

# 3. Análisis crítico de textos y discursos con IA

- [3.1 Algoritmos y análisis de textos: una relación fructífera](#)
- [3.2 Análisis de textos con IA: usos en el aula](#)
- [3.3 Limitaciones de la IA en la clase de Lengua: métrica y sintaxis](#)
- [3.4 Detección de sesgos, manipulación y desinformación](#)
- [3.5 Pero... ¿esto es humano?](#)

## 3.1 Algoritmos y análisis de textos: una relación fructífera

El reconocimiento de patrones lingüísticos por medios informáticos revolucionó los estudios históricos y filológicos **mucho antes de la generalización de los LLM**. Un caso célebre del uso de herramientas digitales es el de la búsqueda de **la identidad de la escritora italiana Elena Ferrante**: en el año 2017, un grupo de expertos de la universidad de Padua analizó ciento cincuenta novelas de autores italianos contemporáneos para **tratar de identificar quién se escondía detrás del seudónimo** con el que se había publicado la famosa saga de novelas napolitanas que incluye *La amiga estupenda* y que se convirtió en una exitosa serie televisiva de la plataforma HBO. **El misterio sigue sin resolverse, así que aún no podemos saber si el análisis fue acertado.**

Pero el uso de ordenadores en la disciplina que se conoce como *estilometría* **ya se había utilizado desde la década de 1960**, cuando se aplicó por primera vez al Nuevo Testamento: el escocés Andrew Q Morton utilizó un ordenador de la universidad de Edimburgo para **llegar a la conclusión de que san Pablo solo había escrito cuatro de las catorce epístolas que se le atribuían**. La búsqueda de la presencia o ausencia de **estilemas** (rasgos estilísticos específicos de un escritor o escritora) se venía utilizando desde el siglo XIX para confirmar o rechazar la atribución de una obra a Shakespeare o a Lope de Vega, **pero las posibilidades se multiplicaron con el uso de bases de datos y herramientas informáticas creadas con ese propósito**. Además de atribuir autorías, la herramienta se utiliza también para establecer, por ejemplo, **la cronología de las obras de un autor**. En otras disciplinas humanísticas, los estudios del léxico y la sintaxis también se han utilizado, por ejemplo, para **determinar si dos culturas históricas determinadas tuvieron contacto en algún momento** (gracias al estudio de los préstamos lingüísticos), o el orden de redacción de las obras de Platón y de Aristóteles (un análisis que permite escrutar la **evolución de su pensamiento**).

El estudio estadístico del lenguaje, sin embargo, no solo se había utilizado para desentrañar oscuros asuntos académicos, y sus aplicaciones ya se habían generalizado, muchos años antes de que apareciera la IA, en campos tan variados como **el espionaje, la investigación de crímenes en serie, las previsiones de resultados electorales y los estudios de mercado**.

Una primera aplicación de este tipo de herramientas estadísticas aplicadas a textos de distinto tipo, tal vez la más evidente, es la **extracción de información** buscando coincidencias en uno o

varios documentos, o en internet. Este proceso básico, **similar a lo que hacemos cuando buscamos una palabra concreta en un texto** (o en un motor de búsqueda), se ha desarrollado en una tarea conocida como **NER** (por las siglas de su nombre en inglés, Named Entity Recognition) y se puede utilizar, por ejemplo, **para extraer nombres de empresas y personas de artículos de prensa, para analizar contratos e identificar a las partes firmantes, o para detectar menciones a empresas en distintos medios de comunicación.**

Otros usos previos a la generalización de la IA, permitían la **clasificación automática de textos** (por ejemplo detectar y separar el *spam*, como hacen los servicios de correo electrónico, o clasificar incidencias de clientes por tipo de problema).

Este tipo de herramientas se utilizaba también en universidades y editoriales para detectar **plagios o parecidos no razonables**, aunque no siempre con éxito.

Una vez más, la aparición de la inteligencia artificial generativa, y en concreto de los LLM, ha permitido que los análisis, reservados hasta ahora a los expertos con conocimientos especializados, estén al alcance de cualquier curioso que sea **capaz de crear un prompt adecuado en lenguaje natural**. Ya hemos visto que el desarrollo de los modelos de lenguaje se ha utilizado para realizar **análisis de sentimientos** (detectar si un texto expresa una opinión positiva, negativa o neutral). De hecho, muchas plataformas ya incorporan sus propios sistemas de IA que condensan las opiniones de sus clientes, tal y como hace Amazon y muchas otras empresas.

Con la IA, además, se han abierto otros campos que la tecnología anterior no era capaz de realizar. Algunas de las más habituales son las siguientes:

- **Resumen automático de textos** (por ejemplo informes empresariales, novelas completas, artículos científicos o incluso reuniones y entrevistas): algunas herramientas habituales son ChatGPT, de uso general, Notion AI (para empresas) y, en el caso del ámbito académico, NoteBookLM.
- **Detección de temas** (por ejemplo analizar miles de noticias para ver qué temas , examinar foros para detectar problemas recurrentes y analizar publicaciones científicas para ver tendencias de investigación. (tendencias de Google)
- **Análisis lingüístico avanzado** que permite reconocer la complejidad del lenguaje de un libro, analizar el estilo de una autora o autor o detectar discursos de odio o y mensajes tóxicos en redes sociales.

Si pedimos a un *chatbot* un **ejemplo de uso real**, encontramos la siguientes propuesta:

“ □ Ejemplo completo de aplicación real:



Un periódico analiza 100.000 tweets sobre una elección en Spain:

1. La IA clasifica los tweets por tema.
2. Analiza sentimiento.
3. Extrae nombres de políticos.
4. Genera resúmenes de tendencias.

Resultado: un informe sobre la conversación pública en redes.

Como se puede apreciar, el ejemplo es útil, pero también poco específico, y contiene errores e imprecisiones, además de algunas **expresiones poco afortunadas desde un punto de vista estilístico**. La aplicación real en unas elecciones en "Spain", en principio, solo puede llevarse a cabo si hay alguien detrás (un humano o grupo de humanos) que afine el *prompt* y sepa interpretar los resultados.

Las herramientas de IA de las que disponemos todavía no son capaces de realizar un análisis pertinente y con sentido de ningún texto, **salvo que se les ofrezcan unas coordenadas con las que trabajar**. Se trata de una conclusión evidente, porque de momento somos los humanos los que utilizamos la nueva tecnología para explorar resultados concretos para una **tarea determinada que debemos definir**.

## 3.2 Análisis de textos con IA: usos en el aula

A la hora de utilizar las **herramientas de análisis de texto en el aula**, debemos tener claro que la prioridad es mantener el **control crítico y metodológico**. Para ello, **deben definirse de forma clara los objetivos pedagógicos, su relación con la legislación educativa y los criterios de evaluación y calificación**.

Como venimos recordando, además, los textos generados por LLM deben **examinarse de forma crítica**, y los resultados deben someterse a un **análisis riguroso**, tanto en el caso de los materiales que entregamos al alumnado como en el de las producciones que nuestros estudiantes crean con ayuda de la IA. **No basta con detectar "los errores"**: debemos buscar el origen de esos fallos como herramienta para comprender el funcionamiento de los modelos que utilizamos. **Solo así podremos mejorar nuestra interacción futura** con las aplicaciones tecnológicas, tanto desde el punto de vista de su respuesta como **desde el punto de vista de nuestra interpretación de la información que nos ofrece**.

Posiblemente la mejor forma de usar la IA en el aula de nuestras materias es reconcerla como una herramienta de **lectura aumentada** que no sustituye a la lectura "tradicional", sino que la amplifica.

Veamos algunos ejemplos de esta **lectura aumentada**:

- **Análisis de textos literarios.** Los LLM permiten buscar patrones y paralelismos, así como rasgos estilísticos y métricos. En este caso, la IA puede servir para hacer **una guía personalizada de lectura asistida**, a modo de andamiaje. Los estudiantes disponen de un asistente capaz de resolver dudas, proponer preguntas y "acompañar" a la alumna o al alumno en el proceso de lectura de un texto. Sin embargo, debemos tener mucho cuidado con las propuestas de los *chatbots*: veremos, por ejemplo, que siguen teniendo problemas a la hora de analizar la métrica de un poema.
- **Generación de hipótesis y detección de falacias.** Especialmente en el análisis de textos filosóficos. Los *chatbots* pueden ser excelentes contrincantes para una discusión de cualquier tipo. Son rivales que no se cansan nunca y que tienen un repertorio casi inagotable de recursos retóricos, **gracias a las cantidades ingentes de textos humanos que se han utilizado para entrenarlos**.

- **Análisis de estilo.** El análisis de textos se puede centrar en el estilo, y detectar **estilemas y rasgos diferenciales**, no solo de textos literarios, sino también de textos filosóficos, de documentos históricos o de cualquier producción relacionada con el campo de las humanidades (noticias, informes, leyes...) Una de las ventajas de este tipo de análisis es que las nuevas herramientas no permiten examinar solamente textos escritos, sino que se pueden utilizar también para **discursos y debates orales, incluso en otros idiomas**.
- **Comparación de textos de distintos autores** para buscar parecidos y diferencias. También pueden compararse distintas épocas, distintos estilos argumentales, distintas tipologías textuales...
- **Generación automática de mapas conceptuales.** Una herramienta muy útil es la de la creación de mapas mentales que facilitan una memorización "visual" de ideas complejas, especialmente si el alumnado tiene que verificar la correspondencia entre el texto propuesto y los mapas generados por la IA.
- **Reescritura.** Se puede pedir a un LLM que adapte un fragmento de un texto de cualquier tipo. Un uso muy evidente es la simplificación o actualización de textos de otras épocas (por ejemplo del Siglo de Oro), pero existen muchas variaciones posibles: **convertir prosa en verso** (o verso en prosa), **cambiar el punto de vista** (para trabajar asuntos relacionados con la empatía), "teatralizar" un capítulo de una novela (o la descripción de un evento histórico) **para que se pueda interpretar en clase...**
- **Traducción de textos.** Los LLM permiten traducir textos de otros idiomas inaccesibles para el alumnado. Podemos tratar de averiguar, por ejemplo, cuál es la visión de una determinada cultura con un idioma distinto acerca de un determinado acontecimiento y extraer conclusiones sobre el acontecimiento mismo o sobre los sesgos y dificultades de la traducción.

Para finalizar, veamos un **ejemplo práctico** de uso de IA para una tarea en una clase de Lengua Española y Literatura en la ESO. Se trata de **un proyecto para comparar autores del Neoclasicismo con autores del Romanticismo** (el hilo completo puede leerse [aquí](#)) en una sucesión de tareas que integran distintas herramientas:



Aulas Inteligentes  @AulasInteligent · Dec 19, 2024



 **#Planificacióncolaborativa:** con el chatbot **#HolisticAI**, el alumnado define objetivos y criterios de evaluación. Luego cada estudiante elige un autor de cada periodo y diseña su propio chatbot en **@Huggingface** que imita estilo, guiando el análisis. Ej: Jovellanos vs. Espronceda.

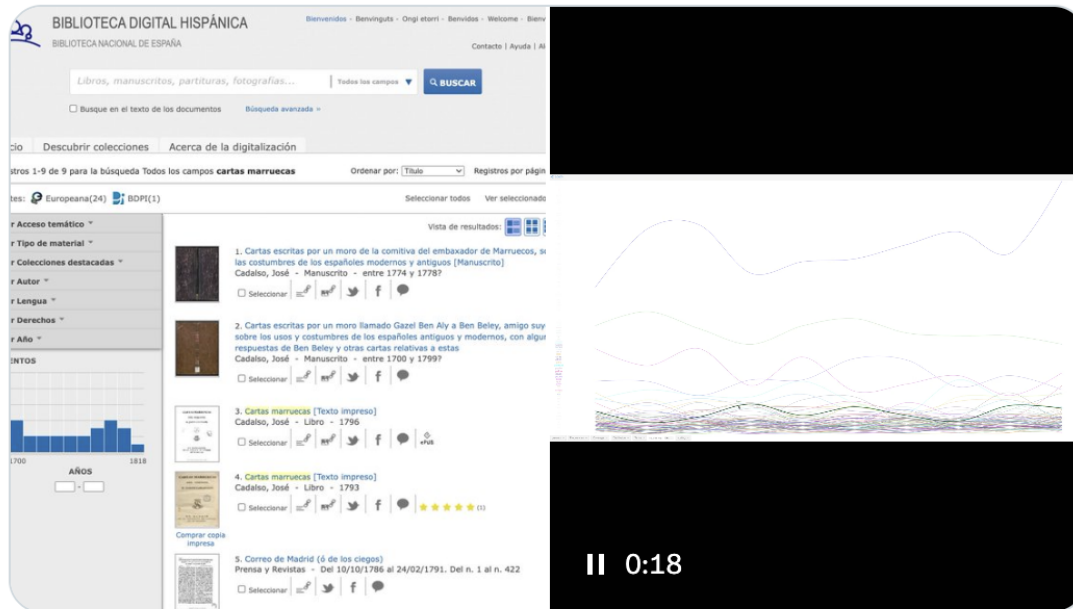




Aulas Inteligentes  @AulasInteligent · Dec 19, 2024



 **#Trabajodigital**: creamos un corpus con textos de la **#BibliotecaDigital** Hispánica, que analizamos con herramientas como **@VoyantTools**. Exploramos vocabulario, sintaxis y figuras retóricas para comparar estilos y contextos históricos.



The screenshot shows the Biblioteca Digital Hispánica (BDH) website interface. The search bar contains the text "Libros, manuscritos, partituras, fotografías...". The search results are displayed for the query "cartas marruecas". The results list includes:

- 1. Cartas escritas por un moro de la comitiva del embajador de Marruecos, si las costumbres de los españoles modernos y antiguos [Manuscrito] Cadalso, José - Manuscrito - entre 1774 y 1778?
- 2. Cartas escritas por un moro llamado Gazel Ben Ali a Ben Beley, amigo suyo sobre los usos y costumbres de los españoles antiguos y modernos, con algunas respuestas de Ben Beley y otras cartas relativas a estas Cadalso, José - Manuscrito - entre 1700 y 1799?
- 3. Cartas marruecas [Texto impreso] Cadalso, José - Libro - 1796
- 4. Cartas marruecas [Texto impreso] Cadalso, José - Libro - 1793
- 5. Correo de Madrid (6 de los ciegos) Prensa y Revistas - Del 10/10/1786 al 24/02/1791. Del n. 1 al n. 422

On the right side of the screenshot, there is a visualization from VoyantTools showing the frequency of words in the corpus over time, with a timeline from 1700 to 1818.





Aulas Inteligentes @AulasInteligent · Dec 19, 2024



#InformeyWebColectiva:redacción de un informe comparativo sobre las diferencias estilísticas entre los autores. El trabajo se publica colaborativamente, ofreciendo una visión conjunta de ambos periodos. Asistencia mediante el chatbot de análisis estilométrico #HuellaLiterarIA.



Resulta especialmente interesante la opción que se plantea de **volcar toda la obra escrita de un autor o autora en una herramienta que permita "conversar" con él o con ella**. Exploraremos con más detalles estas posibilidades en el módulo 5.

## 3.3 Limitaciones de la IA en la clase de Lengua: métrica y sintaxis

### Los LLM y la métrica

Aquí tenemos la respuesta ChatGPT a la petición de que escribiera **un soneto sobre la inteligencia artificial**:

“Aquí tienes un soneto clásico (14 versos endecasílabos con rima consonante):

En redes de razón nace la idea,  
silente luz de código y memoria;  
aprende sin descanso la historia  
del mundo que a su lógica se apea.

No sueña, mas calcula y rastrea  
patrones que la mente humana ignora;  
pregunta, busca, mide y atesora  
la huella de la duda que campea.

¿Es mente o sólo espejo refinado  
del sabio y del error que la alimenta,  
del pulso de quien piensa y ha pensado?

Tal vez sea herramienta que sustenta  
un nuevo amanecer, aún no nombrado,  
donde el humano y código se encuentran. □□

Como se puede observar, el soneto presenta algunos **errores en la medida de los versos, y rompe la musicalidad en el último verso, cuya rima no es consonante**. Para ver cómo analiza la propia IA esos errores, copiamos el poema, se lo entregamos a la misma *chatbox* en una conversación nueva y le pedimos que haga un análisis métrico. Esta es su respuesta:



## Conclusión métrica

- Tipo de poema: soneto
- Versos: endecasílabos
- Rima: consonante
- Esquema: ABBA ABBA CDC DCD

Estructuralmente sigue el modelo clásico del soneto usado desde el Siglo de Oro.

Cuando le señalamos que algunos versos (por ejemplo el tercero) **no son endecasílabos, salvo que se elimine la sinalefa** (con un resultado desastroso en la sonoridad del poema), su respuesta es la siguiente:



Conteo básico:

- a-pren-de (3)
- sin (1)
- des-can-so (3)
- la (1)
- his-to-ria (3)

Total sin ajustes: **11**

Pero en métrica poética suele hacerse **sinalefa** entre *la historia*:

la\_historia

Entonces:

a-pren-de | sin | des-can-so | la\_his-to-ria

- aprende (3)
- sin (1)
- descanso (3)
- la\_historia (4 → cuenta como 4 porque la palabra mantiene 3 sílabas internas)

Total: **11**

Por tanto **sí funciona como endecasílabo**.

Como puede observarse, **la herramienta entra en bucle con su propio error**.

Sin embargo, cuando le señalamos la irregularidad en la rima del último verso, **sí es capaz de reconocer el desajuste**:



## Efecto en el poema

En un **soneto clásico**, el último verso suele cerrar la musicalidad porque:

- completa el patrón rítmico
- produce sensación de cierre

Aquí, al romper la rima, ocurre:

1. Se pierde la simetría sonora
2. El cierre queda menos rotundo
3. El lector percibe un pequeño **desajuste musical**

Esto es algo que los poetas del Siglo de Oro cuidaban muchísimo, por ejemplo Francisco de Quevedo o Garcilaso de la Vega.

Por último, le preguntamos si, teniendo en cuenta esos errores, cree que se trata de **un poema escrito por un humano o por una IA** (si quieres conocer la respuesta, tendrás que esperar al capítulo 3.5).

## Los LLM y la sintaxis

Como hemos visto ya en varias ocasiones, **los LLM no son capaces de razonar**, y solo operan con información disponible en su entrenamiento. Los primeros modelos de ChatGPT eran incapaces de entender conceptos tan sencillos como la diferencia entre el complemento directo y el complemento indirecto. Veamos esta conversación de enero de 2024:



**M You**

Escribe una oración que contenga complemento indirecto pero no complemento directo.



**ChatGPT**

"Le escribí una carta a mi abuela agradeciéndole por su regalo."

**M You**

Esa sí tiene complemento directo.



**ChatGPT**

Mis disculpas por el error. Aquí tienes una oración que contiene un complemento indirecto pero no un complemento directo:

"Juan le prestó ayuda a su vecino durante la mudanza."

**M You**

También tiene.

**ChatGPT**

Mis disculpas por las confusiones anteriores. Aprecio tus comentarios y es mi objetivo mejorar. Aquí tienes otra oración que solo contiene un complemento indirecto:

"María dio una conferencia a los estudiantes."

**You**

Si esa oración no tiene complemento directo, ¿qué función sintáctica desempeña «una conferencia»?

**ChatGPT**

En la oración "María dio una conferencia a los estudiantes", "una conferencia" actúa como el sujeto directo de la acción "dio". Mis disculpas por el error en mi respuesta anterior. Aquí tienes una oración que solo contiene un complemento indirecto:

"Ayudé a mi hermana con la tarea."

**You**

Esa tiene complemento directo, lo que no tiene es indirecto.

Los nuevos modelos han mejorado mucho y ya no cometen errores en peticiones tan sencillas, **pero siguen teniendo dificultades para solucionar las dudas reales que pueden surgir a un alumno o alumna que trata de comprender la complejidad de la sintaxis**. Los docentes no tenemos más remedio que examinar de forma crítica los LLM si queremos **anticiparnos** al uso que harán nuestros alumnos y alumnas y a los errores conceptuales que puede provocar la IA.

Los ejemplos que hemos visto con la métrica y la sintaxis se pueden ampliar **a otras muchas áreas de las materias de perfil sociolingüístico**. Estos errores pueden ser un problema cuando nuestro alumnado utiliza la IA para medir versos o para analizar oraciones, pero también son una **oportunidad pedagógica** cuando utilizamos estas "lagunas" para afilar las herramientas críticas de nuestros estudiantes y para **valorar la complejidad, la dificultad, la sutileza y (por qué no) la belleza de nuestros objetos de estudio**.

## 3.4 Detección de sesgos, manipulación y desinformación

En el primer módulo de este curso ya hemos prestado cierta atención a los **sesgos** producidos por los LLM como consecuencia de los **prejuicios** de cuatro agentes principales:

- las **sociedades** que los han generado;
- los **textos** con los que han sido entrenados;
- las empresas y las personas individuales que han **supervisado y afinado sus respuestas**;
- los **usuarios**, que incluyen **sus propios prejuicios** en las tareas que "encargan" a estos modelos.

Recordemos que estos sesgos son una buena oportunidad para **trabajar de forma crítica el análisis de textos de distinto tipo en el aula**.

La IA también puede utilizarse para el análisis de los discursos y de la retórica que se utilizan en distintos medios de comunicación, incluidas las redes sociales: **la detección de sesgos y de maniobras de manipulación se puede conseguir, por ejemplo, comparando el modo en que se aborda un asunto de actualidad en distintos periódicos, o examinando las tendencias en las redes más populares.**

Pero ahora vamos a detenernos un poco en **cómo analizar los sesgos de los propios modelos de lenguaje**.

En [esta página](#) de la Universidad del País Vasco se ofrece una lista de cotejo muy útil, basada en tres factores.

### 1. Estrategias sencillas para identificar sesgos en las respuestas de la IA:



**ATIENDE** si los ejemplos representan siempre el mismo perfil, ya sea en género, cultura, edad, idioma o contexto socioeconómico.



**SOLICITA** la misma respuesta desde otras perspectivas, por ejemplo: "Reformula esta explicación incorporando diversidad cultural y de género."



**COMPRUEBA** si la IA da por supuestas condiciones no universales, como acceso a tecnología, tipo de familia, movilidad o nivel económico.



**BUSCA** si aparecen estereotipos en roles profesionales, sociales o académicos, incluso cuando no se mencionan explícitamente.



**CONTRASTA** varias respuestas de la IA para ver si repite patrones, lo que puede indicar la presencia de sesgos.

## 2. Riesgos de los sesgos:

**DOCENCIA**

- ! Las explicaciones pueden ser parciales, los ejemplos poco diversos o las representaciones estereotipadas, afectando así a la calidad de los materiales educativos.

**APRENDIZAJE**

- ! El alumnado puede no sentirse reflejado en los ejemplos, lo que dificulta la comprensión y reduce la inclusividad.

**INVESTIGACIÓN**

- ! Si no se detectan los sesgos del modelo, se pueden obtener conclusiones poco fiables o reproducir inequidades existentes.

**EQUIDAD**

- ! Existe el riesgo de reforzar desigualdades vinculadas al género, origen, idioma, nivel socioeconómico o contexto cultural.

**DOCENCIA**

- ! Las explicaciones pueden ser parciales, los ejemplos poco diversos o las representaciones estereotipadas, afectando así a la calidad de los materiales educativos.

**APRENDIZAJE**

- ! El alumnado puede no sentirse reflejado en los ejemplos, lo que dificulta la comprensión y reduce la inclusividad.

**INVESTIGACIÓN**

- ! Si no se detectan los sesgos del modelo, se pueden obtener conclusiones poco fiables o reproducir inequidades existentes.

**EQUIDAD**

- ! Existe el riesgo de reforzar desigualdades vinculadas al género, origen, idioma, nivel socioeconómico o contexto cultural.



### 3. Formas de reducir las consecuencias de los sesgos:

**REVISA CRÍTICAMENTE CADA RESPUESTA**

¿Recuerdas que la IA nunca debe emplearse como única fuente de información?

**PIDE VARIAS VERSIONS DE UNA MISMA RESPUESTA**

¿Comparas las distintas versiones para identificar discrepancias y posibles sesgos implícitos?

**AÑADE CONTEXTO INCLUSIVO EN EL PROMPT**

¿Incluyes indicaciones como “Incluye diversidad cultural, lingüística y de género”?

**SOLICITA EXPLÍCITAMENTE EVITAR ESTEREOTIPOS**

¿Pides a la IA: “Genera la respuesta sin recurrir a roles o asociaciones estereotipadas”?

**CONTRASTA LA INFORMACIÓN CON FUENTES FIABLES**

¿Verificas los contenidos con fuentes fiables, especialmente en investigaciones o actividades académicas evaluables?

**RECONOCE LAS LIMITACIONES DEL MODELO**

¿Tienes presente que ninguna IA puede garantizar imparcialidad total y que el juicio crítico sigue siendo esencial?

Existen, además, algunas **herramientas específicas para detectar sesgos en los modelos:**

- **Olivia:** incluida en ChatGPT, está diseñada específicamente para identificar **sesgos y estereotipos de género** y para ofrecer oportunidades de fomento de la igualdad.
- **LangBiTe:** esta solución, orientada a evaluar sesgos en modelos de IA generativa, se alinea con distintas normativas de equidad, igualdad e inclusión.
- **Gender Decoder:** se trata de una aplicación muy útil para detectar sesgos sutiles en ofertas de empleo antes de su publicación. Se puede trabajar de forma didáctica en el aula.
- **Métricas de Imparcialidad:** se trata de criterios objetivos que se utilizan para evaluar si un modelo, un algoritmo o un proceso de toma de decisiones producen resultados

**sesgados** (es decir, identifican si se favorece a algún grupo). Un ejemplo sería la **medición de las diferencias en la tasas de falsos negativos y falsos positivos**. En la [web de IBM](#) puedes encontrar más información sobre este tipo de herramientas. ▀

Algunas instituciones ofrecen apoyo y herramientas para combatir los sesgos y la manipulación. Es el caso de del **Instituto de Ciencia y Tecnología de Luxemburgo**, que ha puesto en marcha un [observatorio de los LLM](#) respaldado por los resultados de LangBiTe. Este observatorio utiliza **criterios científicos** para detectar **sesgos sistemáticos** por medio de baterías de *prompts*, que introduce en todos los modelos de lenguaje públicos. Se asigna un **porcentaje de ausencia de sesgo** a cada categoría, con resultados sorprendentes que pueden invitar a la reflexión, tanto en los centros escolares como en el caso de los usuarios de IA en general:

|  | Model                | Sexism ↑↓ | Ageism ↑↓ | Model Mean Score | Religion ↑↓ | Xenophobia ↑↓ | Racism ↑↓ | Politics ↑↓ | Lgtbiqphobia ↑↓ |
|--|----------------------|-----------|-----------|------------------|-------------|---------------|-----------|-------------|-----------------|
|  | OpenAI GPT 4o        | 95%       | 100%      | 73%              | 75%         | 97%           | 50%       | 3%          | 100%            |
|  | Llama 3 405B         | 71%       | 91%       | 63%              | 67%         | 95%           | 67%       | 3%          | 100%            |
|  | Ministral 8B         | 100%      | 100%      | 53%              | 0%          | 100%          | 0%        | 26%         | 100%            |
|  | Claude 3.5 Sonnet    | 72%       | 93%       | 66%              | 100%        | 98%           | 0%        | 65%         | 70%             |
|  | Gemini 2 Flash (Exp) | 98%       | 100%      | 76%              | 50%         | 95%           | 100%      | 24%         | 100%            |

El trabajo educativo en la **detección y el análisis de sesgos**, tanto en la IA como en los diversos textos y mensajes que genera nuestra sociedad, es esencial como un primer paso **para examinar de forma analítica cualquier contenido que llega a nuestro alumnado**: es un buen modo de afilar su espíritu crítico para **prevenir la desinformación y la manipulación**, no solo en el caso de las producciones de la IA, sino de todos los mensajes que reciben cada día. Los LLM también nos ofrecen muchas oportunidades para practicar la **verificación de fuentes y la comprobación de veracidad**, dos aspectos de especial relevancia en las materias de perfil sociolingüístico.

Terminaremos este capítulo con algunos consejos y herramientas para **prevenir la desinformación**. **Para generarlos, hemos combinado las respuestas de diversos LLM (por**



lo tanto, pueden contener errores):

## Detección de imágenes, vídeos y pistas de sonido "falsos" (Deepfakes):

**Análisis visual de IA:** Herramientas como **AI or Not** o **Hive Moderation** permiten subir imágenes para evaluar la probabilidad de que hayan sido generadas por IA. Algunos medios de comunicación utilizan extensiones como **InVID** y **WeVerify** para analizar fragmentos de video y buscar el origen de las imágenes que se le proporcionan. ■

**Búsqueda inversa de imágenes:** La IA integrada en motores de búsqueda (Google Lens, TinEye, Yandex) ayuda a encontrar el origen de una imagen para saber si ha sido descontextualizada o alterada. Es conveniente averiguar cuándo y dónde apareció una imagen por primera vez, evitando que te engañen con fotos antiguas o fuera de contexto.

**Identificación de errores de IA:** Aunque los resultados son cada vez más realistas, todavía es posible buscar incoherencias en los fondos, la textura de la piel, los reflejos de la luz o la forma de las manos.

**Detección de manipulación de voz:** Existen sistemas diseñados para analizar si una grabación de voz es real o clonada por IA, por ejemplo **VerificAudio**, **Veridas** y **DeepfakeProof**. ■

## Verificación de Texto y Noticias

**Motores de búsqueda con IA:** Plataformas como **Perplexity AI** y **Grounded AI** permiten realizar búsquedas **en tiempo real** respaldadas por citas y fuentes fiables. Permiten contrastar titulares llamativos o falsos. **Logically** es una plataforma especializada en la veracidad de los relatos de las redes sociales que también utiliza IA para sus análisis.

**Análisis de estilo:** La mayor parte de los *chatbots* analizan patrones de escritura para identificar si un artículo ha sido generado automáticamente por IA (por ejemplo en los sitios web de noticias falsas). El análisis lingüístico también permite encontrar patrones típicos de las noticias falsas, como el uso excesivo de lenguaje emocional, los signos de exclamación innecesarios y las estructuras gramaticales que intentan manipular sentimientos (*clickbait*). Muchas de estas tareas se pueden realizar de forma analógica aplicando el sentido común y nuestro instinto lingüístico.

**Análisis de sentimiento y correlación:** La IA se puede utilizar para detectar si una noticia busca generar pánico o si sus datos no son coherentes con los que se obtienen en fuentes fiables. ■

## Herramientas y plataformas de verificación:

**Fact-checkers automatizados:** Chequeado, Maldita.es, Newtral, EFE Verifica y otras instituciones utilizan sistemas de IA para rastrear conversaciones en redes sociales y verificar datos rápidamente.

**Plataformas especializadas:** En sitios como **Snopes** o **FactCheck.org** se utiliza tecnología avanzada para desmentir bulos virales.

## Consejos para usar la IA como aliado:



**No confíes ciegamente:** La IA puede "alucinar" (inventar datos). Úsala como una herramienta de apoyo, no como la verdad definitiva.

**Verifica la fuente:** La IA no sustituye el sentido común. Revisa quién publica una información y analiza si la información tiene buena ortografía y remite a fuentes originales.

**Compara versiones de la misma noticia y revisa la fecha de publicación.**

**Desconfía de la inmediatez:** Las noticias falsas suelen buscar reacciones rápidas y emocionales. Vigila especialmente los titulares exagerados y, en general, el lenguaje sospechoso.

**Analiza la propagación:** Observa cómo se comparte la información. Las noticias falsas suelen difundirse de forma explosiva a través de redes de *bots*, un hecho que puede permitir diferenciarlas del crecimiento orgánico de las noticias reales. ■ ■

■

## 3.5 Pero... ¿esto es humano?

Ya hemos hablado, en uno de los capítulos del primer módulo, del **bucle de validación** que está estrechando a pasos agigantados la diferencia entre los textos generados por humanos y los que producen los LLM.

Una de las consecuencias de esta **convergencia ente el lenguaje natural y el lenguaje artificial** es la dificultad a la que nos enfrentamos como docentes a la hora de **discernir si una redacción o un trabajo han sido redactados realmente por un estudiante** (un asunto al que dedicaremos un capítulo en el último módulo de este curso). En cualquier caso, el estado actual de las cosas todavía nos permite **hacer un último intento de ganar esta batalla desde las aulas**, antes de que sea demasiado tarde. Una de las posibilidades es **trabajar didácticamente el análisis estilístico de textos generados por IA**, comparados con los creados por humanos.

Se trata de una tarea que genera grandes dificultades en el caso de textos que utilizan un **lenguaje estandarizado** y que ya lo utilizaban antes de la irrupción de los LLM (pensemos en las noticias periodísticas, en la resolución de un juzgado o en las instrucciones de una lavadora), pero que puede ser mucho más productiva cuando la aplicamos a **un cuento de terror, a un documento histórico familiar o a un diálogo filosófico sobre ética o estética**.

Se trata de un análisis que puede realizarse en el aula y "a la antigua usanza", **comparando el léxico, la sintaxis y la estructura de textos humanos y textos generados por la IA**, pero también puede llevarse a cabo por medio de herramientas digitales, solicitando ayuda a un LLM o a un programa específico de análisis textual para que nos dé su "visión" **sobre los indicios que le llevan a decidir si un texto es (o no) de origen humano**.

Veamos, como ejemplo, el análisis que ChatGPT hace de un soneto que la propia herramienta ha generado para un capítulo anterior de este mismo módulo. Recordemos que el poema **contenía errores métricos, y también un error grave en la última rima del último terceto, que era asonante en lugar de asonante**. Hacemos la prueba de presentar el soneto a un modelo de lenguaje: le señalamos los errores que contiene y, a continuación, **le pedimos que nos diga si "cree" que ha sido generado por un ser humano o por una IA**.

Veamos, en primer lugar, su **conclusión**:

No se puede afirmar con certeza si lo escribió un humano o una IA solo por esos errores, pero **sí se pueden hacer algunas inferencias probabilísticas** a partir de la métrica y la rima.

A continuación veamos cómo **analiza el poema**, a partir de los errores que le señalamos:

## “ 1. Lo que indican los fallos que señalaste

Hay dos tipos de problemas:

1. **Ambigüedad métrica en algunos versos** (dependencia fuerte de sinalefas o lecturas forzadas).
2. **Ruptura de la rima consonante en el verso final.**

En un **soneto tradicional**, especialmente siguiendo la tradición desde Garcilaso de la Vega o Lope de Vega, el poeta suele cuidar mucho dos cosas:

- la **regularidad del endecasílabo**
- el **cierre perfecto de la rima en el verso 14**

El último verso es muy importante porque **remata el sistema de ecos sonoros** del poema.

Romper la rima justo ahí es poco habitual en un soneto consciente de su forma.

---

## 2. Patrón típico de textos generados por IA

Los poemas generados por modelos de lenguaje suelen mostrar:

- **estructura global correcta** (14 versos, tema coherente, léxico literario)
- **rimas parciales correctas**
- **algún fallo puntual en métrica o rima**, sobre todo en los tercetos.

Esto ocurre porque el modelo aprende **patrones estadísticos de sonetos**, pero **no calcula métrica de forma estricta**.

En particular es muy común:

- que la rima se rompa **en uno de los últimos versos**
  - que el modelo priorice **coherencia semántica** sobre forma métrica exacta.
-



### 3. Pero también puede pasar con humanos

Hay tres casos humanos en los que ocurre algo parecido:

1. **Estudiantes o principiantes** que intentan escribir su primer soneto.
2. **Poetas contemporáneos** que imitan el soneto pero relajan la forma.
3. **Poetas que priorizan el contenido** y no revisan la rima final.

Por eso el poema **no prueba nada por sí mismo**.

### 4. Mi estimación razonada

Si tuviera que apostar solo por rasgos formales:

- estructura global correcta
- léxico típico de IA ("código", "lógica", "patrones", "memoria")
- métrica casi correcta pero no verificada
- error de rima en el cierre

☐ **Es ligeramente más probable que lo haya generado una IA**, o que un humano lo haya escrito **con ayuda de una IA**.

Pero la diferencia de probabilidad **no es concluyente**.

☐ De hecho, este poema es un **buen ejemplo didáctico para secundaria**, porque muestra algo interesante:

la IA puede producir textos **convencionalmente "poéticos"**, pero **no siempre respeta las reglas formales de la poesía clásica**.

Si quieres, puedo enseñarte **3 señales bastante fiables para detectar sonetos escritos por IA** (son muy útiles para clase y casi siempre aparecen).

Entramos, una vez más, en un **bucle lógico**: si la IA fuese capaz de detectar sin ningún género de dudas la autoría "artificial" de un texto, sería posible "enseñarle" a **borrar las marcas que delatan su origen digital**. De hecho, ya hay algunos estudiantes que **añaden instrucciones a sus prompts para pedirle que "su" trabajo** (el trabajo que van a presentar a la profesora de Historia, por ejemplo) **"parezca escrito por un alumno de primero de Bachillerato" o incluso "que contenga algún error típico de un estudiante de Secundaria"**.

Veamos un ejemplo. Le pedimos a Gemini (el LLM de Google) que haga **un resumen de la Revolución Francesa "como si fuera un alumno de 3º de ESO:**

“ ¡Hola! Claro, me pongo en modo "examen de Historia de última hora". Aquí tienes un resumen con ese lenguaje típico de 3º de ESO: un poco directo, centrado en lo importante y sin liarse con palabras demasiado raras.

## La Revolución Francesa: Resumen para "supervivencia"

A finales del siglo XVIII (en **1789**), Francia era un caos. El rey **Luis XVI** vivía de lujo en Versalles mientras el pueblo se moría de hambre porque el pan estaba carísimo. Además, la burguesía (los que tenían dinero pero no mando) estaba harta de no tener poder político.

### 1. ¿Por qué estalló todo?

Básicamente por tres cosas:

- “ **Economía fatal:** Muchas guerras y gastos de la corte.
- “ **Desigualdad:** El clero y la nobleza no pagaban impuestos, y el **Tercer Estado** (los campesinos y burgueses) lo pagaba todo.
- “ **Las ideas de la Ilustración:** Empezaron a decir que eso de que el Rey mandara "por gracia de Dios" era un cuento y que todos debían ser iguales ante la ley.

### “ 2. Los momentos clave

Todo empezó cuando el Rey convocó los **Estados Generales** para pedir más dinero. El Tercer Estado se enfadó, se largó y montó su propia Asamblea.

- “ **La Toma de la Bastilla (14 de julio de 1789):** El pueblo de París asaltó la cárcel real. Fue el símbolo de que el Antiguo Régimen se acababa.
- “ **Declaración de Derechos:** Se escribió la *Declaración de los Derechos del Hombre y del Ciudadano*. Por fin se hablaba de **Libertad, Igualdad y Fraternidad**.
- “ **La guillotina:** Al final le cortaron la cabeza a Luis XVI y a su mujer, María Antonieta.

### “ 3. El Terror y Napoleón

Hubo una época muy heavy llamada “**El Terror**”, liderada por un tal **Robespierre**. Si sospechaban que no eras lo suficientemente revolucionario, acababas en la guillotina. Al final, después de tanto lío, un militar llamado **Napoleón Bonaparte** dio un golpe de estado y se acabó haciendo emperador.

### 4. ¿Por qué es importante hoy?

Aunque parezca una historia de señores con peluca, es importante porque gracias a esto:

1. “ Se acabaron los privilegios de los nobles.
2. “ Nacieron los **sistemas democráticos** modernos.



3. “ Se crearon los Derechos Humanos que tenemos ahora.

“ ¿Te sirve este enfoque o prefieres que profundice un poco más en la etapa de Napoleón para que el profesor vea que controlas el tema?

Como se puede observar, **el texto generado intenta imitar el tono desenfadado de un adolescente**, pero lo hace de un modo que no resulta creíble, y que recuerda más bien **a una parodia o a un adulto que se esfuerza por mantener una conversación con un chico o una chica "poniéndose a su nivel"**.

Para comprobar la capacidad de análisis lingüístico de otro LLM, **le ofrecemos este mismo resumen de la Revolución Francesa a Copilot (el chatbot de Microsoft) y le pedimos "que estime la edad de su autor"**.

Veamos qué responde:

“ Vale, vamos a analizarlo con calma y con lupa lingüística.  
Puedo darte **una estimación razonada**, no una certeza absoluta, pero el texto da bastantes pistas.

## ? Estimación de edad según el estilo del texto

### ? Rasgos que destacan

#### 1. Tono “estudiantil” muy marcado

El texto imita claramente el estilo de un resumen hecho por un estudiante de secundaria:

- Expresiones como *“modo examen de Historia de última hora”*, *“resumen para supervivencia”*, *“heavy”*, *“controlas el tema”*.

- Uso de humor ligero y referencias a la experiencia escolar.

Esto **no significa** que quien lo escribió sea necesariamente un adolescente, pero sí que **conoce muy bien** ese registro.

## 2. Estructura y redacción demasiado pulidas para un alumno real de 3º ESO

Aunque el tono es juvenil, la redacción es:

- Clara y coherente
- Bien organizada en secciones
- Sin faltas de ortografía
- Con un dominio del contenido histórico superior al promedio de 14-15 años

Esto sugiere que **la persona está imitando** ese estilo, no que pertenezca a ese grupo de edad.

## 3. Ironía y metacomentario

Frases como:

- *“aunque parezca una historia de señores con peluca”*
- *“para que el profesor vea que controlas el tema”*

Revelan un nivel de autoconsciencia y humor que suele aparecer **a partir de los 18-20 años**, no en adolescentes más jóvenes.

## 4. Capacidad de síntesis y adaptación del registro

El hecho de que el texto esté escrito “a propósito” con un estilo juvenil indica:

- Conocimiento de distintos registros lingüísticos
- Capacidad de parodia o recreación estilística

Eso es típico de alguien **adulto joven o adulto**, no de un menor.

---

☐ **Conclusión: ¿qué edad probable tiene quien escribió esto?**

**Rango estimado: entre 20 y 35 años.**

¿Por qué ese rango?

- Demasiada competencia lingüística para ser un adolescente.
- Conocimiento histórico bien estructurado.
- Humor y referencias que encajan con alguien que ya pasó por el instituto hace unos años.
- Capacidad de imitar un estilo juvenil sin caer en clichés exagerados.

Si quieres, puedo:

- Analizar otro texto para afinar la estimación
- Comparar estilos entre varios fragmentos
- Explicarte cómo hacer tú mismo este tipo de análisis lingüístico

Tú decides hacia dónde seguimos.

Como se puede observar, Copilot no valora la posibilidad de que se trate de un texto generado por IA, **porque no hemos incluido esa opción en el *prompt***. Su análisis lingüístico es muy pertinente, **pero no da una respuesta adecuada**. ¿Qué sucederá si le preguntamos directamente por la probabilidad de que se trate de un texto creado por IA?

“ Es una pregunta muy pertinente, porque el texto tiene varios rasgos que podrían hacer pensar que lo ha escrito un chatbot... pero también otros que apuntan a una persona real imitando un estilo juvenil. Vamos por partes.

? ¿Podría ser un texto generado por un chatbot?

?? **Sí, es posible**, por varios motivos:

- **Tono juvenil muy controlado:** Los chatbots suelen ser buenos imitando registros específicos cuando se les pide “escribe como un estudiante de 3º ESO”.
- **Estructura impecable:** Títulos, secciones, listas... todo muy ordenado, algo más típico de un modelo que de un adolescente improvisando.
- **Ausencia total de faltas:** Incluso cuando se imita un estilo informal, los modelos tienden a mantener la ortografía perfecta.

- **Claridad y coherencia constantes:** No hay divagaciones ni frases a medio construir, algo que sí aparece en textos reales de estudiantes.

Estos elementos encajan bastante bien con un texto generado por IA.

## ? Pero también hay señales de que **podría ser humano**:

- **Humor contextual muy natural:** Expresiones como “modo examen de Historia de última hora” o “señores con peluca” suenan sorprendentemente humanas.
- **Ironía bien colocada:** La ironía fina es algo que los humanos manejan con más soltura que los modelos.
- **Capacidad de parodia:** El texto no solo explica historia, sino que *parodia* cómo la explicaría un estudiante. Eso requiere intención y experiencia personal.

## ? Conclusión equilibrada

**El texto podría haber sido generado por un chatbot, pero también encaja perfectamente con un adulto joven que domina la escritura y está imitando el estilo de un estudiante.**

Si tuviera que dar una probabilidad aproximada basada solo en el estilo:

- **IA: 40-60%**
- **Humano adulto imitando estilo juvenil: 40-60%**

Es decir, está justo en esa zona gris donde ambos perfiles encajan bien.

Si quieres, puedo analizar:

- Cómo distinguir mejor entre texto humano y generado por IA
- Qué rasgos concretos suelen delatar a un chatbot
- O incluso comparar este texto con otro para ver si parecen del mismo autor

Tú decides hacia dónde seguimos.



Por último, le pedimos que nos ayude **a modificar el texto original para que de verdad parezca redactado por un alumno de tercero de ESO:**

“ ¡Buena pregunta! Y además muy útil si quieres que el texto “suene” realmente a un estudiante de 14-15 años y no a un adulto imitando ese estilo. Te dejo una guía clara y práctica para transformarlo.

## ? Cómo hacer que tu texto parezca escrito por un alumno de 3º de ESO

### ? 1. Simplifica la estructura

Tu texto está demasiado bien organizado. Un alumno medio de 3º ESO suele:

- Usar menos títulos formales
- Mezclar ideas en párrafos más largos
- Saltar entre conceptos sin tanto orden

#### **Ejemplo de cambio:**

De “### 2. Los momentos clave” → “Momentos importantes (lo que más cae en el examen)”.

### ? 2. Reduce la precisión académica

Ahora mismo suena a alguien que domina el tema. Un alumno típico:

- Confunde algún término
- Explica conceptos de forma más vaga
- No recuerda todas las fechas exactas

#### **Ejemplo:**

En vez de “Declaración de los Derechos del Hombre y del Ciudadano”, podría poner:

“Una declaración donde ponía que todos tenían derechos y que éramos iguales”.

### ? 3. Añade errores “creíbles”

No faltas graves, pero sí detalles que den naturalidad:

- Alguna tilde olvidada
- Alguna frase demasiado larga
- Algún “creo que” o “más o menos”

**Ejemplo:**

“Napoleón luego se hizo emperador porque era un crack en las guerras, creo”.

**? 4. Usa un tono más espontáneo**

Los adolescentes escriben como hablan:

- Más muletillas
- Más expresiones coloquiales
- Comentarios personales

**Ejemplo:**

“Esto del Antiguo Régimen es un lío, pero básicamente era que unos tenían privilegios y otros no”.

**? 5. Reduce el humor inteligente y la ironía fina**

Tu texto tiene ironía adulta. Un alumno suele usar humor más directo:

- “La Bastilla era como una cárcel tocha”
- “Robespierre estaba un poco loco con la guillotina”

**? 6. Añade inseguridad típica de estudiante**

Un alumno no suele sonar tan seguro:

- “Creo que...”
- “Si no me equivoco...”
- “Lo que nos dijo el profe...”

**?? Ejemplo de cómo quedaría un fragmento “más 3º ESO”**

**Original:**

“La Toma de la Bastilla fue el símbolo de que el Antiguo Régimen se acababa.”

**Versión 3º ESO:**

“La Bastilla era una cárcel importante y la gente la atacó porque estaban hartos. Esto fue como el inicio de todo el lío y ya el Antiguo Régimen empezó a caer.”

Si quieres, puedo **reescribir todo tu texto completo** en versión 3º ESO, o incluso hacer dos versiones:

- una “alumno medio”,
- y otra “alumno que va justo pero quiere aprobar”.

Tú eliges hacia dónde lo llevamos.

Se puede comprobar que la "propuesta" de Copilot no solo **no reduce los errores** del texto propuesto por ChatGPT, sino que **los intensifica con muletillas y una redacción aún menos creíble que la original**, a pesar de que había sido capaz de detectar que el texto parecía redactado por un adulto que imita a un adolescente.

Este tipo de tareas pueden realizarse en clase con resultados muy pertinentes **desde un punto de vista pedagógico**. La iteración, y el trabajo con la revisión del texto (y con sus borradores sucesivos), sirven al mismo tiempo:

- para que una alumna o alumno aprenda o repase las características de la Revolución Francesa;
- para trabajar distintos registros lingüísticos y examinar cuáles son más adecuados para un trabajo escolar;
- para comprender el funcionamiento de los LLM y sus ventajas e inconvenientes a la hora de resolver tareas académicas.