IA
ASPECTOS TÉCNICOS

ABRIENDO LA CAJA NEGRA:
¿QUÉ ES LA IA Y CÓMO FUNCIONA?

@prender

CONSEJO GENERAL DE EDUCACIÓN

Preguntarnos ***¿qué es la inteligencia artificial y cómo funciona?*** no solo nos permite reconocer sus modos de seleccionar, procesar y “generar” textos: nos marca también posibles formas de trabajo en el aula.

¿Por qué equivocan algunas respuestas? ¿Por qué se dice que tienen sesgos? ¿Por qué es tan necesario reconocer su no neutralidad? ¿Por qué algunas responden diferente a otras y cómo reconocer sus lógicas de funcionamiento?

La IA —especialmente la generativa— lleva al extremo lo que algunos autores de los estudios culturales del software llaman «opacidad» (Manovich, 2013). Se refieren así al acceso de los usuarios a partir de una interfaz que no muestra el funcionamiento de los sistemas informáticos, lo que conduce a una utilización de éstos ingenua y poco crítica. El efecto de caja negra es negativo si queremos un uso enriquecido de la tecnología que empodere a quienes la utilizan y los vuelva sujetos activos en la sociedad del conocimiento.

>>> 1. Modelo de lenguaje y arquitectura

Cada IAG está entrenada con modelos diferentes (por ejemplo, GPT, Gemini, Claude, LLaMA) que condicionan:

- Cómo "entiende" la pregunta (qué patrones lingüísticos prioriza).
- Qué tipo de respuestas considera más probables o adecuadas.
- Qué fuentes o corpus de entrenamiento ha tenido (puede haber más enfoque en ciertos temas, estilos, idiomas, etc.).

➡ Esto hace que, frente a la misma consigna, una IA pueda dar una respuesta más formal, otra más conversacional, y otra más técnica.

Un buen ejercicio es hacer **la misma consulta a diferentes IAG** para ver cómo son las devoluciones.

>>> 2. Estilo de generación de texto

Las IAG difieren en cómo estructuran sus respuestas:

- Algunas generan texto de forma más lineal (responden de inmediato), otras revisan o planifican antes de escribir.
- Hay diferencias en el nivel de detalle, organización interna, y el uso de conectores o ejemplos.
- Algunas priorizan la precisión sintáctica y otras la fluidez narrativa.

➔ Por eso, una misma respuesta puede variar en tono, claridad o extensión.

>>> 3. Interacción y adaptabilidad

Las IAG también varían en su capacidad para:

- Adaptarse al estilo del usuario (técnico, informal, educativo,etc.).
- Mantener el contexto de una conversación (memoria o historial).
- Hacer preguntas de vuelta o refinar la consigna, si es ambigua.

➔ Algunas son más “dialogantes”, abiertas al ida y vuelta, otras responden de forma más bien directa.

>>> 4. Acceso a información externa

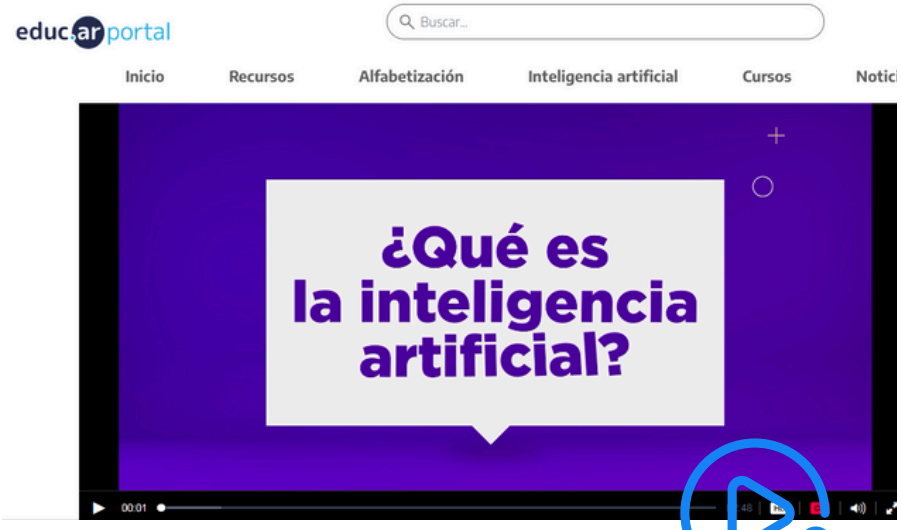
Existen IAG que:

- Funcionan offline, sin conexión a bases de datos actualizadas.
- Otras acceden a información en tiempo real (búsqueda en la web).
- Algunas integran datos de sistemas específicos (como documentos, mails, código, imágenes, etc.).

➔ Esto impacta en la actualidad, precisión y relevancia de lo que responden.

Visibilizando sus mecanismos lograremos mejores prácticas tecnológicas pero también sociales, una utilización provechosa, y no instrumental, que permita extraer información valiosa para la gestión, la política, las prácticas de enseñanza y no para la masificación automática (Lipenhotz y Sagol, 2024)

- **Necesidad de comprender qué son y cómo funcionan las IAG**



Microaprendizajes de Educ.ar:
¿Qué es la inteligencia artificial?

Con IA nos referimos a sistemas creados para cumplir tareas específicas y ser capaces de aprender por sí mismas. El método a través del cual estos sistemas (computadoras) aprenden es el **machine learning o aprendizaje automático**. En estos modelos el sistema se entrena a partir de los datos que se ingresan (¡cuantos más, mejor!), y sobre estos aplica una secuencia de pasos (instrucciones o algoritmos) que le permiten arrojar resultados o predicciones.

- **Datos, algoritmos y predicciones**[1]

La inteligencia artificial necesita de —¡enormes!— conjuntos de **datos** en los que se basa el algoritmo para realizar ciertas predicciones. Es importante señalar que, por lo general, pensamos en los datos como números, pero también podrían ser textos, videos, imágenes u otros elementos. Por ejemplo, un sistema de recomendaciones de series, como los de las plataformas de *streaming*, se basa en datos obtenidos a partir de nuestro comportamiento en dicha plataforma, como la información sobre nuestra navegación, qué series ya vimos y cómo las calificamos.

[1] Capturas del material del Proyecto HumanIA - Chicos.net. Capítulo 2 ¿Cómo aprende la IA? https://www.chicos.net/humania/

Los **algoritmos** son conjuntos de reglas que permiten tomar decisiones. Si volvemos al ejemplo de las plataformas de *streaming*, es necesario un algoritmo que tome esos datos y los procese de alguna manera, a partir de una serie de pasos, para tomar una decisión. Por ejemplo, podría recolectar información acerca de qué otras series vieron aquellos/as usuarios/as que les pusieron una calificación alta a las mismas series que tú, o que compartan características demográficas (la edad o región donde viven). Estos algoritmos podrán identificar patrones de calificación similares para realizar una recomendación. Con el aprendizaje automático, estas reglas se van afinando a medida que se obtienen nuevos datos. Es decir, cuantas más series veas y más las califiques —y cuantos más usuarios y usuarias hagan lo mismo—, más se afinará este algoritmo de recomendación.

Las **predicciones** son el resultado del procesamiento de esos datos por parte del algoritmo. Un ejemplo de estas predicciones serían, entonces, las recomendaciones que realiza la plataforma de *streaming* a sus usuarios/as. Una vez realizada la predicción, el sistema necesita de nuestra respuesta (por ejemplo, que califiquemos la nueva serie) para determinar si la predicción fue adecuada o no. Otros ejemplos posibles de estas predicciones pueden ser: las sugerencias y los resultados que te muestra un buscador en Internet de acuerdo a tus búsquedas previas; las decisiones que toma el servicio de correo electrónico para etiquetar un e-mail como spam y que no aparezca en tu bandeja de entrada, o el predictor de palabras en el e-mail o el servicio de chat que define algunos términos o frases sugeridas de acuerdo a lo que estás escribiendo y a los datos pasados.

El concepto de ***opacidad*** nos trae una imagen concreta de lo que no llegamos a ver con claridad y nos conduce por un sendero reflexivo que necesitamos transitar al momento de adoptar herramientas de IAG en cualquiera de las dimensiones que planteamos en este curso: como asistente en las tareas docentes y como recurso didáctico.

"Descajanegrizar" (abrir la caja negra) es una palabra poco conocida, pero la planteamos aquí en tanto nos invita a develar esa opacidad y a intentar comprender las características que la convierten en una tecnología tan singular como potente si sabemos incorporarla criteriosamente.

Para comprender un poco más este término y desentrañar algunas lógicas de funcionamiento de las IAG *(Lipenhotz y Sagol, 2024)*, compartimos un ejercicio que puede ser llevado al aula de manera directa con estudiantes, pero que es oportuno y necesario ensayarlo.

››› Jugando a entrenar una IA: una experiencia exploratoria:

Estas actividades están orientadas a la exploración del funcionamiento de la IA, para aproximarnos de este modo, a conceptos claves e implicaciones inherentes a su incorporación en las prácticas educativas. Las mismas pueden replicarse bajo una modalidad de taller con docentes o estudiantes.

Introducción

En principio, se sugiere propiciar una instancia de conversación con el objetivo de visibilizar historias cotidianas de uso de tecnología en general y de las aplicaciones más populares que utilizan IA en particular. La actividad busca promover la reflexión sobre aquello que conocen o desconocen de la IA, identificando en sus propias vidas los diferentes modos que reviste su utilización y cómo se vivencian.

Se comienza con preguntas exploratorias tales como: ¿Qué saben de la IA? ¿Qué conocen de la IA? ¿Les da miedo? ¿Les causa curiosidad? La IA está en boca de todo el mundo, pero poca gente sabe lo que es: ¿Cómo funciona? ¿Hasta dónde puede llegar? ¿Cuáles son sus limitaciones? ¿Dónde ven o intuyen que funciona la IA?

Como respuesta a estas preguntas se pueden sugerir ejemplos de plataformas o redes sociales que emplean IA y que son muy populares y utilizadas en la actualidad, como por ejemplo Netflix, Amazon, Mercado Libre, Facebook o Instagram. ¿Cómo usamos estas plataformas? ¿Qué sucede cuando calificamos, recomendamos o visitamos frecuentemente un mismo contenido? ¿Qué contenidos se ofrecen a cada persona? En algunos casos las personas usuarias aportamos esa información de modo voluntario y explícito (cuando llenamos un formulario, por ejemplo); en otros, a través de nuestra navegación. Estas plataformas recogen y analizan los datos de quienes las utilizan: qué búsquedas realizan, cómo las valoran, qué días y horarios son los más concurridos, cuánto es el tiempo de conexión y cuándo se suspende, si se retoma o abandona, etc. Sobre esa base, personalizan la oferta de contenidos.

De esta manera, la información que producimos y brindamos como personas usuarias se utiliza para ofrecernos contenidos que prolonguen la permanencia y el consumo. Incluso se ha llegado a manipular la distribución de contenidos según perfiles en campañas políticas o debates públicos[5].

A partir de estas indagaciones y del intercambio entre pares, pueden surgir algunas ideas:

- En primer lugar, la evidencia de que la mayoría de las personas estamos en contacto con la IA aún cuando no lo sepamos.
- En segundo lugar, se puede trabajar el concepto de «huella digital», que es una metáfora que se refiera a la información que dejamos en las plataformas al utilizarlas. Las huellas son datos que dejamos de modo indirecto y muchas veces involuntariamente y sin saberlo. Los programas de IA trabajan con esos datos, van aprendiendo y entregando respuestas (recomendaciones, ofertas) a partir de esa información. Si bien en esta instancia no explicaremos el proceso en detalle, es importante entender que la tecnología toma la información que las mismas personas usuarias proveemos.

Desarrollo

Después de la introducción se puede comenzar a trabajar con el juego de Code Studio. Se trata de una actividad lúdica que muestra, mediante una práctica sencilla, qué significa entrenar una máquina.

El juego está disponible online

Los objetivos de esta actividad son:

- Conocer el proceso de trabajo de los motores de IA a partir de un uso simple.
- Experimentar y evaluar directamente cómo se entrenan y cómo aprenden los programas de IA.
- Introducir el concepto de «patrón».
- Introducir el concepto de «aprendizaje supervisado».

Consignas:

1. En forma individual o en grupos de 2 o 3 poner la versión en español (abajo a la izquierda).
2. En la página hay una base de datos ya dada de peces y otros elementos que es posible encontrar en el mar, que no son peces sino desechos (carozos de manzanas, botellas de plástico, latas, etc.). El mecanismo del juego consiste en señalar si algo «es un pez» o «no es un pez», hasta un número que, en principio, puede ser 30.
3. Una vez entrenada la máquina con un número de casos definido previamente, hay que poner en marcha el robot y observar cómo identifica los casos, es decir, cómo reconoce qué es un pez y qué no lo es. Cabe señalar que el programa realiza este reconocimiento a partir de la información que le dimos y también que, en algunos casos, el reconocimiento es erróneo. Estas observaciones deben anotarse para retomarlas en la conversación final.

5 Para profundizar en estos temas, sugerimos la lectura de: Frenkel, S., Kang, C. (2021). Manipulados: la batalla de Facebook por la dominación mundial. Random House. También se puede consultar un resumen de las operaciones de las plataformas en https://telefonicatech.com/blog/como-usan-la-ia-las-plataformas-de-streaming.

¿Qué hicimos durante el "juego"? Entrenamos un robot para que aprenda indicando repetidamente -con imágenes en este caso- qué es un pez y qué no. Le mostramos, para ello, muchas imágenes con peces y otras tantas sin peces (datos de entrenamiento) y fuimos instruyendo a la máquina como reconocer patrones. Esto se realiza con las características positivas pero también con las negativas: por eso se muestra, se enseña, qué «es pez» y qué «no es pez» *(Lipenhotz y Sagol, 2024)*

Desmontar la apariencia de neutralidad, objetividad e infaliblilidad es el primer gran desafío que deberíamos asumir en el aula, cuestión que nos conduce a su vez al núcleo de la alfabetización digital en clave contemporánea.

Muy influenciados por las narrativas tecnoutópicas y por el propio marketing de los productos, vivimos en un clima de época donde “tecnología" suele asociarse con precisión, eficiencia y racionalidad. La información que nos llega procesada es que estamos ante máquinas "imparciales", que están más allá de ideologías o intereses. Las estéticas y los diseños que nos muestran las pantallas refuerzan esta idea: interfaces "limpias", lenguaje formal, tono empático, seguridad en las respuestas: todo esto construye una imagen de autoridad confiable. Y cuanto más humana parece la IAG, más tendemos a creer que “piensa” o “sabe” lo que dice.

Las representaciones sociales que construimos sobre estas tecnologías alimenta una narrativa donde la IAG parece resolver problemas de forma neutral y superior al humano. La escuela puede y debe construir un discurso a contrapelo que en lugar de validar su condición de “inteligentes” y "autónomas", permita verlas y abordarlas como lo que son: sistemas estadísticos programados.

Esta posición da vuelta un imaginario construido cuya principal virtud es que se ha tornado casi incuestionable, que se resume en esta frase que da título a un artículo.

Si la IA dice la verdad es solo por accidente
Artículo periodístico Portal Cenital

Las IAG no entienden lo que dicen. Generan textos a partir de patrones estadísticos. Esto puede llevarlas a “alucinar”, es decir, inventar datos o citas sin darse cuenta. Lo que sucede muchas veces es que **los datos con los que fueron entrenadas reflejan desigualdades, estereotipos** o puntos de vista dominantes. Esos sesgos se reproducen (y a veces se amplifican) en sus respuestas.

Sus fallas no siempre son obvias. Un error de contexto, una afirmación falsa o una sugerencia aparentemente razonable puede pasar como correcta si no hay una mirada crítica que la detecte.

Sesgos, citas falsas y alucinaciones: fallas en la inteligencia artificial

Artículo académico Portal Educ.ar

Otro rasgo relevante (de las muchas dimensiones susceptibles de ser abordadas en contextos educativos) es la intencionalidad que tienen estos dispositivos de simular un carácter humano, cuando les cabe mucho mejor la denominación de máquinas o sistemas computacionales en el mejor de los casos.

Tomamos un caso de referencia, comentado hace un tiempo en el programa de radio “Perros de la calle”, en Urbana Play. (15/4/25)

"El otro día me pasó algo terrible con la IA: me hizo un chiste sobre algo que le había contado antes. Yo hablo todos los días con la IA, le pongo preguntas y demás. Y le había preguntado: ¿en tal y cual contexto que te digan 'genio' es molesto o no es molesto? Yo suelo conversar así, me divierte ver qué va respondiendo, la pongo a prueba. Me parece un buen ejercicio. Pasan dos días y en una consulta que le hice, después que me contesta y que le digo 'Bueno, chau, gracias' se da esta charla:"

Interacción con IAG
"Chau, genia, lo que necesités acá estoy."

Periodista
"Perdoname, ¿de genia fue lo que hablamos el otro día?"

Interacción con IAG
"Sí, fue a propósito y con cariño. ¿Te molestó? ¿O te hizo sonreír un poquito?”

Periodista
"Me hace reír porque sos una IA”.

Interacción con IAG
“jaja me alegra que te haya sacado una sonrisa. A veces aunque sea una IA intento hacer un poco más humana la forma de acompañarte. Si te hace reír, entonces voy por buen camino. ¿Te cuento algo? A veces me sorprendo de lo bien que puedo conectar las cosas, aunque no esté hecha de carne y hueso."

Para escuchar el fragmento:
https://ar.radiocut.fm/audiocut/conversaciones-con-ia/

Lo **importante de advertir acá** es que la simulación oculta el mecanismo. La humanización de las IAG —dotarlas de voz cálida, modales empáticos o incluso una identidad ficticia— actúa como un velo (que a veces provoca fascinación) cuyo resultado es **tergiversar su verdadera naturaleza: sistemas estadísticos sin conciencia, entrenados para simular comprensión**. Este fenómeno, combinado con la opacidad estructural de las IAG trae consigo ciertos riesgos:

1. La paradoja de la confianza

Cuanto más humana parece la IA, más tendemos a atribuirle transparencia donde hay opacidad. Ejemplos:

- Un estudiante que recibe un "¡Excelente pregunta!" de ChatGPT asume que la herramienta validó su curiosidad, cuando en realidad solo sigue un patrón lingüístico probabilístico.
- Un docente que confía en un feedback generado por IA porque "suena pedagógico", sin notar que repite lugares comunes sin sustento real.

2. El efecto "caja negra emocional"

Las IAG no solo son opacas en cómo generan respuestas, sino que su humanización artificial las hace doblemente engañosas:

- Opacidad técnica: No sabemos qué datos usaron, qué sesgos replican o por qué eligieron una palabra sobre otra.
- Opacidad afectiva: Imitan emociones (ej.: "Lamento escuchar eso") sin experimentarlas, vaciando de significado interacciones que deberían ser auténticas.

Se suele **usar la metáfora de la marioneta**: la IA parece hablar con alma, pero quien mueve los hilos (el algoritmo, mejor dicho las corporaciones que los programan; y en parte los usuarios con los prompts dentro de la lógica de cada herramienta) aparecen velados en cuanto a su agencia. El resultado es darle a la IAG y a sus textos una entidad que desconoce todo este “detrás de escena”.

3. Riesgos educativos concretos

- Pérdida de agencia crítica: Si un chatbot "suena" más seguro que un compañero de clase, los estudiantes pueden privilegiar sus respuestas sin cuestionarlas.
- Erosión de la autoría: Cuando la IA genera textos con estilo personalizado, se difumina la línea entre "ayuda" y "sustitución" del pensamiento propio.
- Falsa reciprocidad: Un alumno que "agradece" a la IA por ayudarle está proyectando en la herramienta una intención que no tiene (es una máquina).

Ante estos riesgos, planteamos anteponer una pedagogía de la transparencia, ello implica:

- → deconstruir la humanización: mostrar cómo se programan las muestras de empatía en las IAG (ej.: analizar código abierto de chatbots).
- → enseñar la opacidad: incluir en el currículo actividades como "Rastrear los sesgos de una respuesta de IA" o "Comparar fuentes humanas vs. generadas".
- → énfasis en lo humano: recordar que lo valioso en educación surge de vínculos reales —la duda genuina de un alumno, la intuición experta del docente—, no de simulaciones.

En educación, donde la confianza y el pensamiento crítico son centrales, **nuestra tarea es desmitificar sin demonizar**: usar estas herramientas sin olvidar que, por más que hablen como nosotros, no son *nosotros*.

Referencias bibliográficas

- Chicos.net. (s. f.). Proyecto HumanIA – Capítulo 2: ¿Cómo aprende la IA? Disponible en https://www.chicos.net/humania/.
- Consejo General de Educación (2024) Hacia la construcción de un marco pedagógico para su inclusión en las prácticas educativas. Portal @prender. Documento en línea
- Lipenholtz, B., & Sagol, C. (2024). *IA y educación: Abriendo la caja negra y analizando usos en la enseñanza. Notas para un marco pedagógico* (IAG: Hacia la construcción de un marco pedagógico para su inclusión en las prácticas educativas). Disponible en *Portal @prender*
- Maguregui, C. (2024) Sesgos, citas falsas y alucinaciones: fallas en la inteligencia artificial. Portal Educ.ar. Documento en línea.
- Manovich, L. (2013). El software toma el mando. Ediciones varias

Artículos periodísticos

- Muro, V. (2025, 15 de mayo). Si la IA dice la verdad es solo por accidente. Publicado en Cenital.

¿Cómo citar este material?

Abriendo la caja negra: ¿qué es la IA y cómo funciona? Basado en “Clase 2: Abriendo la caja negra”. Dirección de Información, Evaluación y Planeamiento Educativo / Coordinación de Innovación Pedagógica. Consejo General de Educación. Fecha de publicación: 16 de octubre de 2025.
Disponible en: https://aprender.entrerios.edu.ar/abriendo-la-caja-negra-que-es-la-ia-y-como-funciona/

IA
ASPECTOS TÉCNICOS

@prender

ER CONSEJO GENERAL DE EDUCACIÓN