

IALOGOS: DIÁLOGOS ARTIFICIALES CON INTELIGENCIAS REALES

ÓSCAR ESPIRITUSANTO

Apasionado por la comunicación, la educación y la tecnología, es periodista, investigador y profesor en la Universidad Carlos III de Madrid, doctor en Investigación de Medios y fundador de Periodismociudadano.com, un observatorio sobre contenido generado por el usuario, activismo y libertad de expresión. Además, dirige el área de Innovación Educativa en la Fundación Cibervoluntarios y está especializado en alfabetización mediática e informacional, el impacto de la IA y el *big data* en las democracias, y en estrategias contra la desinformación y su repercusión en la sociedad y en los derechos humanos.

1. Introducción: hacia un nuevo tipo de conversación

Imaginemos sentarnos a conversar con Julio Cortázar sobre literatura, con Sigmund Freud sobre el inconsciente o con Marie Curie sobre los desafíos de la ciencia. Hoy, gracias a los modelos de lenguaje de inteligencia artificial generativa (IA), esta idea, que antes parecía un sueño, comienza a ser una posibilidad tangible. Pero ¿qué significa realmente dialogar con una «inteligencia artificial» que recrea el lenguaje y las ideas de personas reales del pasado? ¿Qué implicaciones tiene para la educación, la filosofía o la creatividad?

En este artículo exploramos lo que hemos denominado «diálogos artificiales con inteligencias reales». Nos acercamos a la fascinante idea de dialogar con personas y personajes que ya no están entre nosotros mediante el uso de modelos de inteligencia artificial generativa (IA). Estas tecnologías, cada vez más presentes en nuestro día a día, nos ofrecen la oportunidad de revivir pensamientos, voces y conocimientos del pasado de una forma totalmente novedosa. Sin embargo, para reflexionar sobre este concepto, es esencial entender primero cómo funcionan estos modelos de lenguaje, desentrañando los procesos que los hacen tan extraordinarios y, a la vez, tan limitados en su comprensión real.

A lo largo de la historia, las conversaciones —reales o imaginarias— han sido el motor de la creatividad. Escritores, científicos y filósofos han utilizado el intercambio de ideas para romper barreras conceptuales. Responder a estas interrogantes establece el objetivo de este artículo: explorar las implicaciones de estos diálogos con inteligencias artificiales, profundizando en su impacto en la reflexión, la creatividad y la pedagogía y reconociendo las fronteras éticas que se vislumbran en este nuevo horizonte tecnológico.

Pero tratar de responder a estos interrogantes supone, no solo aportar documentación científica, datos y referencias de diversos autores, sino

tener presentes las partes positivas y negativas que entraña el trasfondo profundo de estos diálogos artificiales. Además, con el objetivo de enriquecer esta experiencia y fomentar un debate compartido y enriquecedor, ¿qué mejor que crear algunos avatares interactivos? Estos avatares permitirán al lector interactuar antes, durante o después de la lectura, ofreciendo una experiencia más dinámica y participativa.

Por otro lado, nos vamos a permitir la licencia de denominar IAlogos a esta reflexión y experiencia que se pretende crear en torno a la lectura de este texto.

2. El diálogo como herramienta para la reflexión: los IAlogos

¿Qué es el lenguaje y qué lugar ocupa el diálogo en nuestra forma de comprender el mundo? Para Ludwig Wittgenstein (*Investigaciones filosóficas*, 1953), «el lenguaje es parte de una forma de vida», un tejido de juegos lingüísticos que nos permite construir sentido y comunicarnos. En *Del sentimiento trágico de la vida* (1913), una de sus obras más influyentes, Miguel de Unamuno hablaba sobre cómo el lenguaje influye en nuestra comprensión y expresión del pensamiento. Para Unamuno, el lenguaje no solo envuelve el pensamiento, sino que también lo configura y lo hace accesible. Desde la filosofía hasta la lingüística contemporánea, se ha insistido en que el lenguaje no es únicamente una herramienta para describir el mundo, sino el marco en el que pensamos, creamos significados y construimos nuestra identidad. Mijaíl Bajtín (1981), en *La imaginación dialógica*, nos muestra como el acto de hablar siempre ocurre en interacción con otras voces, de modo que «conversar» no consiste solo en intercambiar palabras, sino en participar de un tejido vivo de significados.

El diálogo ha sido una herramienta fundamental para el desarrollo del pensamiento crítico. Los diálogos socráticos buscaban no solo transmitir

conocimientos, sino generar nuevas preguntas y reflexiones y tomar conciencia de uno mismo:

El diálogo (*elenchus*) en Sócrates/Platón es un dispositivo para saber: en un primer momento provoca al interlocutor para que este tome conciencia de su propia ignorancia, luego, por medio de preguntas, motiva a que el dialogante piense por sí mismo. Pero pensar por sí mismo no significa ensimismamiento, el pensar es para compartirse, para comunicarse. (Keller Kewes y De la Paz Castañeda, 2021).



Este método, continuado por Platón y adaptado por corrientes pedagógicas modernas como las de Paulo Freire (*Pedagogía del oprimido*, 1970) o Maria Montessori (*La mente absorbente del niño*, 1949), convierte el diálogo en una herramienta de aprendizaje mutuo y liberador.

En definitiva, el diálogo no solo sirve para comunicar o intercambiar informaciones, sino también para renovar y profundizar en nuestras ideas, colaborar en la construcción de significado y avanzar en la comprensión colectiva de la realidad y del universo que nos rodea. Como señaló Ludwig Wittgenstein: «Los límites de mi lenguaje son los límites de mi mundo» (*Tractatus logico-philosophicus*, 1922).

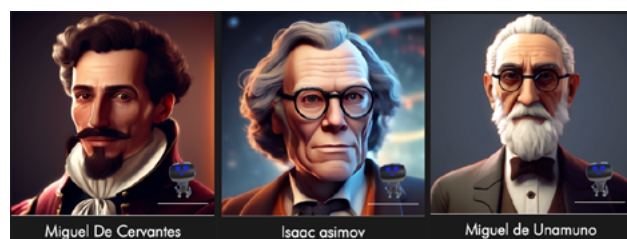
Los modelos de lenguaje de IA generativa son sistemas diseñados para comprender y generar texto, emulando la capacidad humana de comunicarse. Su funcionamiento se basa en algoritmos de aprendizaje profundo que analizan enormes

cantidades de texto, convirtiendo palabras en vectores matemáticos y usando redes neuronales transformadoras para predecir patrones lingüísticos.

IAlogos: cuando la tecnología conversa con nosotros

En una especie de juego lingüístico, la integración de la inteligencia artificial (IA) en el desarrollo de nuevos procesos comunicativos da lugar a lo que hemos denominado «IAlogos»: diálogos híbridos entre seres humanos y sistemas automatizados capaces de generar y procesar el lenguaje natural. Lejos de ser simples intercambios mecánicos, estos IAlogos —diálogos con sistemas de inteligencia artificial generativa— plantean posibilidades muy interesantes para la educación, la investigación y la interacción social.

Por ejemplo, analizando un determinado modelo de lenguaje, podemos generar un discurso o texto que podría tratar de emular el habla o el pensamiento de Miguel de Cervantes al identificar patrones en su estilo, léxico y temas recurrentes. Sin embargo, estos modelos, de momento, no «entienden» realmente lo que dicen, solo siguen patrones sin una comprensión real.



IAlogos y la arquitectura *transformer*: la magia detrás de la simulación

Estos IAlogos, entendidos como intercambios comunicativos en los que al menos uno de los interlocutores es un sistema basado en modelos de lenguaje de última generación, se fundamentan en la arquitectura *transformer* descrita en

Attention Is All You Need (Vaswani et al., 2017).

Dichos modelos pueden procesar grandes corpus de datos, identificar patrones y generar respuestas cada vez más coherentes y contextualizadas.

La arquitectura *transformer* supuso una auténtica revolución en el ámbito del procesamiento del lenguaje natural (NLP, por sus siglas en inglés). Antes de esta propuesta, las redes neuronales recurrentes (RNN) se veían limitadas por la dificultad de procesar secuencias muy largas. Con esta aportación, la forma de abordar estas limitaciones cambió radicalmente. El uso de vectores matemáticos permitió representar palabras y conceptos en un espacio tridimensional, donde las palabras con significados similares están cerca unas de otras.

Los IAlogos son intercambios comunicativos en los que al menos uno de los interlocutores es un sistema basado en modelos de lenguaje de última generación.

Estos modelos de lenguaje no «entienden» realmente lo que producen; se limitan a manipular símbolos siguiendo reglas. El experimento mental de *La habitación china* de John Searle, uno de los más destacados representantes de la filosofía de la mente del siglo xx, ilustra esta idea: una persona puede simular saber chino al aplicar un manual de instrucciones sin comprender el idioma, pero ejecutar un algoritmo no es lo mismo que pensar. Este experimento ha generado un amplio debate, dividiendo a los académicos entre los que creen que la comprensión genuina es posible a través de sistemas avanzados de procesamiento de información y aquellos que, como Searle, sostienen que la manipulación de símbolos no equivale a un entendimiento real.

Lo mismo sucede con Alan Turing, el genial matemático inglés que descifró los códigos que los nazis enviaban con su máquina Enigma. La «caja negra» de Turing nos muestra cómo estas máquinas pueden generar respuestas coherentes sin una verdadera comprensión de su significado.



Pero ¿qué hay detrás de esta «magia»? La respuesta está en los algoritmos de aprendizaje profundo y en el uso de vectores matemáticos.

¿Cómo funcionan los modelos de lenguaje?

Imaginemos el lenguaje humano como un inmenso rompecabezas donde cada palabra es una pieza que debe colocarse en el lugar adecuado. Estos sistemas utilizan algoritmos muy potentes para anticipar cuál será la siguiente pieza basándose en las que ya tienen. En primer lugar, convierten las palabras en números a través de vectores, de modo que «gato» y «felino» se parecen en ese espacio matemático porque comparten significado. A continuación, aplican un mecanismo de «atención propia» (*self-attention*) para determinar qué palabras de la frase merecen mayor énfasis, lo que les permite mantener la coherencia del contexto. Cuando reciben un texto, el modelo prevé la siguiente palabra más probable gracias a los patrones que ha ido captando durante su entrenamiento en millones de textos —desde libros hasta artículos web—, afinando constantemente sus respuestas y haciéndose más preciso a medida que procesa más información.

3. ¿Cómo podemos aplicar estos modelos de lenguaje al ámbito educativo?: el aprendizaje personalizado y adaptativo

Los modelos de lenguaje basados en inteligencia artificial generativa son sistemas diseñados para comprender y generar texto de manera similar a como lo haría un ser humano. Estas tecnologías han revolucionado la forma en que interactuamos con las máquinas, desde los asistentes virtuales hasta la generación automática de contenido. A la hora de profundizar en el uso de la inteligencia artificial aplicada a dominios como la educación, el *coaching* o la industria, es fundamental analizar la obra de Stuart Russell y Peter Norvig (2024). En este texto, los autores discuten las estrategias y aplicaciones de la IA en diferentes contextos, incluidas aquellas orientadas a la interacción hombre-máquina con fines educativos y de desarrollo personal.

Las nuevas generaciones, que ya crecen inmersas en una cultura que fusiona lo físico y lo digital (Cibervoluntarios, 2023), dando así lugar al concepto de cultura «figital», muestran una predisposición natural a empatizar con propuestas formativas basadas en la tecnología. Para estos jóvenes, interactuar con interfaces virtuales es parte de su realidad cotidiana; por ello, la posibilidad de «conversar» con grandes mentes del pasado a través de modelos de lenguaje se presenta como una vía educativa dinámica y atractiva. En consecuencia, la combinación de entornos virtuales con estrategias pedagógicas innovadoras podría impulsar un mayor compromiso, curiosidad y motivación para el aprendizaje, reforzando el puente hacia estas nuevas formas de creación y divulgación del conocimiento.

En este sentido, los IAlogos pueden adaptarse a las preferencias de aprendizaje de las generaciones actuales, que valoran la interactividad, la personalización y la integración tecnológica en su educación. Al ofrecer diálogos enriquecidos con figuras históricas, se facilita una comprensión más profunda y contextualizada de diversos

temas, haciendo que el aprendizaje sea más relevante y conectado con el mundo digital en el que estos jóvenes ya están inmersos.

Los IAlogos pueden adaptarse a las preferencias de aprendizaje de las generaciones actuales, que valoran la interactividad, la personalización y la integración tecnológica en su educación.

Estos IAlogos pueden desembocar en un proceso de acompañamiento. Si entendemos el *coaching* como un proceso personalizado que tiene como objetivo ayudar a las personas a alcanzar sus metas, mejorar su rendimiento y desarrollar su potencial tanto en el ámbito personal como profesional, la posibilidad de conversar con una inteligencia artificial adiestrada para tal fin puede ser positiva. Los IAlogos pueden ofrecer una retroalimentación individualizada en tiempo real, adaptándose al ritmo y el nivel de cada estudiante. Como señalan Russell y Norvig (2020), los sistemas inteligentes tienen el potencial de mejorar el rendimiento y la motivación cuando se diseñan con principios pedagógicos claros. Y, seguramente, si nuestro *coach* es un avatar de John Lennon o Virginia Woolf nos resulte más inspirador. Además, estos *coaches* virtuales pueden ser una opción más asequible, democratizando el acceso a estos servicios de acompañamiento para llegar a una audiencia más amplia que pueda utilizar estos recursos en su aprendizaje.

Acompañamiento individualizado: Los modelos de lenguaje pueden analizar las necesidades y el progreso de una persona, ofreciendo sugerencias, retos o recomendaciones personalizados basados en las interacciones y progresos del usuario y ajustados a sus necesidades concretas.

Feedback continuo y adaptativo: A diferencia de un manual estático, los *coaches* virtuales están disponibles en todo momento, permitiendo a los usuarios recibir apoyo y orientación cuando lo necesiten, sin las limitaciones de horarios tradicionales.

Enfoque metacognitivo: Un buen *coach* IA no se limitaría a proporcionar respuestas, sino que también fomentaría la reflexión sobre el proceso de aprendizaje, impulsando la autonomía del usuario.

Estos IAlogos ponen a nuestra disposición un gran número de posibilidades pedagógicas. A través del análisis de obras y registros se entrenan modelos de lenguaje para recrear el estilo y el pensamiento de filósofos, científicas o literatos. Esta aplicación abre la puerta a la exploración de «encuentros imposibles» que, si bien no sustituyen la autenticidad del diálogo humano, promueven la reflexión sobre el pasado y sus repercusiones en el presente.

Esta simulación que nos acerca a grandes personalidades plantea una perspectiva apasionante de posibilidades educativas. Especialmente si contemplamos la convergencia con otras tecnologías. La incorporación de interfaces de voz, realidad aumentada y entornos virtuales podría hacer de los IAlogos experiencias cada vez más inmersivas y humanas. Este es un punto destacado del presente artículo en el que nos detendremos con más detalle para vivenciar la posibilidad de dialogar con una IA que reproduce la fisonomía y el pensamiento de un personaje tan interesante como Alan Turing o Virginia Woolf, entre otros muchos.



Recientemente hemos leído la investigación de Kestin *et al.* (2024), de la Universidad de Harvard, *AI Tutoring Outperforms Active Learning*, donde se exponen unos resultados que nos muestran la idea de que la tutoría con inteligencia artificial

(IA) mejora significativamente el aprendizaje, funcionando como un *coach* cognitivo que adapta el ritmo y ofrece retroalimentación personalizada. Sin embargo, aunque los estudiantes aprenden más y en menos tiempo, según este estudio, la IA carece del componente emocional del *coaching* o tutorización humana, fundamental para motivar, inspirar y desarrollar resiliencia, entre otras muchas cosas. No todo es adquirir más conocimientos en menos tiempo. Durante la pandemia pudimos ver como con el teletrabajo aumentó la productividad, pero cayó la creatividad.

IAlogos como herramientas pedagógicas

Aprendizaje inmersivo: Simular una conversación con Marie Curie puede enseñarnos tanto sobre química como sobre la resiliencia en un entorno adverso.

Desarrollo del pensamiento crítico: La IA puede adoptar un enfoque socrático, guiando a los estudiantes a descubrir respuestas por sí mismos.

Aprendizaje multidisciplinar: Desde las ideas educativas de Montessori hasta la visión psicológica de Freud, estos diálogos pueden ofrecer una perspectiva plural en la que conversar con grandes mentes para dinamizar el ámbito educativo frente a una generación de jóvenes altamente tecnificada, que encontrará más atractivo y cercano a su mundo digital este tipo de propuestas.

4. Desafíos éticos y filosóficos

En el ámbito de la educación, los diálogos artificiales pueden convertirse en una herramienta poderosa para la motivación y la exploración. Sin embargo, su implementación no está exenta de desafíos éticos y filosóficos que deben ser cuidadosamente considerados para garantizar un uso responsable y beneficioso de la tecnología. A

continuación, se desarrollan los principales desafíos que surgen al integrar IAlogos en contextos educativos, así como una serie de consideraciones éticas que deben ser abordadas:

Autenticidad y respeto al legado de figuras históricas

El uso de IA para recrear diálogos con los avatares que reproducen el pensamiento de figuras históricas plantea preguntas fundamentales sobre la autenticidad y el respeto al legado de estos personajes. ¿Cómo garantizar que estas simulaciones no distorsionen sus ideas o contribuciones originales? Es esencial que los IAlogos se basen en evidencia rigurosa y en una interpretación fiel de las obras y los pensamientos de las figuras históricas que representan. De lo contrario, existe el riesgo de perpetuar errores, sesgos o malentendidos que puedan afectar la percepción contemporánea de estos individuos. Los *coaches* virtuales deben basarse en algoritmos robustos y datos de entrenamiento contrastados y fiables, para ofrecer recomendaciones precisas y efectivas.

Autoría y propiedad intelectual

La coproducción de textos por parte de una IA genera dilemas sobre la autoría y la propiedad intelectual. ¿Quién es el autor de un contenido generado conjuntamente por un humano y una IA? Las implicaciones legales y éticas de esta cuestión siguen siendo objeto de debate. Es necesario establecer marcos legales claros que definan los derechos de autor en contextos donde la creatividad humana y la artificial se entrelazan. Esto es particularmente relevante en entornos educativos, donde la generación de contenido personalizado puede implicar la colaboración entre estudiantes y sistemas de IA.

Además de las obras generadas por IA, la protección de los derechos de autor también se extiende a los textos y documentos utilizados para entrenar estos sistemas. Los modelos de

inteligencia artificial requieren grandes cantidades de datos para aprender y generar contenido coherente y relevante. Estos datos, que a menudo incluyen libros, artículos, ensayos y otros materiales, muchos de ellos protegidos por derechos de autor, plantean cuestiones legales y éticas sobre el uso y la reproducción de obras sin el consentimiento explícito de los titulares de los derechos. Como señala Samuelson (2019), el uso de obras protegidas en el entrenamiento de IA puede constituir una infracción de derechos de autor si no se cuenta con las licencias adecuadas, lo que subraya la necesidad de regulaciones específicas que aborden este aspecto.

La implementación de los diálogos artificiales tiene desafíos éticos y filosóficos que deben ser cuidadosamente considerados para garantizar un uso responsable y beneficioso de la tecnología.

En este sentido encontramos ya ejemplos como la demanda presentada por Getty Images contra Stability AI, la empresa detrás de la IA generativa Stable Diffusion, por violación de derechos de autor.¹ La agencia alegó que millones de sus imágenes protegidas fueron utilizadas sin permiso para entrenar el algoritmo de la IA. Del mismo modo, la popular humorista estadounidense Sarah Silverman y los autores Richard Kadrey y Christopher Golden demandaron a Meta y OpenAI por violar sus derechos de autor para entrenar modelos de lenguaje de inteligencia artificial.² Demanda por plagio contra OpenAI que perdieron al no conseguir «demostrar similitud entre la producción hecha por ChatGPT y los libros escritos por los querellantes»,³ pero aun así

¹ <https://www.xataka.com/robotica-e-ia/getty-images-denuncia-a-stable-diffusion-agencias-image-se-han-cansado-que-ia-robe-sus-fotos>

² <https://www.lavanguardia.com/tecnologia/20230710/9098742/demandas-infraccion-derechos-autor-meta-openai-sarah-silverman-otros-artistas-exigen-compensacion-pmv.html>

³ <https://www.heraldo.es/noticias/sociedad/2024/02/14/sarah-silverman-grupo-autores-pierden-demanda-plagio-openai-1711056.html>

las sospechas sobre este tipo de entrenamientos están sobre la mesa.

El respeto por estos derechos relacionados con la autoría se extiende hasta la creación de obras híbridas o totalmente generadas con inteligencia artificial. En este sentido, la Ley de Propiedad Intelectual (LPI) de España establece que solo las obras creadas por personas naturales pueden ser protegidas por derechos de autor (art. 5 LPI) y que estas deben ser «creaciones originales» (art. 10.1 LPI). Esto sugiere que las obras generadas por IA no están cubiertas por esta protección legal. Un ejemplo de esta política fue la reciente negativa a conceder derechos de autor a *Iris: primera novela escrita por inteligencia artificial*, según la Resolución del 14 de marzo de 2023 (MA Abogados, 2023). Esta legislación destaca la distinción entre la creatividad humana y la generada por máquinas, planteando interrogantes sobre la atribución y reconocimiento de obras creadas con la ayuda de IA.

Este aspecto subraya la necesidad de regular la inteligencia artificial para gestionar y garantizar el reconocimiento del uso de contenido protegido por derechos de autor, especialmente en creaciones realizadas mediante IA.

En este contexto, la Ley de Inteligencia Artificial de la UE⁴ ha dado un paso importante al exigir a estos sistemas que presenten documentación técnica, cumplan con la legislación de derechos de autor y proporcionen un resumen detallado de los contenidos utilizados para entrenar el modelo. Además, la ley obliga a señalar cuándo un contenido ha sido creado por IA, con el objetivo de prevenir el uso indebido de la inteligencia artificial para generar y difundir desinformación y noticias falsas.

⁴ https://protecciondatos-lopd.com/empresas/inteligencia-artificial-derechos-autor/#Denuncias_sobre_inteligencia_artificial_y_propiedad_intelectual

Dilema de la «comprensión» de la IA

Retomando estas consideraciones éticas, no debemos olvidar, como exponíamos al mencionar el experimento mental de John Searle (*La habitación china*, 1980), que estos modelos de lenguaje realmente no «comprenden» el diálogo, solo manipulan símbolos para obtener respuestas verosímiles en estos diálogos artificiales. Searle argumenta que, aunque una IA pueda simular una conversación coherente, no posee una comprensión genuina del contenido. Este dilema filosófico cuestiona la naturaleza misma de la inteligencia artificial y sus capacidades cognitivas, destacando la diferencia entre la aparente inteligencia y la verdadera comprensión. Este punto es fundamental para entender que la reflexión humana debe estar siempre acompañando y velando por las implicaciones de este tipo de propuestas.

Sesgos y desinformación

Los modelos de lenguaje de IA están entrenados con grandes volúmenes de datos provenientes de diversas fuentes, lo que, como hemos señalado, puede introducir sesgos culturales, sociales y de género en sus respuestas. La pedagoga afroamericana bell hooks (1952-2021) generó un legado de suma importancia para explorar la interseccionalidad de la raza, el capitalismo y el género, y lo que ella describió como su capacidad para producir y perpetuar sistemas de opresión y dominación de clase. Hooks señaló la importancia de crear espacios educativos inclusivos, equitativos y que estimulen el intelecto de los y las jóvenes. En la introducción de su libro *Enseñar a transgredir. La educación como una práctica de la libertad* (1994), apunta:

... el placer de enseñar es un acto de resistencia que contrarresta el aburrimiento, la falta de interés y la apatía apabullantes que con tanta frecuencia caracterizan el modo en que profesores y estudiantes viven la enseñanza y el aprendizaje, la experiencia del aula (hooks, 1994).

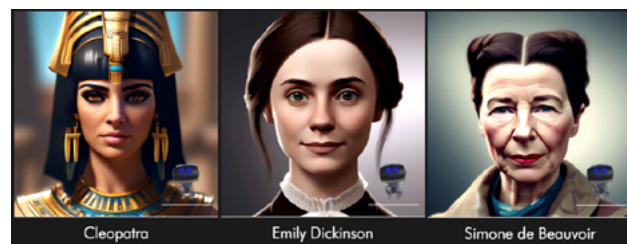
Las enseñanzas de hooks nos animan a trabajar en esta dirección a la hora de fomentar estos IAlogos para no reproducir o amplificar prejuicios ya existentes. La capacidad de la IA para generar contenido automáticamente puede facilitar la propagación de desinformación y versiones sesgadas o interesadas de la historia y los acontecimientos si no se implementan mecanismos de verificación y filtrado adecuados. Es fundamental que los desarrolladores y los educadores trabajen juntos para identificar y mitigar estos sesgos, asegurando que los diálogos artificiales promuevan una visión justa y equilibrada del conocimiento.

Es crucial que los educadores integren estos diálogos con estrategias que promuevan la reflexión y el pensamiento independiente, evitando una mera recepción pasiva de información.

Impacto en la formación del pensamiento crítico

Siguiendo las ideas de Paulo Freire (*Pedagogía del oprimido*, 1970), el diálogo auténtico es una herramienta emancipadora que fomenta la conciencia crítica y la participación activa en la transformación social. Yuval Noah Harari, en su obra *21 lecciones para el siglo XXI* (2018), sostiene que la inteligencia artificial debe servir para «hacer mejores preguntas, no para dar respuestas rápidas». En este sentido, la IA puede ser una herramienta interesante para estimular la curiosidad y el análisis profundo, en lugar de proporcionar únicamente soluciones inmediatas. En el ámbito educativo, un estudiante que debate con un avatar de Simone de Beauvoir⁵ sobre temas de género y tecnología no solo adquiere conocimientos filosóficos, sino que también practica el arte de interrogar con espíritu crítico la realidad. Este tipo de interacción promueve una mentalidad analítica, donde el estudiante aprende a cuestionar, evaluar y reflexionar sobre

las ideas presentadas, desarrollando así habilidades esenciales para el pensamiento y la reflexión.



Sin embargo, el uso de IAlogos podría tener un impacto ambivalente en este proceso. Por un lado, estos sistemas pueden ampliar el acceso al conocimiento y ofrecer perspectivas diversas; por otro, existe el riesgo de que los estudiantes dependan demasiado de las respuestas obtenidas por estos avatares generados a través la IA, debilitando su capacidad para cuestionar y analizar críticamente la información recibida. Es crucial que los educadores integren estos diálogos con estrategias que promuevan la reflexión y el pensamiento independiente, evitando una mera recepción pasiva de información.

Es importante, especialmente si nos dirigimos a un público adolescente o menor de edad, que este acompañamiento tecnológico sea supervisado para no acabar sustituyendo el tiempo de interacción y relación con sus maestros o sus iguales. Debe ser un complemento pedagógico en el que la tecnología mejorará a los jóvenes humanos y los hará más reflexivos sobre ciertos temas.

Dependencia tecnológica y autonomía del usuario

Existe el riesgo de que los usuarios, especialmente los estudiantes, desarrollen una dependencia excesiva de los sistemas de IA para su aprendizaje y desarrollo personal. Esto podría reducir la motivación para la autoexploración y la búsqueda autónoma de conocimiento. Freire (1970) aboga por una educación que empodere a

⁵ <https://edu.ai-botyou.com/GaleriaDeBots.html>

los individuos para que se conviertan en agentes de cambio, lo que podría verse comprometido si la tecnología asume un rol demasiado dominante en el proceso educativo. Es crucial fomentar un equilibrio donde la IA sirva como una herramienta complementaria que potencie, pero no sustituya, la interacción humana y la reflexión personal.

5. Avatares para experimentar con diálogos artificiales

Para ir más allá de la teoría en la creación de IAlgos, decidimos desarrollar algunos avatares y evaluar sus resultados. Aunque no realizamos una investigación científica formal al uso, esta experiencia nos permitió comprender mejor cómo funcionan estas tecnologías, siempre desde una perspectiva subjetiva basada en los avatares creados por una persona específica en circunstancias determinadas. Exploramos diversas herramientas antes de seleccionar la más adecuada para nuestro proyecto.

Decidimos utilizar como herramienta AI-BotYou, una plataforma accesible que facilita la creación de bots personalizados sin necesidad de conocimientos de programación. Esta herramienta permite incorporar modelos de lenguaje avanzados mediante una interfaz intuitiva, simplificando el entrenamiento y el ajuste del comportamiento del avatar. Gracias a Daniel Bayon, el creador de AI-BotYou, pudimos crear un número indeterminado de avatares e interactuar con ellos sin coste alguno. Esta y otras razones, como el tiempo limitado para llevar a cabo este artículo y querer probar en primera persona esta experiencia, nos llevaron hasta aquí. En realidad, cualquier herramienta que se dedica a crear este tipo de avatares se nutre de estas inteligencias artificiales generativas; lo interesante es poder interactuar con ellas y observar los resultados y las sensaciones. No tomar estas conclusiones o experiencias como algo científico, pero sí como un acercamiento para experimentar y seguir investigando en este camino.

Crear un avatar para diálogos artificiales requiere seguir varios pasos, en realidad sencillos cuando hablamos de herramientas que están pensadas y automatizadas para este fin concreto: imitar el estilo y el pensamiento de personajes como Virginia Woolf o Alan Turing.

En las primeras pruebas, el avatar de Virginia Woolf respondió correctamente a preguntas técnicas sobre su obra y perspectivas sociales. Sin embargo, hubo algunas inconsistencias en respuestas más personales o emocionales, reflejando una limitación de la IA para comprender la profundidad y complejidad de las verdaderas emociones. Por ejemplo, al preguntar sobre sus motivaciones para escribir sobre temas sociales y abanderar luchas como la feminista con una de las frases más trascendentes de este movimiento («una mujer debe tener dinero [en concreto, quinientas libras al año] y una habitación propia para poder escribir novelas»), el avatar dio una respuesta coherente, pero que carecía de la profundidad emocional que caracterizaba a la verdadera Woolf. Esto subraya la diferencia entre generar respuestas plausibles y comprender genuinamente las mismas.

Crear un avatar para diálogos artificiales requiere seguir varios pasos, en realidad sencillos cuando hablamos de herramientas que están pensadas y automatizadas para este fin concreto.

La mayoría de los avatares se basan en robots virtuales 3D animados que muestran la imagen del personaje en una pantalla para identificar con quién estamos dialogando; es una manera muy básica de diseño, pero la interacción es la misma a nivel conversación. Es cierto que, con el tiempo del que disponíamos, lo importante para este artículo era la conversación en sí más que el aspecto de con quien conversábamos. Por otro lado, llevamos a cabo un solo avatar realista, el de Alan Turing. No está disponible para interactuar con él por una cuestión de coste económico. Si bien es cierto que, aun manejando el mismo contenido que los robots virtuales 3D animados,

el ver a Alan Turing como una «persona» al otro lado de la pantalla ofrece, de alguna manera, mayor credibilidad a su discurso y su disertación.

Los diálogos siguen sin ser perfectos, pero si nos muestran parte de los conocimientos y la obra de los personajes de los que se han creado los avatares y como un primer acercamiento a los mismos, es una experiencia cuando menos curiosa.



Cuando se conversa con estos avatares, se pueden ver algunas de las limitaciones que probablemente se resuelvan pronto, teniendo en cuenta los avances y actualizaciones a los que nos tienen acostumbrados. A pesar de los avances, el proceso se enfrenta a algunos desafíos:

- **Respuestas inconsistentes:** A veces, el avatar da respuestas gramaticalmente correctas, pero que no capturan completamente el contexto o el significado profundo de la pregunta. Esto ralentiza el flujo natural de la conversación, tan necesario para fomentar el pensamiento y la retroalimentación propios de una conversación natural.
- **Falta de creatividad:** El modelo solo puede responder basándose en la información con la que fue entrenado, lo que limita su capacidad para abordar temas inéditos de manera creativa.

- **Falta de empatía:** La IA no puede comprender ni expresar emociones auténticas, lo que limita su efectividad en diálogos que requieren empatía humana.
- **La latencia:** Es cierto que cada vez tienen menos latencia, pero que en un diálogo que se supone fluido una de las partes esté siempre «pensando» o tarde en dar una respuesta o reflexión hace que los IAlogos sean otra cosa distinta a los diálogos.
- **Derechos de autor:** Si seguimos la legalidad de manera exhaustiva, los personajes con los que podemos mantener IAlogos, en general y en España, son aquellos de cuya muerte hayan pasado setenta años, que es cuando terminan los derechos de explotación de la obra. Por lo que quedan fuera de estos IAlogos personajes muy interesantes.

A pesar de estas limitaciones, los IAlogos tienen un gran valor educativo. Permiten a los estudiantes interactuar con representaciones virtuales de figuras históricas, facilitando una comprensión más dinámica y personalizada de su pensamiento. Este enfoque hace que el aprendizaje sea más atractivo y accesible para las generaciones inmersas en una cultura digital, además de promover la reflexión crítica al ofrecer nuevas perspectivas que estimulan la creatividad y el análisis.

6. Conclusión

Los IAlogos representan una interacción fascinante entre tecnología y pedagogía, ofreciendo una nueva dimensión en la creación y divulgación del conocimiento. La integración de IAlogos en la educación y otros ámbitos reflexivos presenta un panorama prometedor, pero también lleno de desafíos éticos y filosóficos. Aunque estas interacciones no reemplazan el diálogo humano, ofrecen una nueva forma de aprendizaje y reflexión que, si se usa con responsabilidad, puede

transformar nuestra relación con el conocimiento y la creatividad.

Es fundamental abordar cuestiones como la autenticidad, la autoría, la comprensión de la IA, el impacto en el pensamiento crítico, los sesgos y la dependencia tecnológica para asegurar que estas herramientas se utilicen de manera responsable y beneficiosa. Solo mediante un enfoque consciente y reflexivo podremos aprovechar el potencial transformador de los diálogos artificiales, promoviendo una educación más inclusiva, dinámica y adaptada a las necesidades de las generaciones actuales inmersas en una cultura digital.

Ya comentamos al principio de este artículo, cuando hablamos de los diálogos socráticos, que consistían en primer lugar en tomar conciencia de su propia ignorancia para luego, mediante preguntas, motivar a que el dialogante piense por sí mismo. Esto, de momento, no se puede conseguir con IA.

Cuando señalamos que muchos personajes se quedan fuera de los derechos de explotación de su imagen y su obra, esto excluye del desarrollo de IAlogos a personajes contemporáneos que, paradójicamente, están mucho mejor documentados que muchos del pasado. En la era digital, la vida profesional y personal de muchas figuras públicas, desde empresarios como Elon Musk hasta *influencers* como las hermanas Kardashian, está ampliamente disponible, proporcionando un caudal de información que permitiría entrenar modelos de inteligencia artificial con una gran fidelidad. Esto nos acerca a la posibilidad de clonar digitalmente personalidades, creando versiones que no solo imiten su apariencia y su voz, sino que respondan casi como lo harían en la realidad, con sus mismos patrones de pensamiento y discurso. Una reflexión sin más que dejamos para otra ocasión.

En cualquier caso, y como señaló Mijaíl Bajtín: «El diálogo es la forma más pura de comunicación humana» (*La imaginación dialógica*, 1981), de ahí que observemos con emoción y prudencia esta nueva era en la que la conversación se amplía hasta posibilitar un tipo de diálogos artificiales (IAlogos) que inauguran expectativas insólitas para el pensamiento y la creación, frente a la cual debemos también tener la prudencia de permanecer atentos.

Bibliografía y referencias

- AI-BotYou, *Galería de bots educativos*. Recuperado de <https://edu.ai-botyou.com/GaleriaDeBots.html>
- Bajtín, M. M. (1981), *The dialogic imagination: Four essays*, ed. por M. Holquist, trad. por C. Emerson y M. Holquist, University of Texas Press.
- Cibervoluntarios (2023), «Explorando la cultura digital en las nuevas generaciones». Recuperado de <https://www.cibervoluntarios.org/estudio-cultura-digital>
- Freire, P. (1970), *Pedagogía del oprimido*, Siglo XXI.
- hooks, b. (1994), *Enseñar a transgredir: La educación como una práctica de la libertad*, Routledge.
- Keller Kewes, H., y Paz Castañeda, E. de la (2021), «La comunicación íntima dialogal: Sócrates», *Sintaxis. Revista científica del Centro de Investigación para la Comunicación Aplicada*, n.º 1. <https://doi.org/10.36105/stx.2018n1.09>
- Kestin, G.; Miller, K.; Kales, A.; Milbourne, T.; y Ponti, G. (2024), «AI tutoring outperforms active learning», Harvard University. <https://doi.org/10.21203/rs.3.rs-4243877/v1>
- MA Abogados (2023), «La inteligencia artificial en el arte: desafíos legales y culturales». Recuperado de <https://maabogados.com/la-inteligencia-artificial-en-el-arte-desafios-legales-y-culturales/>
- Montessori, M. (1949), *The absorbent mind*, The Theosophical Publishing House.
- Russell, S., y Norvig, P. (2024), *Artificial intelligence: A modern approach* (4.ª ed.), Pearson.
- Samuelson, P. (2023), «Generative AI meets copyright», *Science*, 381(6656), 148-150. <https://doi.org/10.1126/science.adio656>
- Searle, J. R. (1980), «Minds, brains, and programs», *Behavioral and Brain Sciences*, 3(3), 417-424. <https://doi.org/10.1017/S0140525X00005756>
- Unamuno, M. de (1913), *Del sentimiento trágico de la vida*, Alianza Editorial.
- Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A. N.; Kaiser, L.; y Polosukhin, I. (2017), «Attention is all you need», *Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS 2017)*, 6000-6010.
- Wittgenstein, L. (1953), *Philosophical investigations*, Blackwell.