



# Los retos de la lingüística forense en la era de la IA

## The Challenges of Forensic Linguistics in the Age of AI

### Os retos da lingüística forense na era da IA

**Sheila Queralt**

*Laboratorio SQ-Lingüistas Forenses, Barcelona, España*

[sheila.queralt@cllicenciats.cat](mailto:sheila.queralt@cllicenciats.cat)

<https://orcid.org/0000-0002-0641-0727>

#### Resumen

En este trabajo, se analiza los desafíos que enfrenta la lingüística forense en la era de la inteligencia artificial. El objetivo es describir la generación automática de texto y voz mediante IA, y analizar cómo esto afecta la atribución de autoría y la autenticación de pruebas en contextos judiciales. Además, se examinan las herramientas de detección actuales y sus limitaciones. En los resultados, se evidencia que las metodologías tradicionales pueden no ser suficientes para detectar textos generados por IA, y se destaca la necesidad de crear nuevas herramientas y enfoques interdisciplinarios. La investigación subraya la importancia de desarrollar marcos éticos y legales que regulen el uso de la IA en procesos judiciales.

**Palabras clave:** lingüística forense; inteligencia artificial; atribución de autoría; *deepfakes*; textos híbridos.

#### Abstract

In this article, the challenges faced by forensic linguistics in the era of Artificial Intelligence are analyzed. It focuses on describing the automatic generation of text and voice using AI and examining how this impacts authorship attribution and the authentication of evidence in judicial contexts. Additionally, it reviews current detection tools and their limitations. The findings suggest that traditional methodologies may not suffice to detect AI-generated texts, emphasizing the need to develop new tools and interdisciplinary approaches. The research underscores the importance of developing ethical and legal frameworks to regulate the use of AI in judicial processes.

**Keywords:** Forensic Linguistics; Artificial Intelligence; authorship attribution; deepfakes; hybrid texts.

#### Resumo

Neste artigo, analisam-se os desafios enfrentados pela lingüística forense na era da Inteligência Artificial. O foco está em descrever a geração automática de texto e voz por meio da IA e examinar como isso impacta a atribuição de autoria e a autenticação de provas em contextos judiciais. Além disso, examina as ferramentas de detecção atuais e suas limitações. Os resultados indicam que as metodologias tradicionais podem não ser suficientes para detectar textos gerados por IA, destacando a necessidade de desenvolver novas ferramentas e abordagens interdisciplinares. A pesquisa ressalta a importância de desenvolver marcos éticos e legais que regulem o uso da IA em processos judiciais.

**Palavras-chave:** Linguística Forense; Inteligência Artificial; atribuição de autoria; deepfakes; textos híbridos.

Recibido: 11/11/2023

Aceptado: 13/08/2024

Publicado: 30/12/2024

## 1. Introducción

La lingüística forense es una disciplina que utiliza principios lingüísticos para analizar el lenguaje en contextos legales y criminales. Desde su surgimiento en las décadas pasadas, ha sido esencial en casos que involucran la autenticación de documentos, análisis de amenazas, suplantación de identidad y peritajes sobre el estilo lingüístico de los sospechosos (Coulthard y Johnson, 2007). Esta disciplina se ha vuelto indispensable en los sistemas judiciales, donde el análisis del lenguaje es clave para probar la autenticidad de documentos y determinar la autoría de textos en casos de plagio, fraude o amenazas.

Sin embargo, en los últimos años, la inteligencia artificial (IA) ha propiciado cambios radicales. Las herramientas de IA generativa son ahora capaces de producir texto y voz con niveles de precisión y realismo sorprendentes, lo que desafía a los expertos en lingüística forense a encontrar maneras de distinguir entre el lenguaje humano y el generado automáticamente. La capacidad de estas tecnologías para imitar el lenguaje humano complica la distinción entre contenido generado por máquinas y por personas, lo que exige el desarrollo de nuevos métodos y enfoques en el ámbito forense (Franganillo, 2023). Esta situación ha creado una nueva frontera en la disciplina, donde las tecnologías avanzadas como el Procesamiento del Lenguaje Natural (NLP, por sus siglas en inglés) permiten a las IAs imitar estilos de escritura, crear identidades falsas y fabricar pruebas digitales. Estas capacidades plantean desafíos críticos para la autenticidad y la atribución en entornos legales, ya que técnicas como la estilometría adversarial pueden ser empleadas para modificar patrones de escritura y evadir la detección de autoría (Brennan *et al.*, 2012).

El uso de IA para la generación automática de texto y voz no solo ha facilitado el desarrollo de herramientas útiles en múltiples sectores, sino que también ha sido explotado para fines ilícitos. Casos recientes documentan cómo la IA ha sido utilizada para realizar fraudes, suplantaciones de identidad y manipulación de información (Perkins, 2021; Sousa-Silva, 2022). Esto ha llevado a los lingüistas forenses a enfrentarse a nuevas preguntas y desafíos. Como señala Queralt (2023, p. 268), “los avances en los métodos utilizados por los ciberdelincuentes también afectan al trabajo de los lingüistas forenses”. Los métodos tradicionales de análisis forense, que se basan en patrones estilísticos y atributos humanos del lenguaje, deben adaptarse y evolucionar para seguir siendo efectivos en la detección de estos nuevos tipos de fraudes.

Además, los avances en IA han permitido la creación de *deepfakes* de voz, que replican la entonación, el acento y el estilo de habla de una persona con un grado de precisión tal que puede engañar, incluso, a los expertos (Maras y Alexandrou, 2018). Este fenómeno presenta graves riesgos para la integridad de los testimonios y pruebas presentadas en los tribunales.

En el presente artículo, se tiene como objetivo examinar el impacto transformador de la inteligencia artificial en la lingüística forense y analizar los desafíos que surgen en la adaptación de esta disciplina en un entorno tecnológico en constante evolución. En un contexto donde las herramientas de IA son capaces de generar pruebas falsas con un grado de precisión comparable al humano, resulta imperativo considerar cómo los lingüistas forenses pueden garantizar la fiabilidad y validez de sus análisis en el ámbito legal. Asimismo, se plantea la necesidad de identificar y desarrollar nuevos métodos y herramientas que permitan abordar eficazmente estos retos emergentes.

En este artículo, se profundiza en los principales retos que plantea la inteligencia artificial para la lingüística forense, se presentan algunas de las tendencias actuales en cuanto a la generación automática de texto y voz, y se examina cómo la disciplina debe adaptarse para seguir cumpliendo su rol en el ámbito judicial.

## 2. El impacto actual de la inteligencia artificial en la lingüística forense

La lingüística forense ha jugado un papel clave en la resolución de crímenes a través del análisis de textos y del habla humana en contextos legales. Sin embargo, con la creciente prevalencia de la inteligencia artificial (IA) en la generación de texto y voz, se enfrenta a nuevos y complejos retos. Las tecnologías avanzadas de Procesamiento de Lenguaje Natural (PLN), *deep learning* y redes neuronales han revolucionado la manera en que se crea y manipula el lenguaje, lo que afecta profundamente la capacidad de los expertos para distinguir entre el lenguaje generado por humanos y el generado por máquinas (Ji *et al.*, 2024).

En la actualidad, no existen estadísticas uniformes sobre la cantidad exacta de delitos cometidos mediante la generación de voz o texto por IA. Esto se debe a la dificultad de rastrear y clasificar estos delitos, que a menudo pasan desapercibidos o se mezclan con otras formas de delincuencia cibernética. Sin embargo, estudios recientes apuntan a un aumento significativo en el uso de estas tecnologías para actividades delictivas como fraudes, suplantación de identidad, desinformación y manipulación de evidencias. En particular, el informe de *NTT* del primer semestre de 2024 indica que el uso de la inteligencia artificial en ciberataques aumentó un 600% en el primer semestre de 2024 en comparación con la segunda mitad de 2023, duplicando el costo de los ataques a 10.000 millones de euros (Departamento de Cyber Threat Intelligence de NTT DATA Cybersecurity, 2024).

Un ejemplo preocupante es el uso de IA para la creación de *deepfakes*; es decir, contenido multimedia manipulado mediante IA, que ha sido empleado en la manipulación de videos y voces con fines maliciosos. El Federal Bureau of Investigation (FBI) advirtió en 2024 que los cibercriminales están utilizando esta tecnología para ejecutar fraudes, difundir desinformación y realizar ataques de suplantación de identidad con una precisión nunca antes vista (FBI, 2024). La capacidad de las IAs para generar voces con gran realismo y crear vídeos falsos ha generado una crisis de confianza en las pruebas digitales, las cuales son fundamentales en los juicios modernos.

La suplantación de identidad y los fraudes financieros son otros delitos que se han visto exacerbados por la IA. En el ámbito empresarial, los *deepfakes* de voz han permitido a los criminales convencer a empleados de alto rango de realizar transferencias financieras fraudulentas. Un caso notorio ocurrió en 2019, cuando unos delincuentes utilizaron un *deepfake* de voz para imitar al CEO de una compañía alemana, lo que logró que se transfirieran más de 243 000 dólares a una cuenta controlada por los criminales (Sophos, 2019).

La creciente capacidad de los modelos de IA para generar textos sintéticos altamente realistas es otro de los desafíos emergentes. En el caso del phishing, los atacantes utilizan IA para crear correos electrónicos falsos con gramática impecable y un estilo que imita a personas conocidas por la víctima, lo que aumenta la efectividad del engaño. Estas técnicas, alimentadas por la capacidad de las IA de analizar grandes volúmenes de datos, hacen que los delincuentes puedan crear mensajes

extremadamente personalizados y difíciles de detectar por los sistemas de seguridad tradicionales (Europol, 2024).

La difusión de desinformación es otro frente crítico donde la IA ha transformado las amenazas existentes. Las campañas de desinformación, alimentadas por textos generados automáticamente, han alcanzado una escala sin precedentes, con el potencial de influir en elecciones, manipular la opinión pública y generar conflictos sociales. El uso de modelos avanzados de IA para generar contenido automatizado (como el mencionado anteriormente) plantea riesgos para la percepción pública, como la creación de noticias falsas o publicaciones que manipulan opiniones sociales (Zeng *et al.*, 2024). En 2024, se documentó un caso donde una red de bots operada por una agencia gubernamental rusa utilizó IA para generar perfiles falsos en redes sociales, con el objetivo de difundir noticias falsas y propaganda política (Departamento de Justicia de EE.UU., 2024).

Ante este panorama, es evidente que los delitos relacionados con la generación de texto y voz mediante IA presentan retos únicos para las autoridades, las empresas y los lingüistas forenses. A medida que las tecnologías de IA avanzan, los métodos tradicionales de análisis forense no son suficientes para hacer frente a las nuevas formas de fraude y manipulación digital. La adaptación a este nuevo entorno requiere un enfoque multidisciplinario que involucre tanto a expertos en IA como a profesionales del ámbito legal y forense, para abordar los desafíos éticos, sociales y legales que plantea su implementación (Floridi *et al.*, 2021).

### 3. Hallazgos recientes y tendencias en el uso de la IA para delitos

La investigación reciente en el ámbito de la inteligencia artificial ha revelado tendencias preocupantes sobre cómo las tecnologías avanzadas se están utilizando con fines delictivos. Estas tendencias presentan nuevos retos para la lingüística forense, ya que el uso de IA en la creación de textos y voces sintéticas está complicando los métodos tradicionales de atribución y autenticación. A continuación, se describen los hallazgos más destacados y las tendencias actuales en el uso malicioso de la IA.

#### 3.1. Phishing o suplantación de identidad

El phishing ha sido uno de los métodos más comunes de ciberdelincuencia durante años, pero la integración de herramientas de inteligencia artificial ha incrementado notablemente su efectividad. Los cibercriminales están utilizando sistemas de procesamiento del lenguaje natural (PLN) para crear correos electrónicos y mensajes altamente personalizados, diseñados para engañar a las víctimas (Hadnagy y Fincher, 2015). La IA puede analizar grandes cantidades de datos sobre el lenguaje de las personas y replicar su estilo, de manera que logra suplantaciones casi indetectables.

Un informe reciente del FBI reveló que los ataques de phishing impulsados por IA se han vuelto mucho más complejos, con mensajes creados con gramática impecable y realismo en la estructura, lo que aumenta la probabilidad de éxito de los ataques. Estos mensajes se utilizan para robar datos sensibles, como contraseñas y credenciales financieras, y para autorizar transferencias fraudulentas. La creciente personalización de los ataques mediante clonación de voz también ha sido reportada, donde las IA son capaces de generar voces que imitan a personas conocidas por las víctimas (FBI, 2024).

### 3.2. Deepfakes y desinformación

La tecnología de *deepfake* ha sido un área de especial preocupación debido a su capacidad para manipular tanto imágenes como audio. Los *deepfakes* permiten a los actores malintencionados crear vídeos falsos que representan a personas que dicen o hacen cosas que nunca ocurrieron. Esto ha tenido implicaciones graves en la difusión de desinformación, la difamación de figuras públicas y la alteración de procesos políticos.

Un informe de Europol en 2024 indicó que los *deepfakes* se están utilizando para influir en elecciones y desacreditar a personas públicas. Esta tecnología ha sido empleada en varios países para fabricar vídeos que parecen auténticos, lo que ha dificultado la identificación de contenidos falsos, incluso, para expertos en verificación de datos (Europol, 2024). La accesibilidad de estas tecnologías ha permitido que actores con recursos limitados puedan crear *deepfakes* sofisticados, lo que incrementa la proliferación de este tipo de delitos.

Además, la difusión de noticias falsas y teorías de conspiración a través de textos generados por IA también ha sido identificada como una tendencia alarmante. En las redes sociales, los bots impulsados por IA han sido utilizados para generar miles de publicaciones que fomentan la desinformación y manipulan la opinión pública. Por ejemplo, en 2024, el Departamento de Justicia de EE.UU. reveló una operación de una “granja de bots” rusa que usaba IA para crear perfiles falsos y difundir propaganda política en plataformas sociales como X (anteriormente Twitter), lo que afectó directamente el discurso público (Departamento de Justicia de EE.UU., 2024).

### 3.3. Fraude financiero

Otro aspecto crítico es el uso de la inteligencia artificial para cometer fraudes financieros. Un ejemplo notorio de esto es el uso de *deepfakes* de voz en fraudes corporativos. En 2019, un grupo de criminales utilizó un software de inteligencia artificial para imitar la voz de un ejecutivo de una empresa alemana, y engañó a un subordinado en el Reino Unido para que realizara una transferencia de 243 000 dólares a una cuenta controlada por los delincuentes (Sophos, 2019). Lo alarmante es que la IA no solo imitó el tono y el timbre de la voz del ejecutivo, sino también su acento y las pausas características de su forma de hablar, lo que convenció por completo a la víctima.

Este tipo de fraude ha ido en aumento, ya que la IA permite a los delincuentes crear voces extremadamente convincentes, reduciendo el tiempo y los costos que antes se necesitaban para llevar a cabo este tipo de ataques. La facilidad con la que se pueden fabricar estos engaños mediante IA ha llevado a un incremento en el número de fraudes financieros reportados en varios países, lo que pone en riesgo no solo a las empresas, sino también a los sistemas financieros globales (ASIS International, 2024).

### 3.4. Manipulación de mercados y opinión pública

La manipulación de mercados y la influencia en la opinión pública mediante IA también ha ganado relevancia. Los sistemas de IA pueden generar reseñas falsas, comentarios positivos o negativos sobre productos, e incluso artículos de opinión sesgados. Esto se hace con el fin de influir en las decisiones de compra o en la percepción pública de empresas o individuos.

Un caso reciente implicó a la empresa Rytr, que fue acusada por la Federal Trade Commission (FTC) de facilitar la creación de reseñas falsas a través de su herramienta de generación automática de texto. Estas reseñas, generadas sin base en experiencias reales, engañaron a muchos consumidores, lo que afectó sus decisiones de compra (FTC, 2024).

### 3.5. Amenazas y extorsión

Finalmente, el uso de la IA para generar amenazas y extorsiones es una tendencia que también ha emergido. En abril de 2023, en Arizona, un grupo criminal utilizó una tecnología de clonación de voz para imitar los gritos de una adolescente y extorsionar a su madre, exigiendo un rescate de un millón de dólares. La madre recibió una llamada en la que creyó escuchar a su hija, pero afortunadamente logró verificar que su hija estaba a salvo antes de realizar el pago. Este caso destaca los peligros del uso de IA en actividades de extorsión y secuestro virtual (Gómez, 2023).

## 4. Desafíos del análisis forense de textos generados por inteligencia artificial

La creciente sofisticación de los modelos de inteligencia artificial ha introducido importantes desafíos para los lingüistas forenses. El análisis tradicional del lenguaje, que solía centrarse en identificar patrones estilísticos únicos de los autores humanos, ahora enfrenta un escenario radicalmente diferente. Los modelos de inteligencia artificial generativa, como los basados en redes neuronales profundas y el aprendizaje no supervisado, son capaces de producir textos que no solo imitan el lenguaje humano, sino que en muchos casos lo hacen de manera indistinguible. Esto genera varios problemas para los peritos lingüistas, tanto en términos de atribución como de autenticación de textos (Gehrmann *et al.*, 2019).

Los principales desafíos que enfrentan los lingüistas forenses en relación con la inteligencia artificial incluyen la imitación de patrones lingüísticos humanos, que dificulta la distinción entre textos generados por humanos y máquinas; la manipulación automatizada de textos, que complica la identificación de fuentes originales; la falta de herramientas específicas para el análisis forense, especialmente en textos híbridos; la detección de errores y anomalías semánticas, que suelen ser sutiles en textos generados por IA; y la ausencia de consensos metodológicos, lo que limita la fiabilidad de los análisis en contextos legales. Estos problemas se desarrollan y analizan en detalle a continuación.

### 4.1. Imitación de patrones lingüísticos humanos

Los modelos avanzados de IA, como GPT-3, GPT-4 y otras variantes de transformers, han demostrado una capacidad notable para generar textos que imitan los patrones lingüísticos humanos, tanto a nivel gramatical como estilístico. Estos sistemas son entrenados con grandes cantidades de datos textuales, lo que les permite aprender no solo la sintaxis del lenguaje, sino también características sutiles como el uso de conectores, la estructura narrativa y el tono (Brown *et al.*, 2020).

El principal problema para los lingüistas forenses es que estos modelos son capaces de producir textos que pueden ser confundidos con escritos humanos. En muchos casos, la capacidad de la IA para generar textos coherentes y bien estructurados hace difícil determinar si el autor real es una persona o una máquina. Esto plantea un reto significativo en la autenticación de documentos y en casos donde la autoría de un texto es clave para el desarrollo de un proceso judicial. Los avances en el

análisis de autoría forense han enfrentado desafíos con la introducción de tecnologías modernas, que complican la atribución precisa de autoría y la autenticidad de los textos presentados como evidencia (Grant, 2022; Fobbe, 2019).

Por ejemplo, en un caso de fraude empresarial, el uso de IA para generar correos electrónicos de phishing personalizados ha dificultado a los expertos determinar si el mensaje fue realmente escrito por un empleado o por un modelo de IA que suplantó la identidad del remitente.

#### 4.2. Manipulación de texto automatizado

La IA no solo es capaz de generar texto de manera autónoma, sino que también puede modificar textos preexistentes. Este tipo de manipulación textual automatizada complica aún más la labor de los expertos forenses. Los algoritmos pueden reescribir o parafrasear textos, lo que hace extremadamente difícil identificar la fuente original de un documento o establecer conexiones entre textos aparentemente diferentes, pero que han sido manipulados por la IA. La creciente interacción entre humanos y máquinas en la creación de textos plantea desafíos importantes en contextos legales, donde distinguir entre contribuciones humanas y generadas por IA resulta crítico (Muñoz-Basols y Fuertes Gutiérrez, 2024). Este fenómeno subraya la necesidad de herramientas avanzadas capaces de identificar estos híbridos textuales.

Además, la edición de textos híbridos —donde los humanos ajustan pequeños fragmentos de textos generados por IA para hacerlos parecer más naturales— introduce otro nivel de complejidad. Zeng *et al.* (2024) destacan que estos textos híbridos, resultado de la colaboración entre humanos e IA, representan uno de los mayores desafíos para los detectores actuales de IA, ya que suelen fallar al enfrentar textos editados o parcialmente generados por humanos debido a la alteración de patrones predecibles.

#### 4.3. Falta de herramientas específicas para el análisis forense de IA

A pesar de los avances en herramientas diseñadas para la detección de texto generado por IA, muchas de estas herramientas no están optimizadas para su uso en el campo de la Lingüística Forense. Herramientas como GPTZero y DetectGPT han demostrado ser útiles para identificar textos generados por IA con altos niveles de precisión cuando el texto es completamente sintético. Sin embargo, su efectividad disminuye drásticamente cuando se trata de textos híbridos, donde partes han sido creadas por IA y otras han sido editadas por humanos (Chakraborty, *et al.*, 2023; Zeng, *et al.*, 2024). Es más, el estudio de Mindner *et al.* (2023), que investiga la detección de textos reescritos por IA, concluye que la clasificación es más difícil cuando la IA ha sido instruida para crear el texto de manera que un humano no reconozca que fue generado por una IA, lo que sugiere que la edición humana de textos generados por IA puede confundir todavía más a los detectores.

Además, aunque existen herramientas de análisis estilométrico como JGAAP y Stylometry with R, estas están más enfocadas en la atribución de autoría humana y no en la detección de textos generados por IA. Por lo tanto, su uso en casos que involucran IA sigue siendo limitado. Esto resalta la necesidad de desarrollar nuevas herramientas que no solo sean capaces de identificar patrones estilísticos, sino también de reconocer irregularidades y anomalías propias de textos generados por IA.

#### 4.4. Detección de errores y anomalías semánticas

Aunque los textos generados por IA son generalmente coherentes y gramaticalmente correctos, los modelos actuales aún presentan algunas anomalías semánticas que pueden ser detectadas por lingüistas forenses entrenados. Estos errores incluyen incoherencias en el uso de expresiones idiomáticas, cambios abruptos de contexto y una tendencia a repetir ciertas estructuras sintácticas o frases. Los humanos tienden a introducir una mayor variabilidad en su estilo y contenido, mientras que las IA suelen generar patrones más predecibles (Giulianelli *et al.*, 2023).

No obstante, estos errores no siempre son suficientes para establecer una identificación concluyente. En muchos casos, las correcciones automáticas y los algoritmos de automejora utilizados por las IAs más avanzadas minimizan estos errores, lo que propicia que los textos generados por máquinas se acerquen aún más a los escritos por humanos. Esto reduce significativamente la capacidad de los lingüistas forenses para identificar anomalías de manera confiable.

#### 4.5. Falta de consensos en metodologías y estándares

Un problema adicional es la ausencia de consensos claros en las metodologías utilizadas para analizar textos generados por IA. A diferencia de la atribución de autoría tradicional, que cuenta con décadas de investigación y estándares consolidados, la detección de textos generados por IA es un campo emergente que aún carece de acuerdos metodológicos sólidos. Los diferentes enfoques y herramientas a menudo arrojan resultados inconsistentes, lo que compromete la fiabilidad de los análisis en contextos legales.

La variabilidad en los resultados de las herramientas disponibles hace que sea difícil presentar pruebas concluyentes en los tribunales. Esto pone en riesgo la validez de las pruebas basadas en análisis lingüísticos basados exclusivamente en herramientas computacionales y que no incluyen un análisis cualitativo realizado por un experto.

### 5. Herramientas tecnológicas para la detección de textos generados por IA

Con el aumento de la generación automática de texto mediante inteligencia artificial, la necesidad de herramientas especializadas para detectar y analizar estos textos ha crecido exponencialmente. En el campo de la lingüística forense, estas herramientas son esenciales no solo para identificar textos creados por IA, sino también para atribuir autoría y garantizar la integridad de las pruebas presentadas en los tribunales. Sin embargo, la precisión y fiabilidad de estas herramientas varía significativamente, especialmente en el contexto de textos híbridos, donde los humanos han editado o ajustado textos generados por IA. Herramientas como GLTR, descritas por Gehrmann *et al.* (2019), han sentado las bases para la creación de detectores modernos como GPTZero, que emplean análisis estadísticos para diferenciar entre texto humano y generado automáticamente.

A continuación, se presenta una descripción de algunas de las herramientas más utilizadas en general para la detección de textos generados por IA, aunque no necesariamente optimizadas ni ampliamente adoptadas en el ámbito de la Lingüística Forense.

### 5.1. GPTZero

GPTZero es una de las herramientas más populares para la detección de textos generados por modelos de IA como GPT-3 y GPT-4. Desarrollada para identificar características estadísticas que varían entre el texto humano y el texto generado automáticamente, GPTZero ha demostrado ser eficaz en la detección de textos íntegramente creados por IA, alcanzando una precisión de hasta el 99 % en ciertos casos.

Sin embargo, la precisión de esta herramienta disminuye considerablemente cuando se trata de textos híbridos. Los textos editados por humanos presentan patrones de estilo que confunden a los modelos de detección, lo que puede generar una alta tasa de falsos positivos (textos humanos clasificados erróneamente como generados por IA) y falsos negativos (textos generados por IA que pasan como humanos). Esto limita su utilidad en contextos donde las pruebas han sido parcialmente manipuladas o ajustadas (Lau y Zubiaga, 2024). De hecho, autores como Ypma *et al.* (2023) apuntan que la carencia de estándares internacionales en el uso de herramientas de IA para análisis forenses complica la validez y fiabilidad de las pruebas presentadas en tribunales ya que los resultados pueden variar significativamente según las herramientas y metodologías empleadas.

### 5.2. DetectGPT

DetectGPT es otra herramienta desarrollada específicamente para identificar textos generados por modelos basados en transformers. Este sistema utiliza una combinación de análisis estilométrico y análisis probabilístico para identificar patrones textuales que son típicos de los modelos de IA. Al igual que GPTZero, DetectGPT es muy preciso en la detección de textos que no han sido modificados, pero su rendimiento disminuye en textos híbridos o editados por humanos (Weichert y Dimobi, 2024).

Uno de los puntos fuertes de DetectGPT es su capacidad para identificar irregularidades en la distribución léxica y la estructura gramatical de los textos generados por IA. Sin embargo, su precisión varía dependiendo de la longitud del texto y la complejidad del lenguaje utilizado. Textos más cortos y simples suelen ser más difíciles de clasificar con precisión, lo que puede afectar la fiabilidad de los análisis forenses en ciertos casos.

### 5.3. JGAAP (Java Graphical Authorship Attribution Program)

JGAAP es una herramienta de análisis estilométrico diseñada para la atribución de autoría, utilizada en el análisis forense de textos humanos. Aunque no fue específicamente diseñada para la detección de IA, JGAAP puede ser útil para comparar características textuales en grandes corpus de datos y establecer diferencias estilísticas entre autores humanos. Esto puede ser valioso en el contexto de la lingüística forense, especialmente en casos donde se sospecha que un texto ha sido manipulado por diferentes individuos (Juola, 2015).

No obstante, su capacidad para detectar textos generados por IA es limitada, ya que no está optimizada para identificar los patrones específicos generados por los modelos de IA modernos. JGAAP se centra más en los cambios de estilo entre autores humanos, por lo que su uso para analizar textos creados o modificados por IA sigue siendo un desafío.

#### 5.4. Stylometry with R

El paquete Stylometry with R es ampliamente utilizado para el análisis estilométrico en la comparación de textos. Aunque esta herramienta no está diseñada específicamente para la detección de IA, su capacidad para identificar diferencias sutiles en la estructura gramatical, la longitud de las oraciones y el uso del léxico la hace útil para analizar textos humanos en contextos literarios y académicos (Eder *et al.*, 2016).

Sin embargo, al igual que JGAAP, Stylometry with R no ha sido adaptado para el análisis de textos generados por IA. Los modelos de IA, especialmente aquellos basados en transformers, tienden a generar textos con distribuciones de palabras que se alejan de los patrones humanos, pero el sistema no está optimizado para detectar estas diferencias. Como resultado, Stylometry with R es más útil en la atribución de autoría humana que en la detección de IA.

#### 5.5. TAP (Text Analysis Portal)

TAP es otra herramienta utilizada en el análisis de grandes conjuntos de datos textuales. Aunque está orientada a la investigación en lingüística (Gillespie y McKee, 1998) y no específicamente a la detección de IA, TAP permite la comparación de características lingüísticas entre diferentes corpus de texto. Esto puede ser útil para analizar patrones repetitivos o anómalos en textos donde se sospecha que se ha utilizado IA.

Sin embargo, TAP tampoco cuenta con un índice de precisión específico para detectar IA y, por lo tanto, su uso en el análisis forense de textos generados por máquinas sigue siendo limitado.

#### 5.6. Limitaciones generales

Aunque las herramientas disponibles ofrecen ciertas capacidades para identificar textos generados por inteligencia artificial, presentan limitaciones significativas que afectan su fiabilidad en contextos forenses, donde se requiere un alto grado de precisión y consistencia. Uno de los principales desafíos radica en la detección de textos híbridos, aquellos que combinan características generadas por IA con ediciones realizadas por humanos. Estos textos representan un problema para los algoritmos de detección, que están diseñados para identificar patrones completos de IA pero fallan al enfrentarse a este tipo de composiciones, lo que disminuye considerablemente su efectividad.

Además, la variabilidad de los resultados es una preocupación recurrente. Factores como la longitud, la complejidad y el estilo del texto influyen en el desempeño de las herramientas, lo que genera resultados inconsistentes que pueden comprometer la validez del análisis en contextos legales. Esto es particularmente crítico en casos judiciales, donde la fiabilidad de las pruebas lingüísticas es fundamental para la toma de decisiones.

Por último, la evolución constante de la tecnología exacerba estas limitaciones. Los modelos de IA avanzan rápidamente y están siendo optimizados para evitar los patrones que los detectores forenses suelen identificar. Este desarrollo no solo reduce la efectividad de las herramientas existentes, sino que también plantea la necesidad de actualizaciones constantes y de nuevas metodologías para mantener la capacidad de detección frente a estos modelos cada vez más sofisticados. Zellers *et al.* (2019) destacan que los discriminadores actuales pueden clasificar noticias falsas generadas por modelos neuronales

con una precisión del 73%, asumiendo acceso a un nivel moderado de datos de entrenamiento en textos híbridos o editados por humanos. No obstante, su modelo, el modelo Grover, alcanza una precisión del 92% en la detección de noticias falsas generadas por IA. Esto demuestra la importancia de liberar generadores potentes para mejorar las técnicas de verificación.). En este contexto, resulta evidente la necesidad de una innovación continua en las herramientas de análisis forense.

## 6. Retos futuros de la lingüística forense ante la evaluación de la IA

A medida que la inteligencia artificial continúa evolucionando, la lingüística forense se enfrenta a un panorama cada vez más complejo. Los avances en la generación automática de texto y voz, así como en la manipulación digital, plantean nuevos desafíos para los expertos en la atribución de autoría, la autenticación de pruebas y el análisis de comunicaciones en contextos legales. A continuación, se describen algunos de los retos más significativos que la lingüística forense deberá abordar en el futuro.

Uno de los mayores desafíos es la creciente capacidad de los modelos de IA generativa para imitar con precisión el lenguaje humano. A medida que los modelos avanzan, los textos generados por IA se vuelven cada vez más difíciles de distinguir de los escritos por humanos. Esto se debe a que los modelos de lenguaje grande (LLMs) como GPT-3 no solo aprenden la sintaxis del lenguaje, sino también los matices estilísticos y discursivos que caracterizan la escritura humana (Brown *et al.*, 2020), capacidades que han sentado las bases para desarrollos como GPT-4. Este fenómeno difumina los límites entre el texto humano y el generado por IA, lo que complica la tarea de atribuir autoría en contextos forenses. En el futuro, los lingüistas forenses deberán desarrollar nuevas metodologías que les permitan distinguir de manera más precisa entre los dos tipos de texto. Esto podría implicar la creación de modelos híbridos que combinen análisis estilométrico y técnicas de inteligencia artificial para identificar las sutilezas que aún separan el lenguaje humano del generado por máquinas.

Otro reto clave es la posibilidad de que las IAs se optimicen para evadir los detectores forenses. Un desafío clave es la posibilidad de que las IAs se optimicen para evadir los detectores forenses. Como señalan estudios recientes, los ataques adversariales pueden alterar pequeñas porciones de texto, logrando que las herramientas de detección no identifiquen su origen sintético (Wallace *et al.*, 2019).

Por ejemplo, los modelos de IA pueden parafrasear y reorganizar frases, de manera que las características que suelen ser detectadas por los sistemas forenses se vean atenuadas o eliminadas. Esto presenta un reto significativo, ya que los atacantes pueden seguir generando textos fraudulentos con la seguridad de que pasarán las barreras de detección existentes. Los expertos forenses deberán desarrollar algoritmos de detección más robustos que sean capaces de adaptarse y detectar estos ataques adversariales, un área que requerirá colaboración entre lingüistas y expertos en IA.

El avance constante en la investigación de IA ha dado lugar a la creación de múltiples variantes de los modelos de lenguaje grande (LLMs). Cada una de estas variantes presenta desafíos únicos para la detección forense. Los modelos entrenados con datos específicos, como los adaptados para ciertos dialectos o estilos de escritura, pueden generar textos que parecen aún más humanos que los producidos por los modelos generales (Alhafni, 2024). Estas variantes personalizadas permiten a los delincuentes adaptar los textos generados para ser coherentes con el estilo o el contexto particular de una persona, lo que aumenta la dificultad de detectarlos.

En el futuro, los lingüistas forenses deberán mantenerse al día con los desarrollos tecnológicos y ser capaces de analizar una variedad de modelos de IA con diferentes características y capacidades. Esto implica la necesidad de capacitación interdisciplinaria en inteligencia artificial y el desarrollo de nuevas herramientas especializadas que sean capaces de analizar textos generados por una amplia gama de modelos.

La evolución de la IA ha resaltado la necesidad de una colaboración interdisciplinaria entre expertos en lingüística, inteligencia artificial, derecho y ciberseguridad. Los lingüistas forenses deberán trabajar en estrecha colaboración con desarrolladores de IA para comprender el funcionamiento interno de los modelos generativos y sus limitaciones. Esta colaboración permitirá desarrollar herramientas y técnicas más precisas para detectar textos y voces generados por IA.

Además, los profesionales en lingüística forense necesitarán formación especializada en tecnologías de IA para poder realizar análisis más efectivos. Los modelos de IA se están volviendo cada vez más complejos y requieren un conocimiento profundo de su funcionamiento para poder detectar anomalías en el texto o las voces generadas. La formación en procesamiento de lenguaje natural (PLN) y técnicas de aprendizaje automático será fundamental para mantenerse al día con estos avances.

Actualmente, no existe un consenso global sobre los estándares forenses para el análisis de textos generados por IA. A diferencia de la atribución de autoría tradicional, que cuenta con décadas de investigación, la detección de textos creados por IA es un campo emergente que carece de metodologías unificadas. La ausencia de estándares globales dificulta la validez y fiabilidad de los análisis forenses en los tribunales, ya que los resultados pueden variar dependiendo de las herramientas y enfoques utilizados (Dinis-Oliveira y Azevedo, 2023).

En el futuro, será fundamental el desarrollo de marcos normativos y metodológicos que establezcan las mejores prácticas para el análisis de textos generados por IA. Esto incluirá la creación de protocolos replicables que puedan ser aceptados en tribunales y en investigaciones judiciales, asegurando que los análisis sean consistentes y fiables.

Finalmente, uno de los retos más importantes que enfrenta la lingüística forense es el impacto ético y legal del uso de IA en contextos legales. A medida que las pruebas lingüísticas generadas por IA se utilizan en investigaciones y juicios, surge la pregunta de cómo garantizar que estas pruebas sean tratadas de manera justa y ética. La transparencia en el uso de herramientas de IA y el respeto por los derechos de privacidad de los individuos son cuestiones fundamentales que deberán ser abordadas.

Además, la creciente sofisticación de las IA plantea el riesgo de que se utilicen de manera maliciosa para incriminar a personas inocentes mediante la creación de pruebas falsas. En este sentido, es vital que los expertos forenses y los legisladores trabajen juntos para crear marcos legales sólidos que regulen el uso de estas tecnologías de inteligencia artificial. Estos marcos deben garantizar que las pruebas generadas o analizadas mediante IA se utilicen de manera ética y responsable en los sistemas judiciales, protegiendo la integridad del proceso legal (Floridi *et al.*, 2021).

## 7. Conclusiones

La lingüística forense se encuentra en un punto de inflexión debido a los rápidos avances en inteligencia artificial. Las tecnologías de generación automática de texto y voz han desafiado los métodos tradicionales de análisis forense, lo que pone en duda la capacidad de los expertos para atribuir autoría y autenticar pruebas en contextos legales. Los modelos de lenguaje grande (LLMs), como GPT-3 y GPT-4, han alcanzado un nivel de sofisticación que permite generar textos y voces con un realismo tal que es casi indistinguible del lenguaje humano. Esto plantea una serie de desafíos para los lingüistas forenses, quienes deberán adaptar sus técnicas y herramientas para mantenerse al día con los avances tecnológicos.

A lo largo de este artículo, se han explorado los principales retos que enfrenta la lingüística forense en la era de la IA. Se ha demostrado que la difuminación de los límites entre el lenguaje humano y el generado por máquinas, junto con la posibilidad de ataques adversariales y la proliferación de textos híbridos, complica significativamente la tarea de identificar y autenticar textos en un contexto judicial. A medida que los modelos de IA continúan evolucionando, la capacidad de estos sistemas para evadir detectores y generar pruebas falsas se incrementará, lo que subraya la importancia de desarrollar nuevas metodologías y herramientas forenses más robustas.

En el aspecto tecnológico, se ha destacado la necesidad de un enfoque interdisciplinario para abordar estos retos. Los lingüistas forenses deberán colaborar estrechamente con expertos en IA, científicos de datos y juristas para desarrollar soluciones que permitan detectar, analizar y atribuir textos generados por IA con un alto grado de precisión. La formación especializada en procesamiento de lenguaje natural (PLN) y aprendizaje automático será crucial para que los peritos lingüistas comprendan el funcionamiento interno de los modelos generativos y puedan mejorar la fiabilidad de sus análisis.

Desde una perspectiva legal y ética, es imperativo que los legisladores y los expertos forenses trabajen juntos para crear marcos normativos que regulen el uso de la IA en contextos legales. La creciente capacidad de la IA para generar pruebas falsas, como *deepfakes* de voz y textos falsificados, presenta un riesgo significativo para la integridad de los sistemas judiciales. La transparencia y la responsabilidad y uso de estas tecnologías son pilares fundamentales para asegurar que estas herramientas se alineen con principios éticos y promuevan la equidad en sus aplicaciones, incluyendo el ámbito judicial (Floridi *et al.*, 2021). En este sentido, las directrices de la Unesco destacan que la adopción de sistemas de IA en los tribunales debe realizarse de manera que se respeten los derechos humanos y se garantice la justicia y la equidad en los procesos legales. Además, subrayan la importancia de la transparencia en la toma de decisiones basadas en IA para fortalecer la confianza pública en los sistemas judiciales (Unesco, 2024). Es esencial que se desarrollen protocolos replicables para el análisis de textos generados por IA que puedan ser aceptados en los tribunales y otras investigaciones judiciales.

La evolución de la lingüística forense en la era de la inteligencia artificial no es opcional, sino una necesidad urgente. La disciplina debe transformarse para enfrentar los retos que plantea el uso malicioso de la IA y mantenerse relevante en un entorno digital cada vez más complejo. El futuro de la lingüística forense dependerá de su capacidad para adaptarse y para incorporar nuevas tecnologías y enfoques. Solo mediante la integración de la IA con las metodologías tradicionales de análisis forense

y el desarrollo de nuevas técnicas de detección, los lingüistas forenses podrán seguir desempeñando su papel en la administración de justicia en un mundo donde la manipulación digital y la generación de contenido sintético están en constante expansión.

En definitiva, la colaboración interdisciplinaria entre expertos en lingüística, inteligencia artificial, derecho y ciberseguridad se perfila como la clave para enfrentar estos desafíos. Aunque las respuestas aún están en construcción, es necesario avanzar rápidamente en el desarrollo de soluciones tecnológicas y marcos legales que aseguren que los sistemas judiciales puedan adaptarse y mantenerse firmes frente a los cambios que trae consigo la inteligencia artificial. La capacidad de la lingüística forense para evolucionar y adaptarse será crucial para garantizar que la justicia y la verdad prevalezcan en un entorno donde las tecnologías emergentes están redefiniendo las reglas del juego.

## Referencias

- Alhafni, B., Kulkarni, V., Kumar, D., y Raheja, V. (2024). Personalized Text Generation with Fine-Grained Linguistic Control. *arXiv*. <https://doi.org/10.48550/arXiv.2402.04914>
- ASIS International. (2024). *Putting Generative AI to Use for Crime: Fraud, Disinformation, Exploitation, and More*. ASIS. <https://www.asisonline.org>
- Brennan, M., Afroz, S., y Greenstadt, R. (2012). Adversarial stylometry: Circumventing authorship recognition to preserve privacy and anonymity. *ACM Transactions on Information and System Security (TISSEC)*, 15(3), 1-22. <https://doi.org/10.1145/2382448.2382450>
- Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D., Wu, Z., Winter, C., y Amodei, D. (2020). Language models are few-shot learners. *arXiv*. <https://doi.org/10.48550/arXiv.2005.14165>
- Chakraborty, S., Bedi, A. S., Zhu, S., An, B., Manocha, D., y Huang, F. (2023). On the possibilities of ai-generated text detection. *arXiv*. <https://doi.org/10.48550/arXiv.2304.04736>
- Coulthard, M., y Johnson, A. (2007). *An Introduction to Forensic Linguistics: Language in Evidence*. Cambridge University Press.
- Dinis-Oliveira, R. J., y Azevedo, R. M. (2023). ChatGPT in forensic sciences: a new Pandora's box with advantages and challenges to pay attention. *Forensic Sciences Research*, 8(4), 275-279. <https://doi.org/10.1093/fsr/owad039>
- Departamento de Cyber Threat Intelligence de NTT DATA Cybersecurity. (2024). *Informe de tendencias y ciberamenazas del 1er semestre de 2024*. NTT DATA. <https://es.nttdata.com/newsfolder/los-ciberataques-supondran-un-coste-de-10000-millones-en-2024>
- Departamento de Justicia de EE.UU. (2024, 9 de julio). *Justice Department Leads Efforts Among Federal, International, and Private Sector Partners*. United States Attorney's Office. <https://www.justice.gov/usao-ndil/pr/justice-department-leads-efforts-among-federal-international-and-private-sector>
- Eder, M., Rybicki, J., y Kestemont, M. (2016). Stylometry with R: A package for computational text analysis. *R Journal*, 8(1), 107-121. <https://journal.r-project.org/archive/2016/RJ-2016-007/index.html>
- Europol. (2024). *New report finds that criminals leverage AI for malicious use – and it's not just deep fakes*. <https://www.europol.europa.eu/media-press/newsroom/news/new-report-finds-criminals-leverage-ai-for-malicious-use-%E2%80%93-and-it%E2%80%99s-not-just-deep-fakes>
- FBI. (2024, 8 de mayo). *FBI Warns of Increasing Threat of Cyber Criminals Utilizing Artificial Intelligence*. <https://www.fbi.gov/contact-us/field-offices/sanfrancisco/news/fbi-warns-of-increasing-threat-of-cyber-criminals-utilizing-artificial-intelligence>
- Floridi, L., Cows, J., King, T. C., y Taddeo, M. (2021). How to design AI for social good: Seven essential factors. *Ethics, Governance, and Policies in Artificial Intelligence*, 26, 125-151. <https://link.springer.com/article/10.1007/s11948-020-00213-5>

- Fobbe, E. (2020). Text-linguistic analysis in forensic authorship attribution. *Journal of Language and Law (JLL)*, 9, 111–123. <https://www.languageandlaw.eu/jll/article/view/78>
- Franganillo, J. (2023). La inteligencia artificial generativa y su impacto en la creación de contenidos mediáticos. *Metodos, revista de ciencias sociales*, 11(2), 1-17. <https://doi.org/10.17502/mrcs.v11i2.710>
- Federal Trade Commission. (2024, 25 de septiembre). FTC Announces Crackdown on Deceptive AI Claims and Schemes Federal Trade Commission. Recuperado de <https://www.ftc.gov/news-events/news/press-releases/2024/09/ftc-announces-crackdown-deceptive-ai-claims-schemes>
- Gehrmann, S., Strobel, H., y Rush, A. M. (2019). GLTR: Statistical Detection and Visualization of Generated Text. *arXiv*. <https://doi.org/10.48550/arXiv.1906.04043>
- Gillespie, J. H., y McKee, J. (1998). The Text Analysis Program: Developing students' analytical skills. *ReCALL*, 10(2), 44-53. <https://www.cambridge.org/core/journals/recall/article/abs/text-analysis-program-developing-students-analytical-skills/047569D72795ADEBA82B9D2D97BEC457>
- Giulianelli, M., Baan, J., Aziz, W., Fernández, R., y Plank, B. (2023). What comes next? evaluating uncertainty in neural text generators against human production variability. *arXiv*. <https://doi.org/10.48550/arXiv.2305.11707>
- Gómez, M. J. (2023). “Mamá, ayúdame”: *Inteligencia Artificial imita voz de adolescente para fingir secuestro y exigir rescate*. Recuperado de <https://www.latercera.com/tendencias/noticia/mama-ayudame-inteligencia-artificial-imita-voz-de-adolescente-para-fingir-secuestro-y-exigir-rescate/XTQEUXTJZJH5BJJXHRJX7QID4E/>
- Grant, T. (2022). *The idea of progress in forensic authorship analysis*. Cambridge University Press.
- Hadnagy, C., y Fincher, M. (2015). *Phishing Dark Waters: The Offensive and Defensive Sides of Malicious Emails*. John Wiley & Sons.
- Ji, J., Li, R., Li, S., Guo, J., Qiu, W., Huang, Z., Chen, C., Xinru, J., y Lu, X. (2024). Detecting machine-generated texts: Not just “ai vs humans” and explainability is complicated. *arXiv*. <https://doi.org/10.48550/arXiv.2406.18259>
- Juola, P. (2015). The Rowling case: a proposed standard analytic protocol for authorship questions. *Digital Scholarship in the Humanities*, 30, 100-113. <https://doi.org/10.1093/llc/fqv040>
- Lau, H. T., y Zubiaga, A. (2024). Understanding the Effects of Human-written Paraphrases in LLM-generated Text Detection. *arXiv*.
- Maras, M. H. y Alexandrou, A. (2018). Determining authenticity of video evidence in the age of artificial intelligence and in the wake of Deepfake videos. *The International Journal of Evidence & Proof*, 23, 255-262. <https://doi.org/10.1177/1365712718807226>
- Mindner, L., Schlippe, T., y Schaaff, K. (2023). Classification of human-and ai-generated texts: Investigating features for chatgpt. En *International Conference on Artificial Intelligence in Education Technology* (pp. 152-170). Springer Nature Singapore. <https://doi.org/10.48550/arXiv.2308.05341>

- Muñoz-Basols, J., y Fuertes Gutiérrez, (2024). Oportunidades de la Inteligencia Artificial (IA) en la enseñanza y el aprendizaje de lenguas. En J. Muñoz-Basols, M. Fuertes Gutiérrez, y L. Cerezo (Eds.), *La enseñanza del español mediada por tecnología: De la justicia social a la Inteligencia Artificial (IA)* (pp. 343-360). Routledge.
- Queralt, S. (2023). Los aportes de la lingüística forense contra el cibercrimen. *Del Español. Revista de Lengua*, 1, 259-271. <https://doi.org/10.33776/dlesp.v1.7923>
- Perkins, R. (2021). The Application of Forensic Linguistics in Cybercrime Investigations. *Policing: A Journal of Policy and Practice*, 15, 168–78. <https://doi.org/10.1093/police/pay097>
- Sousa-Silva, R. (2022). Fighting the Fake: A Forensic Linguistic Analysis to Fake News Detection. *International Journal for the Semiotics of Law - Revue internationale de Sémiotique juridique*, 35, 2409–2433. <https://doi.org/10.1007/s11196-022-09901-w>
- Sophos. (2019, 5 de septiembre). Scammers deepfake CEO's voice to talk underling into \$243000 transfer. *Sophos News*. [https://news.sophos.com/en-us/2019/09/05/scammers-deepfake-ceos-voice-to-talk-underling-into-243000-transfer/?utm\\_source=chatgpt.com](https://news.sophos.com/en-us/2019/09/05/scammers-deepfake-ceos-voice-to-talk-underling-into-243000-transfer/?utm_source=chatgpt.com)
- Unesco. (2024). *Directrices de la UNESCO para el uso de sistemas de inteligencia artificial en juzgados y tribunales*. Unesdoc. Recuperado de [https://unesdoc.unesco.org/ark:/48223/pf0000390781\\_spa](https://unesdoc.unesco.org/ark:/48223/pf0000390781_spa)
- Wallace, E., Feng, S., Kandpal, N., Gardner, M., y Singh, S. (2019). Universal adversarial triggers for attacking and analyzing NLP. *arXiv*. <https://doi.org/10.48550/arXiv.1908.07125>
- Weichert, J., y Dimobi, C. (2024). DUPE: Detection Undermining via Prompt Engineering for Deepfake Text. *arXiv*. <https://doi.org/10.48550/arXiv.2404.11408>
- Ypma, R. J. F., Ramos, D., y Meuwly, D. (2023). AI-based Forensic Evaluation in Court: The Desirability of Explanation and the Necessity of Validation. En Z. Geradts y K. Franke (Eds.), *Artificial Intelligence (AI) in Forensic Sciences*. Wiley.
- Zellers, R., Holtzman, A., Rashkin, H., Bisk, Y., Farhadi, A., Roesner, F., y Choi, Y. (2019). Defending against neural fake news. *arXiv*. <https://doi.org/10.48550/arXiv.1905.12616>
- Zeng, Z., Liu, S., Sha, L., Li, Z., Yang, K., Liu, S., Gašević, D. y Chen, G. (2024). Detecting AI-Generated Sentences in Human-AI Collaborative Hybrid Texts: Challenges, Strategies, and Insights. *arXiv*. <https://doi.org/10.48550/arXiv.2403.03506>

### Contribución de la autora

Sheila Queralt ha participado en la elaboración, el recojo de datos, el diseño, la redacción del artículo y da aprobación a la versión que se publica en la revista”.

### Financiamiento

Sin financiamiento

## Conflicto de intereses

La autora no presenta conflicto de interés

**Correspondencia:** [sheila.queralt@cllicenciats.cat](mailto:sheila.queralt@cllicenciats.cat)

## Trayectoria académica de la autora

**Sheila Queralt** es doctora en Traducción y Ciencias del Lenguaje, máster en Lingüística Forense, en Policía Científica e Inteligencia Criminal, en Criminalística, en Ciencias del Grafismo, en Lingüística Teórica y Aplicada, y en Estadística Aplicada a la Investigación. Es licenciada en Lingüística y en Traducción e Interpretación. Es directora del Laboratorio SQ-Lingüistas Forenses, docente en distintas universidades y colabora como experta lingüista con diferentes cuerpos policiales nacionales e internacionales en casos de corrupción, ciberseguridad, narcotráfico, homicidios, terrorismo, entre otros. Es autora de varios libros entre los que destacan *Fundamentos de la Lingüística Forense* (2019), *Atrapados por la lengua* (2020, 3.<sup>a</sup> ed. 2024), *Estafas amorosas* (2022) y *Lingüistas de Hoy* (2023).