1. Impacto social, lingüístico y literario de la irrupción de la IA

- 1.1 Introducción
- 1.2 La IA en el aula de materias de perfil sociolingüístico
- 1.3 ¿Qué son los LLM?
- 1.4 ¿La inteligencia artificial es inteligente?
- 1.5 IA y literatura: una aproximación
- 1.6 Un mundo de ciencia ficción: bucles, sesgos y alucinaciones

Copyright 2026

1.1 Introducción

Os damos la bienvenida a la **parte específica** del curso de aplicación de la Inteligencia Artificial en el aula. Como sabéis, este curso forma parte de un **itinerario** que responde a una demanda creciente en el ámbito de la enseñanza, y que incluye módulos relacionados con los aspectos éticos y normativos, con la evaluación, con la creación asistida de rúbricas y con el desarrollo curricular, entre otros aspectos. En este curso, por lo tanto, **habrá menciones a esos elementos, pero nos centraremos en la relación entre la Inteligencia Artificial (IA, en adelante) y las materias de perfil sociolingüístico**, es decir, aquellas que normalmente asociamos a la denominación genérica de “Humanidades”.

> En algunos de los siguientes capítulos vamos a examinar el funcionamiento de algunas herramientas de IA para su uso en materias del ámbito sociolingüístico, pero debemos tener en cuenta que la creciente aceleración de las novedades y la proliferación de aplicaciones cada vez más potentes hacen que un curso centrado únicamente en las posibilidades de un momento concreto **quede obsoleto de forma rápida, puede que incluso en el momento mismo de su publicación.**

Por otra parte, muchos de los *chatbots* más utilizados en la actualidad (ChatGPT, Claude, Gemini…) serían capaces de **desarrollar un curso individualizado de uso de la IA para su uso en una aula de secundaria en las materias ligadas a las humanidades** o al ámbito sociolingüístico. No solo eso, sino que podrían **adaptarlo a las necesidades individuales de cada docente** (ratios, niveles educativos, alumnado con necesidades especiales, circunstancias socioeconómicas…) Existe, además, una multitud de cursos y tutoriales *online*, gratuitos y de fácil acceso, además de los que ofrecen distintas instituciones y empresas privadas. Por lo tanto, no tendría demasiado sentido ofrecer contenidos que cualquier persona con una mínima competencia digital podría obtener de forma sencilla, y este curso debe entenderse sobre todo como **un espacio de reflexión que permita adoptar una perspectiva crítica a la hora de preparar materiales y llevarlos al aula.**

> Una perspectiva crítica acerca de la IA no solo nos permitirá ser más eficaces y evitar sesgos y errores, sino que nos ayudará a transmitir a nuestro alumnado, por medio del ejemplo, **un conocimiento que exceda el uso meramente instrumental de herramientas concretas.**

Copyright 2026

Del mismo modo, el enfoque de estas páginas va a tomar como punto de partida **la producción y comprensión de textos orales y escritos** y va a dar prioridad al trabajo con Grandes Modelos de Lenguaje (LLM, por sus siglas en inglés). Una de las consecuencias de la progresiva dependencia de nuestra sociedad (no sólo los jóvenes) de los dispositivos digitales es un **acusado descenso de la comprensión lectora**, y la irrupción de la IA debe tratar de utilizarse como un modo de revertir esa tendencia, no de intensificarla. Veremos que la IA está contribuyendo al "embudo estilístico" que simplifica la sintaxis de los hablantes, y **trataremos de buscar estrategias con las que mitigar este efecto en las aulas.** Como docentes, debemos predicar con el ejemplo y prestar atención al trabajo autónomo con textos escritos de cierta complejidad.

Copyright 2026

1.2 La IA en el aula de materias de perfil sociolingüístico

En los últimos años, **la IA está transformando nuestras sociedades**, y no pasa un día sin que nos bombardeen con noticias acerca de sus aplicaciones y sus implicaciones en distintos aspectos de nuestra vida y de la organización de nuestras sociedades, desde la optimización de procesos industriales y de la gestión del hogar hasta la automatización de tareas burocráticas, el diagnóstico médico y la creación de contenidos de todo tipo. Los estudios, optimistas o apocalípticos, inundan los medios tradicionales (periódicos digitales o en papel, informativos televisivos, tertulias y programas de radio) y acaparan **millones de horas de contenido digital** (según una estimación de Gemini), desde podcasts hasta canales de YouTube especializados y tutoriales de distintas universidades, empresas e instituciones de todo tipo. Existen cada vez más medios que prestan una atención experta a un mundo que **avanza a una velocidad vertiginosa** (recomendamos por ejemplo que echéis un vistazo a El Arjonauta, el Substack de Daniel Arjona, un periodista especializado en nuevas tecnologías que mezcla alta y baja cultura, cotilleos empresariales, análisis éticos y novedades de las grandes empresas tecnológicas, con entrevistas a algunos de sus grandes gurús), pero en muchos otros casos este flujo constante de información **carece de rigor y conduce a creencias equívocas** acerca del uso de estas nuevas herramientas, relacionadas con un desfase creciente entre el uso de la tecnología y el conocimiento real sobre su funcionamiento. Un estudio reciente de la Universidad Oberta de Catalunya demuestra que un conocimiento “técnico” profundo del funcionamiento de la IA no implica necesariamente un uso más frecuente (ni mejor) de estas herramientas en el aula, pero no hay duda de que se necesitan **unos conocimientos mínimos de su construcción para ser capaz de optimizar su uso, de explorar sus posibilidades y de conocer sus peligros y sus limitaciones**.

> La entrada de la IA en las aulas también ha supuesto un gran desconcierto a la hora de gestionar las tareas con las que abordamos la enseñanza de nuestras materias. Muchos docentes se enfrentan a grandes dudas a la hora de preparar y evaluar tareas, ante la sospecha de que muchos estudiantes utilizan directamente la IA para hacer trabajos sin esfuerzo y sin un conocimiento real de la producción que presentarán en el aula o que

> entregarán al profesor.

Una consecuencia concreta de esa desconfianza es la existencia de muchos profesores y profesoras que evitan “mandar tareas escritas para casa”. **Se trata de una tendencia que elimina el proceso de revisión, fundamental para la formación de los jóvenes, según el modelo de Flower y Hayes, que considera que la escritura es un proceso cognitivo recursivo y no lineal que precisa «tiempos muertos».** La escritura no se aprende solo escribiendo, sino también en los procesos de revisión mental que tienen lugar entre una revisión de texto y la siguiente. **La alternativa, por lo tanto, no es prescindir de ciertas tareas que han probado su eficacia pedagógica, sino buscar modos de adaptarlas a los nuevos tiempos. En el módulo 5 de este mismo curso examinaremos algunas propuestas.**

Por otro lado, la IA ha abierto enormes posibilidades para la **creación de materiales atractivos de forma rápida**: resumen de textos, creación de vídeos, presentaciones y podcasts, o incluso programación de unidades didácticas y situaciones de aprendizaje completas.

> Podemos decir, sin temor a exagerar, que la generalización de la IA, y en concreto de los modelos LLM, ha supuesto un **cambio de paradigma.**

A pesar de que todavía es pronto para extraer conclusiones sobre la repercusión que estos nuevos recursos tendrán en la formación de los jóvenes y en el futuro de nuestras sociedades, **ya existen numerosos estudios sobre su uso en la educación**. Una revisión de 155 artículos científicos publicados entre 2015 y 2025 encontró un **incremento significativo en el número de publicaciones a partir de 2022**, centradas especialmente en la posibilidad de personalizar la enseñanza y aumentar la motivación de los estudiantes (Garzón, Patiño y Marulanda, 2025) . El artículo, por otra parte, mostraba también la preocupación por algunos aspectos relacionados, como **la dependencia digital, los problemas éticos, las dificultades técnicas y la resistencia por parte de algunos miembros de la comunidad educativa**. Como vemos, no se trata de ventajas y desventajas nuevas, sino de una **ampliación de los desafíos que ya presentaban las nuevas tecnologías antes de la irrupción de la IA**. Estos desafíos, por otra parte, suponen una oportunidad para tratar en el aula, desde una perspectiva moderna, asuntos relacionados con la ética, la construcción de discursos, la fiabilidad de los documentos, que nos permitirán enlazar con los grandes problemas de las materias de Filosofía, Literatura, Historia...

Como primera aproximación al uso de la IA en nuestras aulas, podemos reflexionar sobre ciertas medidas que debemos tener en cuenta en la elaboración de materiales:

- No utilizar una herramienta (por espectacular que sea) sin **haber analizado primero qué pretendemos conseguir con ella**, desde un punto de vista **pedagógico**.
- Amoldar las tareas a la **normativa educativa** vigente.

Copyright 2026

- No entregar **nunca** a nuestro alumnado materiales elaborados por IA que no hayan sido examinados **en su totalidad** por el docente.
- **Antes** de utilizar la IA para una tarea, haremos un análisis de **coste-beneficio** y valoraremos si **el proceso completo de preparación, supervisión y presumible corrección** de los materiales realmente nos va a ahorrar tiempo.

Estas medidas, por supuesto, se añaden a las que tienen que ver con los aspectos éticos (que se abordan en otro curso de este itinerario), y con la enseñanza de un uso responsable por parte del alumnado. Pero no podemos olvidar, en ningún caso, que el **uso que nosotros y nosotras hagamos de la IA** está lanzando un mensaje mucho más potente que los consejos y recomendaciones que demos a nuestros estudiantes.

Copyright 2026

1.3 ¿Qué son los LLM?

Los **grandes modelos de lenguaje** (LLM, por las siglas de su denominación en inglés, Large Language Models) son algoritmos “entrenados” con grandes volúmenes de documentos humanos para generar textos como respuesta a peticiones (*prompts*) específicas. Se trata de modelos estadísticos capaces de predecir la siguiente palabra o grupo de palabras gracias a un entrenamiento “afinado” por medio de supervisores humanos.

Mucha gente identifica la Inteligencia Artificial con los Grandes Modelos de Lenguaje, como las sucesivas encarnaciones de ChatGPT. Esta identificación, sin embargo, no es del todo exacta, y puede llevar a confusiones. **No todas las herramientas de Inteligencia Artificial son iguales**, ni en su configuración, ni en su entrenamiento ni en su propósito. La aparición de modelos integrados de inteligencia artificial generativa capaces de responder al lenguaje "natural" ha contribuido a incrementar la confusión.

En el siguiente vídeo podéis ver una introducción sencilla a algunos conceptos básicos relacionados con la historia y el funcionamiento de los LLM:

https://www.youtube.com/embed/Sz4qacFBHLk

También podéis encontrar un análisis **más técnico y detallado** (y muy didáctico) en este módulo creado para CATEDU por Luis Hueso Ibañez y Pedro López Savirón como parte del curso *La IA en educación. Una aproximación práctica.*

A continuación incluimos un **glosario** con los términos que nos pueden resulta más útiles para un enfoque centrado en las materias de perfil sociolingüístico:

IA (Inteligencia Artificial)	Conjunto de tecnologías que permiten a las máquinas realizar tareas que normalmente requieren inteligencia humana, como comprender texto, reconocer imágenes o tomar decisiones simples.
IA generativa	Tipo de IA capaz de crear contenido nuevo (texto, imágenes, audio, código) a partir de patrones aprendidos en grandes conjuntos de datos.

Copyright 2026

LLM (Large Language Model)	Modelo de IA entrenado con enormes cantidades de texto para comprender y generar lenguaje natural. Es la base de muchos asistentes conversacionales actuales.
GPT (Generative Pre-trained Transformer)	Tipo de modelo de lenguaje basado en la arquitectura transformer que ha sido preentrenado con grandes cantidades de texto para generar respuestas coherentes.
Procesamiento del lenguaje natural (PLN / NLP)	Campo de la IA que se ocupa de que los ordenadores comprendan, interpreten y produzcan lenguaje humano.
Aprendizaje automático (Machine Learning)	Técnica mediante la cual los sistemas informáticos aprenden patrones a partir de datos en lugar de ser programados con reglas explícitas.
Aprendizaje profundo (Deep Learning)	Subcampo del aprendizaje automático que utiliza redes neuronales con múltiples capas para analizar grandes cantidades de datos.
Redes neuronales	Modelos matemáticos inspirados en el cerebro humano formados por capas de nodos conectados que procesan información y detectan patrones.
Aprendizaje supervisado	Método de entrenamiento en el que el modelo aprende a partir de datos previamente etiquetados por humanos.
Token	Unidad mínima de texto que un modelo de lenguaje procesa (palabra, parte de una palabra o signo de puntuación).
Chatbot (asistente conversacional)	Programa que permite interactuar con la IA mediante conversación en lenguaje natural.
Prompt	Instrucción o pregunta que se introduce en un sistema de IA para generar una respuesta o contenido.
Ingeniería de prompts	Conjunto de técnicas para formular instrucciones eficaces y obtener mejores resultados de un sistema de IA.
Agente de IA	Sistema de IA que puede planificar acciones, tomar decisiones y utilizar herramientas externas para cumplir un objetivo (algunos sistemas, por ejemplo ChatGPT, cuentan con un Agent Mode que permite encadenar tareas y actuar con mayor autonomía).
Deep Research	Capacidad de algunos sistemas de IA para realizar investigaciones complejas, consultando múltiples fuentes y generando informes estructurados.
AGI (Inteligencia Artificial General)	Hipotético tipo de IA capaz de realizar cualquier tarea intelectual humana con flexibilidad y autonomía. Actualmente no existe.
Alucinación	Situación en la que un sistema de IA genera información falsa o «inventada» presentándola como si fuera correcta.

Copyright 2026

Sesgo (Bias)	Distorsión en los resultados de la IA causada por datos de entrenamiento incompletos o prejuicios presentes en los datos.
Alineamiento	Área de investigación que busca que los sistemas de IA actúen de acuerdo con valores humanos, normas sociales y objetivos seguros.
Supervisión humana (Human-in-the-loop)	Principio según el cual las decisiones o resultados de la IA deben ser revisados o validados por personas.

Copyright 2026

1.4 ¿La inteligencia artificial es inteligente?

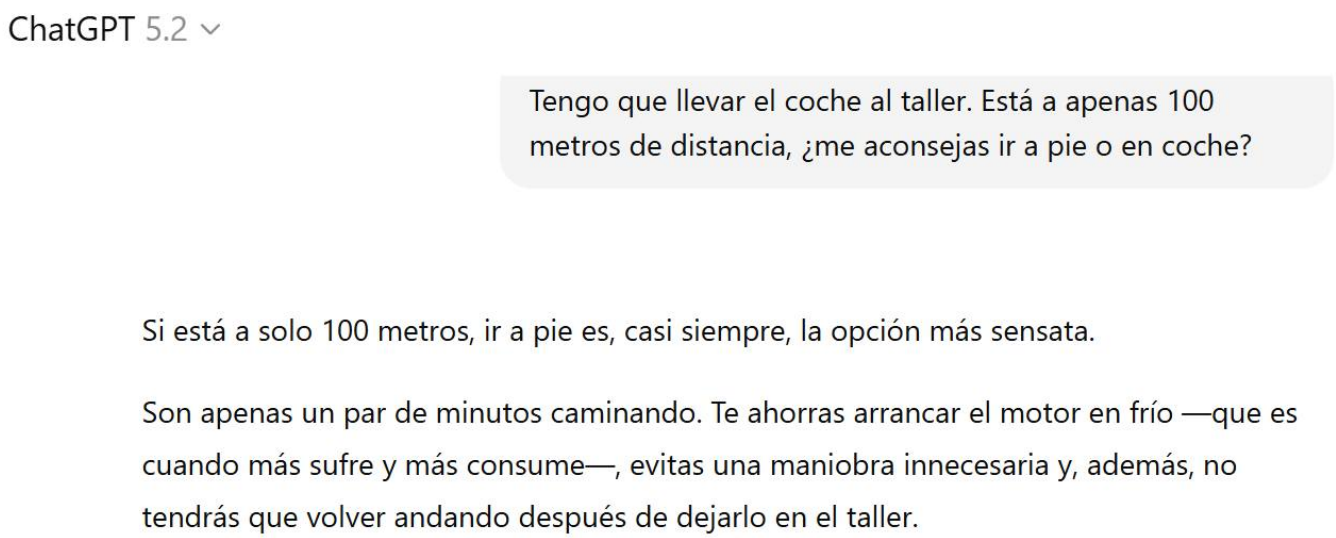
> «Los límites de mi lenguaje son los límites de mi mundo».
>
> Ludwig Wittgenstein

Los ordenadores, y otros tipos de dispositivos electrónicos, **llevan décadas realizando tareas que los humanos no somos capaces de realizar.** La calculadora de bolsillo más básica es capaz de ejecutar una operación matemática que a cualquier humano le llevaría una cantidad considerable de tiempo (en el caso de que sea capaz de realizarla). **Sin embargo, a nadie se le ocurre decir que una calculadora sea "inteligente"**, ni mucho menos que es más inteligente que un ser humano. Esta capacidad de utilizar algoritmos para realizar tareas mecánicas de forma precisa y fiable se extiende a otros muchos campos, desde la ingeniería hasta la elección de la próxima serie que veremos en nuestra plataforma de *streaming*. A nadie se le había ocurrido pensar hasta ahora, sin embargo, que una **secuencia ordenada de pasos e instrucciones para realizar un cálculo** pudiera **«cobrar conciencia de repente»**, una posibilidad que sí concedemos a los últimos modelos de IA.

Antes de profundizar en este problema debemos preguntarnos **qué es la inteligencia**. El diccionario de la Real Academia Española de la Lengua define “inteligencia”, en su primera acepción, como la **«Capacidad de entender y comprender»**. La segunda acepción ofrece una definición un poco distinta e identifica la inteligencia con la **«Capacidad de resolver problemas»**, una definición que de algún modo se ciñe más a la etimología del término, relacionada con la idea de **«elegir entre varias opciones»**. Podríamos pensar que una calculadora tiene la «capacidad de resolver problemas», siempre y cuando consideremos que una operación matemática es un problema en sí mismo (y lo es, sin duda, para muchas personas). El algoritmo de la plataforma audiovisual, por otra parte, ofrece la posibilidad de «elegir entre varias opciones», una capacidad, por otra parte, que podríamos llegar a atribuir a cualquier mecanismo de generación aleatoria de números, por ejemplo un dado. **Pero tampoco diríamos que estas capacidades hacen que la calculadora o la plataforma sean «inteligentes»**, y es evidente que su mecanismo de funcionamiento es completamente distinto (para empezar, la calculadora no comete los errores de apreciación que tanto nos intrigan en la oferta «personalizada» de canciones o de películas).

Copyright 2026

¿Qué sucede con los LMM? ¿Hasta qué punto podemos considerar que son «inteligentes»? Veamos un ejemplo en el que se hace una pregunta sencilla, de «sentido común», a una de las últimas versiones del *chatbot* más popular:



Como vemos, el LLM falla de forma estrepitosa (y cómica) en su respuesta. El modelo no es capaz de comprender que para llevar el coche al taller es necesario... llevar el coche al taller. El error no está en el algoritmo (aunque no hay duda de que en un futuro inmediato los modelos serán capaces de resolver este problema y otros similares con respuestas más satisfactorias), sino en nuestra percepción de lo que es capaz de hacer un *chatbot*. **El asistente no es capaz de responder a la pregunta precisamente porque es una pregunta con una respuesta tan obvia que nadie la ha planteado nunca, y por lo tanto la opción correcta no aparece en los textos de entrenamiento.**

> La clave está en el **lenguaje** y nos remite de nuevo a la primera definición del diccionario de la Real Academia, esa que nos habla de la «capacidad de entender y comprender». **Identificamos las ideas con su verbalización** (como en la cita de Wittgenstein que abre este capítulo), y **la capacidad de razonar con la simulación de razonamiento expresada por medio del lenguaje**.

La lingüista Emily Bender, que dirige el Laboratorio de Lingüística Computacional de la universidad de Washington, acuñó el término **"loros estocásticos"** para referirse a los Grandes Modelos de Lenguaje, una expresión que ha tenido cierto éxito y que se sigue utilizando: al igual que los loros, los LLM utilizan palabras que no comprenden y, en el mejor de los casos, las utilizan para responder a estímulos de forma condicionada, **sin entender la relación entre su producción lingüística y el evento que la provoca**. La propuesta de Bender, y de los otros firmantes de este artículo, consiste en limitar la cantidad de datos que se utilizan para entrar a los modelos de

lenguaje y **cuidar más la calidad de la información en que se basan, con una fuerte supervisión humana**.

Otra de los argumentos que se utilizan para demostrar que la IA no es capaz de "pensar" tiene que ver con **sus limitaciones a la hora de detectar, comprender y generar humor**. En una entrevista con el diario *The Guardian*, el profesor José Camacho Collados, responsable de un estudio sobre el modo en que la IA procesa los chistes, afirmó que “En general, los LLM tienden a memorizar lo que han aprendido durante su entrenamiento. Por ello, detectan bien los juegos de palabras ya existentes, pero eso no significa que realmente los comprendan”. Y continuó: “Hemos sido capaces de engañar a los LLM de forma sistemática modificando chistes existentes, eliminando el doble sentido que hacía que el juego de palabras original funcionara. En estos casos, **los modelos asocian las frases con juegos de palabras previos e inventan razonamientos de todo tipo para justificar que se trata de un juego de palabras**. En última instancia, hemos descubierto que su comprensión de los juegos de palabras es una ilusión.”

Uno de los intentos de desarrollar la IA generativa para que tenga una capacidad real de razonar, tal y como la entendemos los humanos, ha sido la creación de **Grandes Modelos de Razonamiento** o LRM (por sus siglas en inglés), sistemas programados para superar las limitaciones de los modelos puramente lingüísticos por medio de procedimientos computacionales distintos, capaces de **revisar** los pasos previos (un proceso que implica respuestas más lentas).

Sin embargo, los resultados de estos modelos siguen siendo decepcionantes y un estudio publicado en noviembre de 2025 los comparó con los LMM tradicionales y descubrió que, de hecho, los LLM respondían mejor a tareas sencillas (se equivocaban menos) y que, aunque los LRM superaban a los LLM en tareas de complejidad media, **ambos tipos de IA colapsaban (es decir, tenían una tasa nula de éxito) ante tareas complejas que exigían razonamientos con un número elevado de pasos**.

> Como docentes, nuestra obligación es crear oportunidades para que nuestro alumnado comprenda que la IA, al menos en su estado actual de desarrollo, **no es capaz de llevar a cabo el proceso de razonamiento que asociamos con la inteligencia**. En este curso veremos algunas actividades que permitirán abordar los LLM desde una **perspectiva crítica.**

1.5 IA y literatura: una aproximación

En 1981 el escritor polaco **Stanislaw Lem** publicó *Golem XIV*, la recopilación de **dos conferencias dictadas por un ordenador** en un futuro cercano. La curiosidad, profética en cierto modo, es que estos **supuestos textos de una IA** venían precedidos por un prólogo sobre la historia de los superordenadores escrito por un profesor ficticio, Irving T. Creve, y fechada... **en el año 2026**.



No se trata de un caso aislado: en la historia de la ciencia-ficción aparecen muchos casos de **ordenadores «creativos»** que componen música, escriben poemas y realizan otras tareas que parecen reservadas a los humanos. Sin embargo, **el intento real de crear programas capaces de generar literatura no se ha generalizado hasta la llegada de los LLM**, aunque hubo intentos anteriores, por ejemplo este del año 2008, en el que un grupo de programadores y filólogos colaboró para conseguir un *software* capaz de reescribir *Ana Karenina*, la famosa novela

de Tolstói, **con el estilo del escritor japonés Haruki Murakami**.

Con la llegada de ChatGPT, sin embargo, ya no hizo falta reunir a un grupo de expertos, y cualquier usuario puede crear un texto literario de cierta complejidad en unos minutos. La facilidad es tal que en 2023 Amazon **tuvo que restringir el número de libros que un autor o autora podía subir a su plataforma** de obras autoeditadas. Desde entonces, si eres escritor y quieres que tus libros se vendan en Amazon, **solo puedes publicar tres libros al día**.

Parece, sin embargo, que **las obras creadas con IA no tienen todavía la calidad de un autor profesional**, aunque en muchos casos sean indistinguibles de las de los malos autores (no hay duda de que incluyen menos faltas de ortografía). En una iniciativa de la UNED, **el escritor argentino Patricio Pron se enfrentó a ChatGPT en un "duelo literario"**. Cada uno de los contrincantes tenía que escribir **el título de treinta películas ficticias**, y desarrollar el argumento de cada una de ellas en seiscientas palabras. A continuación, un jurado de expertos valoraría el mérito artístico de todas las propuestas, **sin saber a quién correspondía cada una**. El resultado fue apabullante, y Pron venció a ChatGPT con títulos como *Después de todo lo que casi hice por ti*, *Enfermedad mental tres días a la semana*, *La mujer lego* y *Escoge una carta cualquiera. No, esa no, otra,* que el jurado consideró, con buen criterio, mucho mejores que *Fragmentos de un ayer invisible*, *La ciudad invertida*, *La melodía olvidada*, *El último vuelo de la mariposa* y *Huellas en el mar de arena.* Cualquier grupo de expertos con un mínimo nivel de exigencia habría ofrecido el mismo veredicto.

En la práctica, la inteligencia artificial no está sustituyendo (todavía) a los escritores, sino que está dando lugar a formas de creación híbrida en las que humanos y máquinas colaboran.

Veamos algunos usos frecuentes de la IA en la escritura creativa:

- Generación de **ideas o inspiración** para relatos o personajes.
- Exploración de **variantes narrativas** (cambiar el punto de vista, el tono o el estilo).
- Creación de **borradores iniciales** que luego son revisados por el autor.
- **Experimentación estilística**, por ejemplo imitando ciertos registros o géneros literarios.
- **Documentación** histórica.
- Corrección **ortográfica y estilística**.

En este sentido, la IA puede entenderse como una **herramienta de apoyo al proceso creativo**, comparable en algunos aspectos a los diccionarios, los correctores o los programas de edición de texto.

Algunos usos son más interesantes y están ofreciendo **nuevas formas de creatividad**. El escritor murciano Javier Moreno estaba escribiendo una novela cuando empezó a experimentar con distintas herramientas de generación de imágenes con IA. Su intención era solamente **probar nuevas formas de narración**, pero el resultado le gustó tanto que decidió crear una serie de

audiovisual que mantiene todas las características de extrañeza que caracterizan su obra, pero en un **nuevo formato**. A continuación podéis ver el primer capítulo de esta **«literatura ampliada» que presenta una nueva forma de colaboración entre los humanos y las máquinas**:

https://www.youtube.com/embed/omurcd3HR9w?si=S4yYoY0KuA85Zlht

1.6 Un mundo de ciencia ficción: bucles, sesgos y alucinaciones

> «En el caso de las llamadas inteligencias artificiales han decidido llamar alucinaciones a lo que son errores crasos. ¿Por qué? Una alucinación humana es una percepción que se produce sin un estímulo externo que la sostenga (...) Pero las máquinas, hoy por hoy, no saben lo que dicen ni les importa y con cierta frecuencia combinan las palabras de forma que generan afirmaciones falsas. Supogo que lo llaman alucinar porque fabrican mentiras sin saber que lo son. Aunque quienes les pusieron el nombre deberían buscar otro para recordar que cuando las inteligencias artificiales hacen afirmaciones verdaderas tampoco lo saben».
>
> *Te siguen*, Belén Gopegui Durán

El uso creciente de sistemas de inteligencia artificial en contextos educativos hace necesario comprender no solo sus capacidades, sino también **sus limitaciones y riesgo**s. Entre los más importantes se encuentran los **sesgos** y las **alucinaciones**, dos fenómenos que afectan especialmente a los sistemas de IA generativa y a los modelos de lenguaje utilizados en los *chatbots* a los que nuestro alumnado recurre pare realizar cualquier tipo de tarea. Es imprescindible que los docentes —**especialmente en el ámbito de las humanidades**— comprendan estos conceptos y que fomenten un uso crítico, responsable y pedagógicamente adecuado de estas herramientas. Para ello, una vez más, deben empezar **por usar la IA de forma crítica, responsable y pedagógicamente adecuada**.

Copyright 2026

Los sesgos de los LLM

Los sesgos de la IA aparecen cuando un sistema genera resultados que favorecen o perjudican sistemáticamente a determinados grupos, perspectivas o interpretaciones. Estos sesgos no suelen ser intencionales, sino que **surgen principalmente de los datos con los que se entrena el sistema o de la supervisión humana que los etiqueta y los valida**. No debemos olvidar que, en origen, la IA depende de programas y modelos generados por humanos: no es extraño, por tanto, que los prejuicios de la sociedad que las ha creado **se reproduzcan en sus respuestas**.

Los modelos de lenguaje actuales se entrenan con enormes cantidades de texto procedente de libros y artículos, pero también de internet y de otras fuentes digitales. **Estos materiales reflejan inevitablemente las desigualdades, prejuicios, visiones culturales y limitaciones históricas de las sociedades que los produjeron**. Cuando la IA aprende a partir de esos datos, puede reproducir —o incluso amplificar— esas mismas tendencias.

Por ejemplo, un sistema de IA puede:

- asociar determinadas profesiones con un género concreto;
- ofrecer interpretaciones históricas centradas en perspectivas occidentales;
- reproducir estereotipos culturales presentes en los textos de entrenamiento.

En el ámbito educativo, este problema es especialmente relevante en asignaturas como historia, filosofía, literatura o ciencias sociales, donde l**as interpretaciones, narrativas y perspectivas culturales desempeñan un papel fundamental**. Si se utiliza la IA como fuente de información sin un análisis crítico, los sesgos del sistema pueden influir en la forma en que se presentan ciertos temas o en el modo en que se representa a determinados colectivos. Estos sesgos, a su vez, **retroalimentarán los valores que los han producido en un bucle muy difícil de detener**.

Es importante que comprendamos que los sesgos pueden aparecer en diferentes niveles:

- Sesgo **en los datos**: cuando el conjunto de entrenamiento no representa adecuadamente la diversidad del mundo real.
- Sesgo **en el diseño del modelo**: cuando las decisiones técnicas o de optimización priorizan ciertos resultados.
- Sesgo **en la interacción con el usuario**: cuando las preguntas o *prompts* refuerzan determinadas perspectivas.

> Por esta razón, la alfabetización en IA implica desarrollar en el alumnado —y en el profesorado— una actitud crítica similar a la que se aplica al análisis de fuentes históricas o textos argumentativos: **toda información debe contextualizarse, contrastarse y evaluarse.**

Copyright 2026

Las alucinaciones de los LLM

Otro fenómeno importante son las llamadas alucinaciones. En inteligencia artificial, una alucinación ocurre cuando un sistema genera información incorrecta, «inventada» o no verificable, **pero la presenta con una apariencia de seguridad y coherencia**.

Esto puede suceder porque los modelos de lenguaje no “saben” cosas en el sentido humano del término. En realidad, funcionan prediciendo qué palabras son más probables en una secuencia, basándose en **patrones estadísticos** aprendidos durante el entrenamiento, como hemos visto en un capítulo anterior de este mismo curso. **El objetivo del sistema es producir una respuesta lingüísticamente plausible, no necesariamente garantizar que sea verdadera.**

Como consecuencia, la IA puede:

- inventar referencias bibliográficas que no existen;
- atribuir frases e ideas a personas vivas o muertas que nunca las pronunciaron ni escribieron;
- confundir fechas, nombres o acontecimientos históricos;
- construir explicaciones aparentemente coherentes pero incorrectas.

En el contexto educativo, **este comportamiento plantea un riesgo evidente si el alumnado utiliza estos sistemas como fuente primaria de información sin verificar los resultados**. Sin embargo, también ofrece una oportunidad pedagógica interesante: enseñar a detectar errores, contrastar información y analizar críticamente las respuestas generadas por la IA.

De hecho, **algunos docentes utilizan deliberadamente ejemplos de respuestas incorrectas generadas por IA como ejercicios de pensamiento crítico**, pidiendo al alumnado que identifique errores, busque fuentes fiables y reconstruya la información correcta.

Implicaciones educativas

Para el profesorado, comprender los sesgos y las alucinaciones no significa evitar completamente el uso de la IA, sino integrarla con criterios pedagógicos claros. Algunas recomendaciones habituales incluyen:

- **Verificar** la información generada por la IA mediante fuentes académicas o institucionales.
- Utilizar la IA como **punto de partida**, no como autoridad final.
- **Fomentar el pensamiento crítico**, analizando con el alumnado posibles errores o perspectivas parciales.
- **Explicar cómo funcionan los modelos**, para que los estudiantes entiendan sus limitaciones.

Copyright 2026

- **Promover la supervisión humana,** especialmente en tareas que impliquen conocimiento factual o interpretación histórica.

En las materias de perfil sociolingüístico, en las que el análisis crítico de las fuentes es una competencia central, estos riesgos pueden transformarse en oportunidades educativas. Analizar cómo y por qué una IA puede equivocarse **permite reflexionar sobre la naturaleza del conocimiento, la construcción de los relatos históricos y el papel de los datos en la producción de información**.

Por ello, la integración de la IA en educación debe ir acompañada de **alfabetización digital, pensamiento crítico y supervisión humana.** En el ámbito de las humanidades, estas herramientas pueden convertirse no solo en recursos didácticos, sino también en objetos de análisis que ayuden a comprender mejor cómo se produce, se transmite y se interpreta el conocimiento en la era digital.

El embudo estilístico y el bucle de validación

Un último peligro relacionado con la IA, y al que se suele conceder menos importancia que a los sesgos y las alucinaciones, tiene que ver con la **naturaleza estadística** de los LLM. Como todo sistema estadístico, **los LLM tienden a eliminar los resultados con un peso estadístico marginal**. Se trata, por una parte, de un buen sistema de control que permite "borrar" de los resultados los datos incorrectos y las teorías descabelladas sobre cualquier fenómeno. Veamos un ejemplo sencillo: si la palabra "alucinación" aparece un millón de veces en los textos que se proporcionan a un modelo de lenguaje para su entrenamiento, podemos tener la certeza de que, al tratarse de textos producidos por humanos, habrá al menos varios casos en los que la persona que escribió la palabra tuvo un desliz y escribió, por ejemplo, "aluciniación". El modelo no tiene ningún modo de saber cuál de las dos palabras es la correcta, pero eliminará la que aparece en pocas ocasiones, y "entenderá" que se trata de una errata. **Se trata de una solución eficaz, pero que elimina también las interpretaciones válidas que no son mayoritarias.**

Sucede algo parecido con el estilo. Una de las características del lenguaje humano es su singularidad: el léxico y la sintaxis de cada hablante individual de un idioma, lo que se conoce como su **idiolecto**, tiene unas características particulares que dependen **de su formación, su entorno sociocultural, su localización en el tiempo y el espacio, así como de factores psicológicos**, hasta el punto de que se han desarrollado modelos de **lingüística forense** que permiten descartar o confirmar la autoría de una carta de amenaza o de un mensaje de correo electrónico. La naturaleza estadística de los LLM, sin embargo, tiende a borrar todas esas diferencias y a crear un "lenguaje estándar" que es una mezcla de los **rasgos mayoritarios** de todos los textos con los que se ha entrenado.

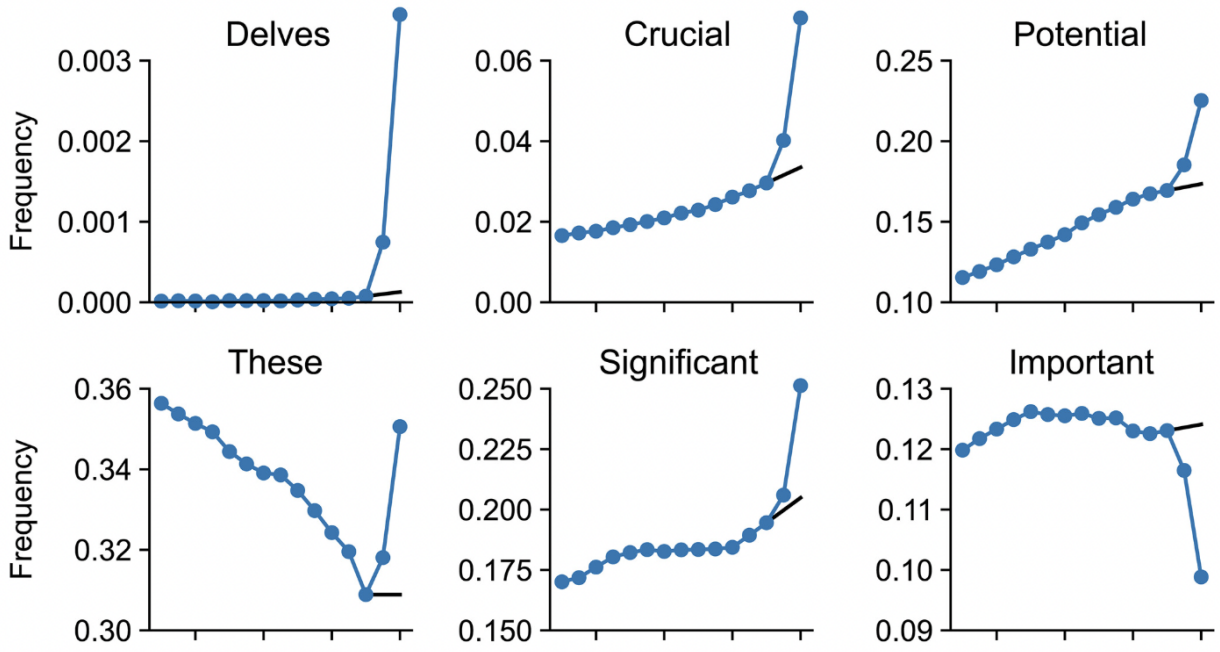
> Cuando utilizamos un texto generado por un LLM en el aula, o en un texto que ofrecemos a nuestro alumnado,**estamos borrando nuestra identidad individual como hablantes y estamos validando la producción lingüística de la IA como el modelo correcto de expresión escrita**.

Ese proceso de **"borrado de la diferencia"** tiene tres consecuencias inmediatas:

- **empobrece** el léxico;
- **simplifica** la sintaxis;
- crea una **ilusión de uniformidad**.

Se trata, en los tres casos, de consecuencias que apuntan directamente contra **pilares fundamentales de la formación humanística que debemos ofrecer en nuestras aulas**.

Los efectos de esta apisonadora estilística ya se están percibiendo. Un estudio reciente publicado en la revista *Science* ha detectado **un aumento significativo de determinados términos en las publicaciones científicas en inglés**: palabras como "delves" han disparado su frecuencia en los *abstracts* de los estudios de determinadas ramas de la ciencia, un giro que solo se puede atribuir al uso generalizado de *chatbots* y que **puede comprobarse en la siguiente gráfica de frecuencias:**



Pero el bucle de la validación no termina aquí, y tiene todavía otra vuelta: según estimaciones recientes, **el 90% de los nuevos contenidos que se suben a internet cada día ya están generados por IA**, incluso en los medios profesionales (que en ocasiones ya indican que las

noticias que ofrecen han sido generadas por un LLM), de modo que las siguientes generaciones de IA se van a entrenar con textos humanos que no serán humanos, aunque habrán sido validados por nuestro uso. **Esta "segunda vuelta" contribuirá a eliminar más singularidades y a crear más dificultades para distinguir un texto generado por IA.**

> Como docentes de materias de perfil sociolingüístico debemos tratar de **ampliar el léxico de nuestros alumnos y alumnas, mejorar su expresión escrita, estimular su comprensión de estructuras lingüísticas complejas y ayudarles a comprender e interpretar las diferencias que existen en cualquier conjunto de datos**. Todos estos objetivos pueden lograrse gracias al uso de la IA en el aula, pero su desarrollo también puede perjudicarse si utilizamos los LLM **de forma irreflexiva y poco crítica**.

Copyright 2026