precisión tiene la IA a la hora de identificar textos generados por IA y si los detectores de IA son fiables en situaciones reales. Aunque los datos muestran que muchos sistemas obtienen buenos resultados en pruebas de referencia académicas, la variación entre los distintos estudios sugiere que los resultados de la detección de IA deben interpretarse en su contexto, en lugar de tratarse como una puntuación de precisión única y universal.

Si bien los estudios académicos ofrecen una visión general de la precisión de la detección de IA, una comparación más detallada de las herramientas individuales ayuda a responder preguntas prácticas, como si el detector de IA de Turnitin es preciso y cómo se compara su rendimiento con el de otros sistemas de detección de IA.

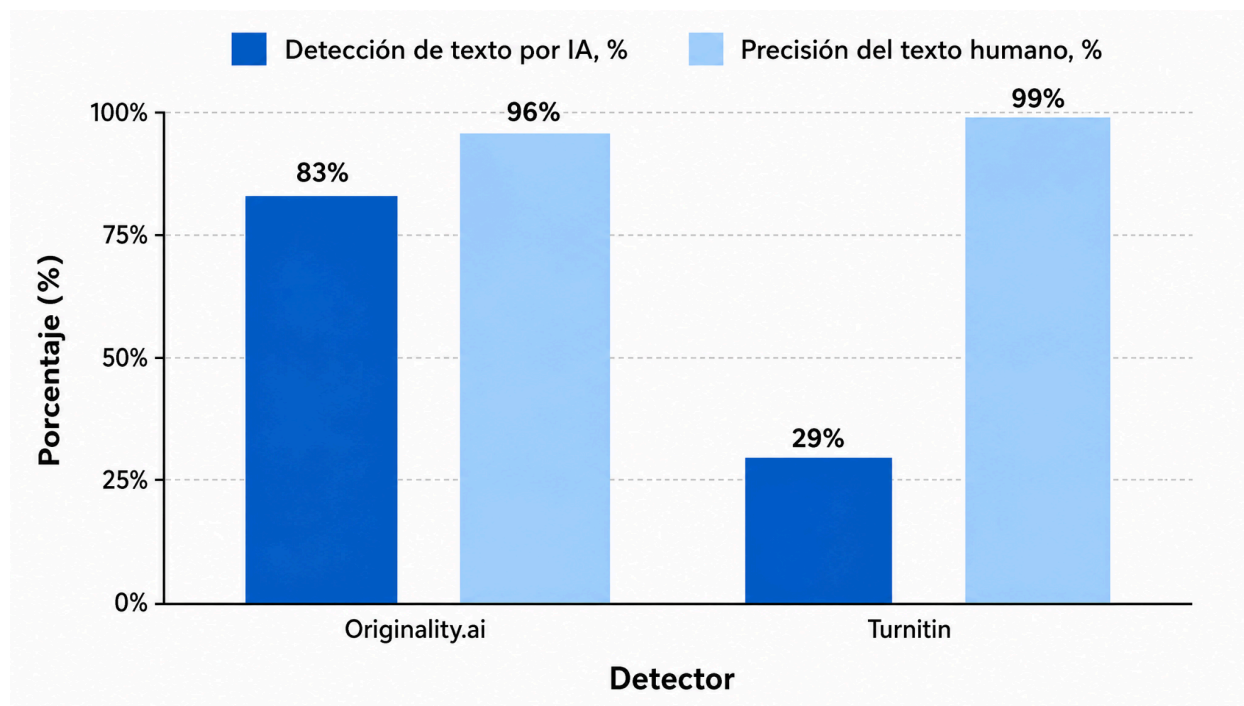
Precisión de la detección de IA de Turnitin en comparación con otros detectores de IA

El siguiente gráfico compara el rendimiento de detección de dos detectores de IA en dos métricas: la capacidad de identificar correctamente el texto generado por IA y la precisión en la clasificación del contenido escrito por humanos.

Esta comparación ayuda a abordar preguntas comunes como qué grado de precisión tiene el detector de IA de Turnitin y si su rendimiento se ajusta a la precisión de detección de IA de Turnitin descrita en evaluaciones académicas. Los resultados proceden de un estudio que probó ambos detectores en conjuntos de datos que contenían textos generados por IA, escritos por humanos y híbridos.

Las tasas de precisión de los detectores de IA varían en función del conjunto de datos, el tipo de texto y la metodología de evaluación. La siguiente comparación refleja los resultados de un estudio específico, no un punto de referencia universal.

Turnitin frente a otros detectores de IA



Originality.ai detectó el texto generado por IA con una precisión del 83 %, frente al 29 % de Turnitin en la misma evaluación.

- Turnitin identificó correctamente el texto escrito por humanos en el 93 % de los casos, un porcentaje ligeramente inferior al de Originality.ai, que fue del 96 %.
- La diferencia en la capacidad de detección de IA entre los dos sistemas alcanzó los 54 puntos porcentuales (83 % frente a 29 %) en este conjunto de datos.

¿Qué precisión tiene el detector de IA de Turnitin en comparación con otras herramientas?

Detector	Detección de texto por IA, %	Precisión del texto humano, %
Originality.ai	83 %	96 %
Turnitin	29 %	93 %

La comparación pone de relieve cómo los resultados pueden variar significativamente entre las herramientas de detección de IA a la hora de evaluar textos generados por IA. En este conjunto de datos, la precisión de detección de IA de Turnitin para identificar contenido generado por IA fue considerablemente inferior a la del sistema alternativo, a pesar de que ambas herramientas demostraron una precisión relativamente alta al clasificar textos escritos por humanos.

Estos hallazgos contribuyen a los debates en curso sobre si el detector de IA de Turnitin es preciso e ilustran por qué la precisión de detección de IA de Turnitin debe interpretarse en el contexto de condiciones de prueba específicas. En general, los datos sugieren que el rendimiento de la detección depende no solo de la herramienta en sí,

sino también del conjunto de datos y la metodología de evaluación utilizados en el estudio.

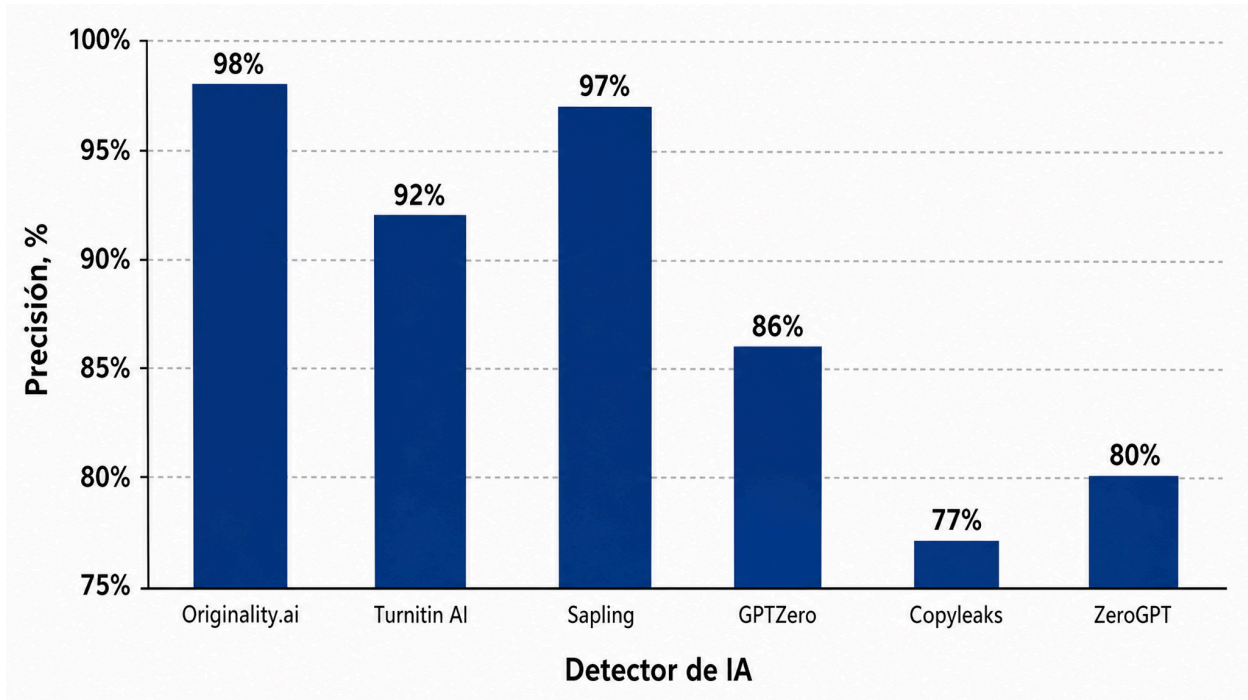
Tras examinar la precisión de la detección de IA de Turnitin en comparación con otro detector, el siguiente paso es analizar el rendimiento de varias de las principales herramientas de detección de IA en una comparación directa.

Comparación de la precisión de los principales detectores de IA

El siguiente gráfico compara la precisión de detección de IA declarada de varias herramientas de detección de IA ampliamente utilizadas. Esta comparación ayuda a responder a preguntas comunes, como cuál es el detector de IA más preciso y qué grado de precisión tienen los detectores de IA cuando se evalúan en diferentes plataformas.

Las tasas de precisión de los detectores de IA varían significativamente en función del conjunto de datos, el tipo de texto y la metodología de evaluación. La siguiente comparación agrupa los resultados comunicados en diferentes estudios académicos y pruebas comparativas.

Debido a estas diferencias, el gráfico debe interpretarse como una visión general comparativa y no como una clasificación definitiva del rendimiento de los detectores.



Originality.ai presenta la mayor precisión registrada, con un 98 %, seguido de cerca por Sapling, con un 97 %.

- Turnitin AI registra una precisión del 92 %, lo que lo sitúa entre los detectores de mayor rendimiento y las herramientas de menor rendimiento.
- Copyleaks y ZeroGPT muestran niveles de precisión más bajos, del 77 % y el 80 %, respectivamente, mientras que GPTZero registra una precisión del 86 % en los resultados evaluados.

¿Cuál es el detector de IA más preciso? Comparación de la precisión entre herramientas

Detector de IA

Precisión, %

Originality.ai	98 %
Turnitin AI	92 %
Sapling	97 %
GPTZero	86 %
Copyleaks	77 %
ZeroGPT	80 %

La comparación ilustra que la precisión de detección de IA declarada puede variar considerablemente entre los distintos detectores de IA. Mientras que algunas herramientas declaran niveles de precisión superiores al 95 %, otras muestran un rendimiento más cercano al 77-86 %, dependiendo de la evaluación.

Estas diferencias ayudan a explicar por qué cuestiones como cuál es el detector de IA más preciso siguen estando abiertas a interpretación. Dado que los resultados de la detección dependen de los conjuntos de datos, los métodos de evaluación y los tipos de texto que se analizan, los valores de precisión indicados deben considerarse indicadores comparativos más que garantías fijas de rendimiento.

Más allá de la comparación de detectores individuales, otro factor importante que afecta a la precisión de la detección de IA es el tipo de modelo de IA que generó el texto.

Precisión de la detección de IA por fuente de LLM

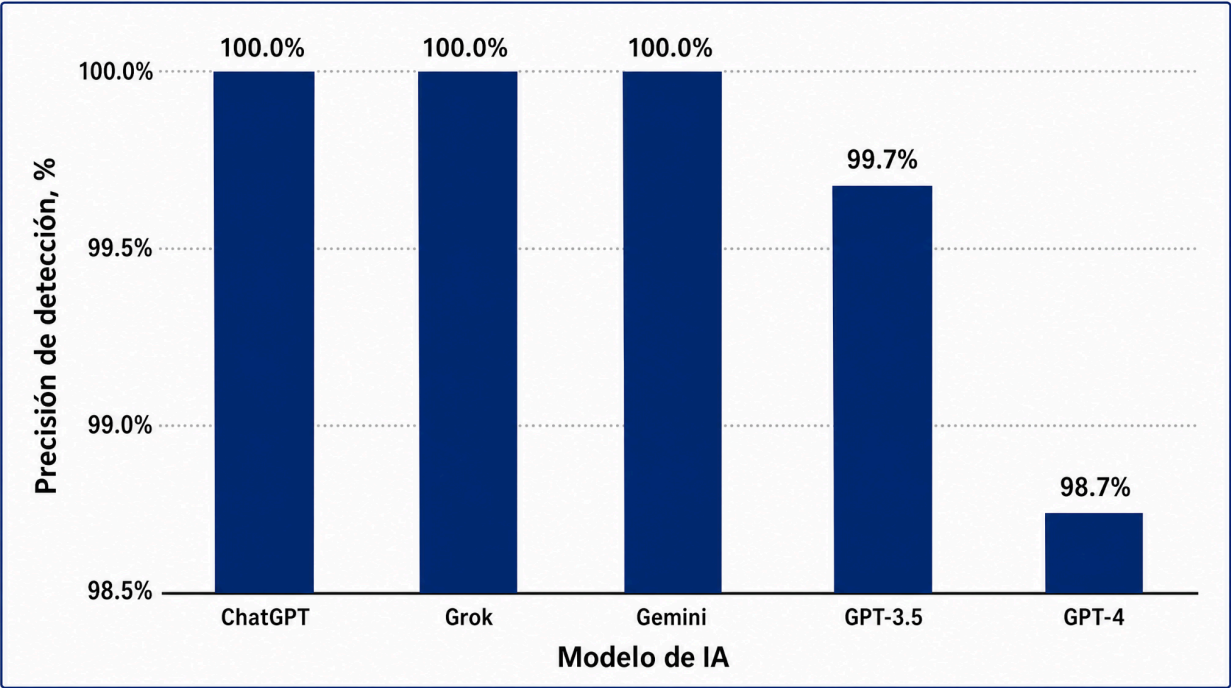
El siguiente gráfico muestra la precisión con la que los sistemas de detección de IA identifican el texto generado por diferentes modelos de lenguaje a gran escala (LLM). Evaluar el rendimiento de la detección en múltiples modelos ayuda a abordar cuestiones más amplias, como la precisión de la IA a la hora de identificar texto generado por IA procedente de diferentes fuentes.

La precisión comunicada puede variar en función del conjunto de datos, la estructura del texto y la metodología de evaluación utilizada en cada estudio. Como resultado, el rendimiento de la detección puede diferir al analizar los resultados de diferentes modelos de IA.

Los valores del gráfico representan los resultados comunicados de evaluaciones específicas, más que un único punto de referencia estandarizado.

Comprender estas diferencias ayuda a explicar por qué los debates sobre si los detectores de IA son precisos suelen depender del modelo de IA que se esté probando.

Precisión de la detección de IA por fuente de LLM



El texto generado por IA de ChatGPT, Grok y Gemini se detectó con una precisión del 100,0 % en las pruebas evaluadas.

- La precisión de detección de GPT-3.5 alcanzó el 99,7 %, lo que indica una identificación casi perfecta del contenido generado por IA.
- GPT-4 mostró una precisión de detección ligeramente inferior, del 98,7 %, aunque se mantuvo por encima del nivel del 98 %.

Cómo varía la precisión de detección entre los diferentes modelos de IA

Modelo de IA	Precisión de detección, %
ChatGPT	100,0 %

Grok	100,0 %
Gemini	100,0 %
GPT-3.5	99,7 %
GPT-4	98,7 %

Los resultados sugieren que los sistemas de detección de IA pueden alcanzar una alta precisión a la hora de identificar texto generado por las principales plataformas de LLM. En el conjunto de datos analizado, la precisión de la detección osciló entre el 98,7 % y el 100 %, dependiendo del modelo de IA.

Estos hallazgos aportan un contexto adicional a cuestiones como si los detectores de IA son precisos y con qué eficacia los sistemas de detección distinguen el contenido generado por IA. Si bien los resultados presentados demuestran una gran capacidad de detección en varios de los principales LLM, las variaciones entre los conjuntos de datos y los métodos de evaluación implican que el rendimiento en el mundo real puede variar en función de la fuente de texto específica que se analice.

Además del modelo de IA que genera el texto, otro factor que influye en la precisión de la detección de IA es el tipo de contenido que se analiza.

Precisión de la detección por tipo de texto

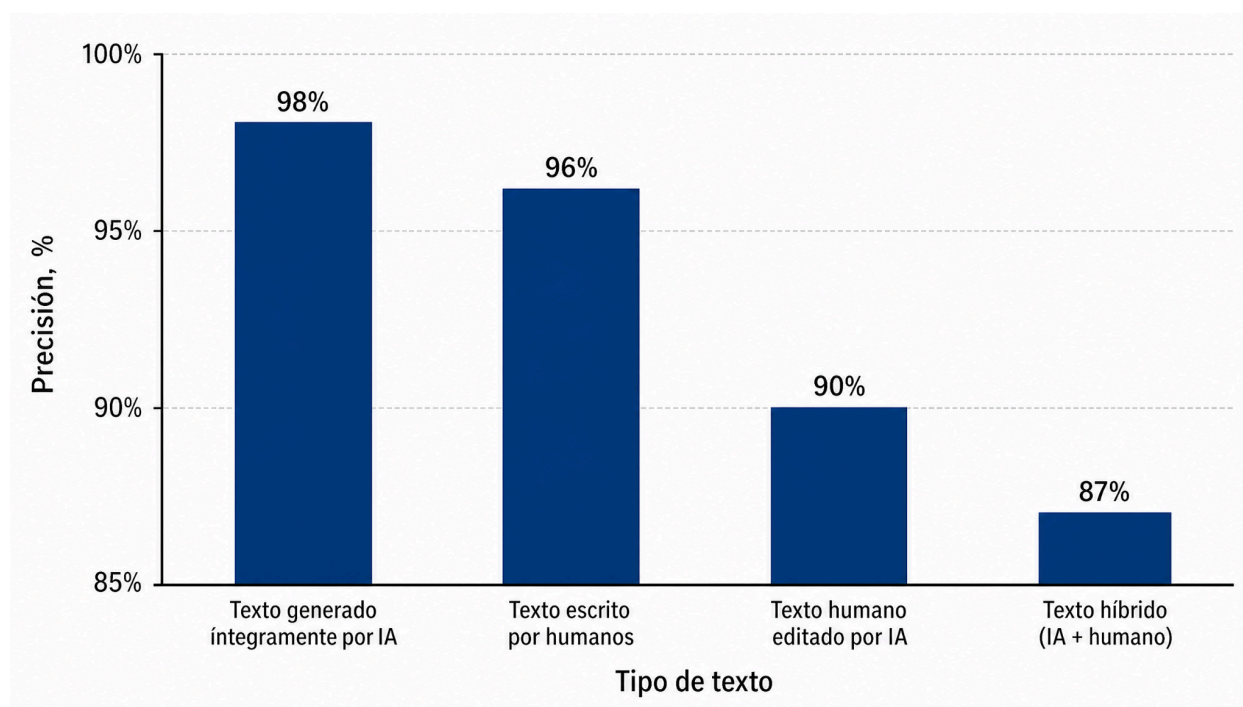
El siguiente gráfico compara el rendimiento de la detección en diferentes categorías de texto, incluyendo contenido generado íntegramente por IA, texto escrito por humanos, texto editado por IA y contenido híbrido entre IA y humanos.

Estas distinciones son importantes a la hora de evaluar la precisión de los detectores de IA, ya que el nivel de participación de la IA en el texto puede afectar significativamente a los resultados de la detección. Por ejemplo, identificar texto generado íntegramente por IA suele ser más fácil que detectar contenido parcialmente editado o híbrido.

Las tasas de precisión de los detectores de IA varían en función del conjunto de datos, la estructura del texto y la metodología de evaluación. Los valores que se muestran en el gráfico reflejan los resultados de una evaluación específica, no un punto de referencia universal.

Comprender estas diferencias también ayuda a explicar los debates en curso sobre si los detectores de IA son fiables a la hora de analizar textos mixtos o parcialmente asistidos por IA.

Precisión de detección por tipo de texto



- La precisión de detección alcanzó el 98 % para el texto generado íntegramente por IA, lo que representa el mejor rendimiento entre los tipos de texto evaluados.

- Los detectores de IA clasificaron correctamente el 96 % de los textos escritos por humanos, lo que indica una tasa relativamente baja de falsos positivos en este conjunto de datos.
- La precisión descendió al 90 % en el caso de los textos redactados por humanos y editados por IA, y al 87 % en el de los contenidos híbridos (IA y humanos), lo que demuestra que la autoría mixta es más difícil de detectar.

Cómo afecta el tipo de texto a la precisión de la detección de IA

Tipo de texto	Precisión, %
Texto generado íntegramente por IA	98 %
Texto escrito por humanos	96
Texto humano editado por IA	90
Texto híbrido (IA + humano)	87

Los resultados muestran que la precisión de la detección de IA varía en función de cómo se haya creado el contenido. Los sistemas funcionan mejor al analizar texto generado íntegramente por IA, donde la precisión de detección alcanzó el 98 %, mientras que el contenido mixto o parcialmente editado presenta mayores retos.

Estos hallazgos aportan un contexto adicional a cuestiones como la precisión de los detectores de IA y su fiabilidad a la hora de evaluar textos del mundo real. A medida que aumenta el uso de herramientas de edición asistidas por IA, distinguir entre contenido escrito por humanos, editado por IA y contenido híbrido puede convertirse en un factor clave que influya en el rendimiento de los sistemas de detección de IA.

Si bien las métricas de precisión ofrecen una visión general de la precisión de la detección de IA, evaluar los falsos positivos y los falsos negativos ayuda a responder otra pregunta importante: ¿pueden equivocarse los detectores de IA?

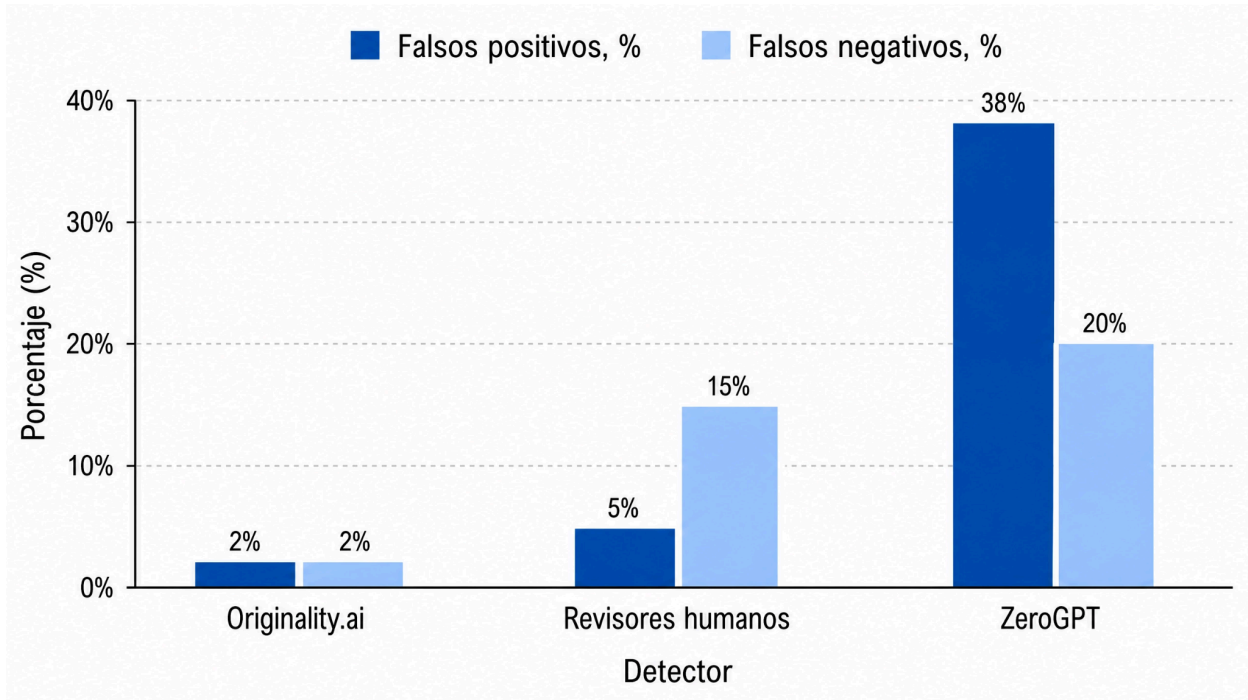
Tasas de falsos positivos y falsos negativos en la detección de IA

El siguiente gráfico compara las tasas de error de diferentes evaluadores, incluidos los detectores de IA automatizados y los revisores humanos. Estos resultados suelen aparecer en investigaciones que evalúan las metodologías de los artículos sobre detectores de IA y verificadores de IA, donde se utilizan métricas de la matriz de confusión, como los falsos positivos y los falsos negativos, para medir la fiabilidad.

Los falsos positivos se producen cuando un texto escrito por humanos se identifica erróneamente como generado por IA, mientras que los falsos negativos se producen cuando el contenido generado por IA se clasifica como escrito por humanos. Ambos tipos de errores influyen en la fiabilidad de los detectores de IA en aplicaciones del mundo real.

Las tasas de precisión de los detectores de IA varían en función del conjunto de datos y del método de evaluación. Los valores que se presentan a continuación reflejan los resultados comunicados en estudios específicos, más que un único punto de referencia estandarizado.

Índices de falsos positivos y falsos negativos



Originality.ai registró las tasas de error más bajas, con un 2 % de falsos positivos y un 2 % de falsos negativos en el conjunto de datos evaluado.

- Los revisores humanos mostraron una tasa de falsos negativos más alta, del 15 %, lo que significa que se pasó por alto el texto generado por IA en el 15 % de los casos.
- ZeroGPT mostró las tasas de error más altas, con un 38 % de falsos positivos y un 20 % de falsos negativos en la misma evaluación.

¿Pueden equivocarse los detectores de IA? Tasas de error en los sistemas de detección de IA

Detector

Falsos positivos, %

Falsos negativos, %

Originality.ai	2 %	2
Revisores humanos	5	15 %
ZeroGPT	38 %	20

Los datos muestran que evaluar la precisión de la detección de IA requiere algo más que un simple porcentaje de precisión. Los falsos positivos y los falsos negativos proporcionan información adicional sobre el rendimiento de los sistemas de detección a la hora de clasificar tanto textos generados por IA como escritos por humanos.

En este conjunto de datos, los detectores automatizados mostraron tasas de error que oscilaron entre el 2 % y el 38 %, mientras que los revisores humanos registraron una tasa de falsos negativos del 15 %. Estas diferencias ayudan a explicar por qué los debates sobre si los detectores de IA pueden equivocarse siguen siendo relevantes en la investigación académica y técnica.

En general, los resultados publicados en múltiples artículos sobre detectores de IA y estudios sobre verificadores de IA sugieren que tanto las herramientas automatizadas como los revisores humanos pueden cometer errores de clasificación, lo que refuerza la importancia de interpretar los resultados de la detección de IA en el contexto de conjuntos de datos y métodos de evaluación específicos.

Tras examinar la precisión de la detección, la variación de los conjuntos de datos y las tasas de error, resulta útil analizar cómo se comparan, en general, las diferentes herramientas de detección de IA en cuanto a la precisión comunicada.

Los detectores de contenido de IA más precisos

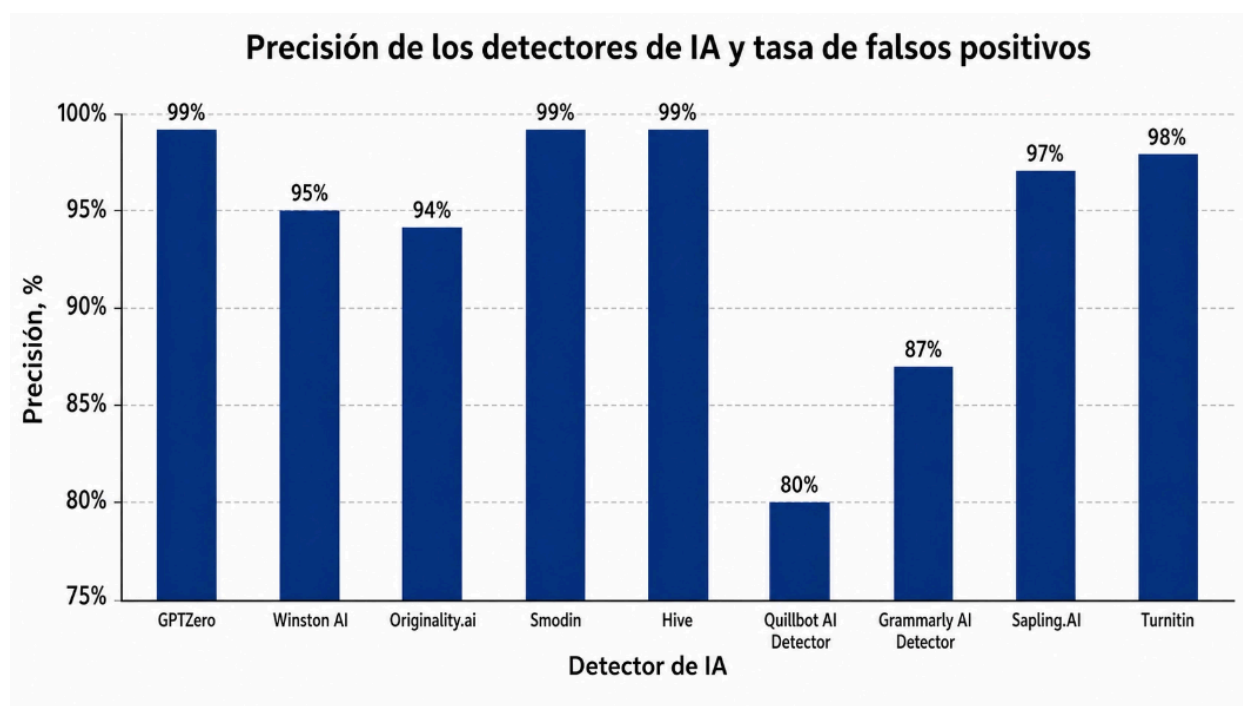
El siguiente gráfico compara varias herramientas de detección de IA ampliamente utilizadas basándose en los valores de precisión comunicados a partir de diferentes conjuntos de datos y evaluaciones. Esta comparación ayuda a abordar la pregunta

habitual de cuál es el detector de IA más preciso, al resumir los indicadores de rendimiento comunicados para las diferentes herramientas.

Las tasas de precisión de los detectores de IA varían significativamente en función del conjunto de datos, el tipo de texto y la metodología de evaluación. La siguiente comparación agrega los resultados comunicados en diferentes evaluaciones y pruebas de referencia.

Debido a estas diferencias metodológicas, las comparaciones deben interpretarse como una visión general y no como una clasificación definitiva del rendimiento de los detectores.

Los detectores de contenido de IA más precisos



GPTZero, Smodin y Hive registran los niveles de precisión más altos, con un 99 %, según los datos disponibles.

- Turnitin registra una precisión del 98 %, mientras que Sapling.AI muestra una precisión del 97 % con tasas de falsos positivos relativamente bajas.

- Algunas herramientas presentan valores de precisión más bajos, como Quillbot AI Detector, con un 80 %, y Grammarly AI Detector, con un 87 %.

¿Cuál es el detector de IA más preciso? Comparación de la precisión de las principales herramientas

Detector de IA	Precisión, %	Tasa de falsos positivos
GPTZero	99 %	Baja
Winston AI	95 %	Moderado
Originality.ai	94 %	Moderado-alto
Smodin	99 %	Moderado
Hive	99 %	Muy bajo
Detector de IA Quillbot	80 %	Moderado

Detector de IA de Grammarly	87 %	Alto
Sapling.AI	97 %	Bajo
Turnitin	98 %	Bajo

La comparación muestra que la precisión de detección de IA declarada varía sustancialmente entre las distintas plataformas de detección de IA. Algunos sistemas declaran niveles de precisión cercanos al 99 %, mientras que otros se sitúan más cerca del rango del 80-90 %, dependiendo de las condiciones de evaluación.

Estas diferencias ayudan a explicar por qué sigue siendo difícil responder de forma definitiva a preguntas como cuál es el detector de IA más preciso. El rendimiento de la detección depende no solo de la herramienta en sí, sino también del conjunto de datos, la metodología de prueba y el tipo de contenido que se analiza.

Por lo tanto, las comparaciones entre detectores deben interpretarse en el contexto de los estudios publicados, en lugar de considerarse garantías fijas de rendimiento para todos los casos de uso.

Conclusiones

- Los datos disponibles sobre la precisión de la detección de IA indican que los sistemas modernos de detección de IA pueden alcanzar un rendimiento relativamente alto en entornos de investigación controlados. En los estudios académicos analizados, los valores de precisión comunicados oscilaron entre el 83 % y el 100 %, y la mayoría de los estudios reportaron resultados entre el 96 % y el 99 %, lo que sugiere que muchos detectores de IA son capaces de identificar de forma fiable el texto generado por IA en condiciones de prueba específicas.

- Las comparaciones entre herramientas individuales muestran que la precisión de la detección de IA varía sustancialmente entre las distintas plataformas de detección. En los conjuntos de datos evaluados, la precisión reportada osciló entre el 77 % para Copyleaks y el 98 % para Originality.ai, mientras que Turnitin AI reportó una precisión del 92 % y GPTZero del 86 %, lo que indica que diferentes herramientas pueden producir resultados notablemente diferentes.
- Un análisis adicional muestra que la precisión de detección de Turnitin AI puede variar en función del conjunto de datos y del método de evaluación. En el conjunto de datos examinado en este artículo, Turnitin detectó texto generado por IA con una precisión del 29 %, frente al 83 % de Originality.ai, mientras que ambas herramientas demostraron una precisión relativamente alta a la hora de identificar texto escrito por humanos (93 % y 96 %, respectivamente).
- El rendimiento de la detección también varía en función del modelo de IA que generó el texto. En los resultados evaluados, los detectores de IA identificaron el texto producido por ChatGPT, Grok y Gemini con una precisión del 100 %, mientras que se registraron tasas de detección ligeramente inferiores para GPT-3.5 (99,7 %) y GPT-4 (98,7 %), lo que demuestra que los resultados de la detección pueden diferir en función de la fuente del LLM.
- El análisis muestra además que la estructura del texto y el tipo de autoría influyen en el rendimiento de la detección. La precisión de la detección alcanzó el 98 % para el texto generado íntegramente por IA, el 96 % para el texto escrito por humanos, el 90 % para el contenido humano editado por IA y el 87 % para la escritura híbrida entre IA y humanos, lo que indica que el contenido mixto o parcialmente asistido por IA puede resultar más difícil de clasificar correctamente para los detectores.

- El análisis de la tasa de error indica que tanto los sistemas automatizados como los evaluadores humanos pueden producir errores de clasificación. En el conjunto de datos analizado, las tasas de falsos positivos oscilaron entre el 2 % y el 38 %, mientras que las de falsos negativos oscilaron entre el 2 % y el 20 %, y los revisores humanos mostraron una tasa de falsos negativos del 15 %, lo que ilustra que los resultados de la detección mediante IA no están exentos de errores.
- En general, los resultados sugieren que la precisión de la detección mediante IA depende en gran medida del conjunto de datos, la metodología de evaluación, la fuente del modelo de IA y el tipo de texto que se analiza. Por lo tanto, los valores de precisión comunicados deben interpretarse como resultados dependientes del contexto, más que como garantías universales de rendimiento, a la hora de evaluar si los detectores de IA pueden identificar de forma fiable el contenido generado por IA.

Fuentes

- Barlow, Written. “Los 9 mejores detectores de IA con mayor precisión en 2026”. Recursos de detección de IA | GPTZero, 2 de enero de 2026, <https://gptzero.me/news/best-ai-detectors/>. Consultado el 16 de marzo de 2026.
- Gillham, Jonathan. “Estudios sobre la precisión de la detección de IA: metaanálisis de 13 estudios – Originality.AI”. Originality.AI, <https://originality.ai/blog/ai-detection-studies-round-up>. Consultado el 16 de marzo de 2026.

- “Tenemos una precisión del 99 % en la detección de IA: estudio de Originality.Ai – Originality.AI”. Originality.AI, <https://originality.ai/blog/ai-accuracy>. Consultado el 16 de marzo de 2026.