

Diez preguntas frecuentes y urgentes sobre Inteligencia Artificial

Marcos J. Gómez

Universidad Nacional de Córdoba,
Fundación Sadosky

mgomez@fundacionsadosky.org.ar

Julián Dabbah

Fundación Sadosky

jdabbah@fundacionsadosky.org.ar

Pablo E. 'Fidel' Martínez López

Universidad Nacional de Hurlingham¹
Universidad Nacional de Quilmes,
Fundación Sadosky

fidel@fundacionsadosky.org.ar

Mara Borchardt

Fundación Sadosky

mborchardt@fundacionsadosky.org.ar

Resumen

A partir de la publicación de ChatGPT-3 cobraron protagonismo en la vida cotidiana, los medios de comunicación, la industria y la academia las soluciones basadas en Inteligencia Artificial (IA). Resultaron tan exitosas que mucha gente empezó a realizar preguntas sobre esta tecnología. La Inteligencia Artificial es, para la mayoría de las personas, una caja negra en la que se deposita la esperanza de que permitirá solucionar problemas que hoy no se pueden resolver con facilidad y que esto sucederá en el corto plazo. Esta idea es alentada por los medios de comunicación y la publicidad de las empresas que desarrollan e implementan soluciones basadas en IA.

Este documento se propone responder 10 preguntas en torno a la IA para explicar cuestiones clave vinculadas a sus características y poner en tensión la posición antes descrita brindando información sobre las formas en que se produce esta tecnología, algunos de los problemas más salientes que conllevan las soluciones basadas en IA y las implicaciones éticas de su desarrollo y uso, abordando el impacto en el ambiente, entre otras temáticas urgentes. El propósito es contribuir a la democratización del conocimiento especializado para que cada vez más personas puedan reflexionar informada y críticamente sobre el impacto de la IA. En relación a las preguntas suscitadas por su auge en el sistema educativo, se aporta una mirada desde la perspectiva de los saberes a enseñar sin abordar recomendaciones sobre su uso específico para apoyar tareas de gestión o pedagógicas.

La elaboración de este documento se realizó con la colaboración de los siguientes especialistas, con quienes mantuvimos comunicaciones personales en referencia a este tema: la Dra. Luciana Benotti (UNC), la Dra. Cecilia Martínez (UNC), el Dr. Nicolás Wolovick (UNC), el Dr. Juan Echagüe (empresa Practia Global), el Lic. Gabriel A. Baum (UNLP, UNQ, UNRC y varias otras universidades nacionales), la Dra. Manuela Cerdeiro (Facultad de Ciencias Exactas y Naturales de

¹ Universidad Nacional de Hurlingham (UNAHUR), CIDIA: Centro de Investigación y Desarrollo en Informática Aplicada. Hurlingham, Argentina.

la UBA y miembro de la Fundación Sadosky) y la Dra. Victoria Dumas (directora del programa de Ciencia de Datos e Inteligencia Artificial de la Fundación Sadosky).

Las respuestas a las **10 preguntas** elegidas no buscan dar por cerrado el tema. Se profundizan y muchas veces se complejizan los asuntos abordados presentando nuevas aristas que promuevan seguir investigando, pensando y discutiendo sobre aquello que aún es incierto.

Índice

Resumen.....	1
Índice.....	2
Introducción.....	3
Pregunta 1: ¿Qué es la Inteligencia Artificial?.....	5
Pregunta 2: ¿Cuándo surgió la IA?.....	7
Pregunta 3: ¿Cómo se construyen las aplicaciones basadas en IA?.....	9
Pregunta 4: ¿Cuánto intervienen las personas en la creación de sistemas de IA?.....	11
Pregunta 5: ¿Cuál es el verdadero costo de la IA ?.....	12
Pregunta 6: ¿Los sistemas de IA razonan?.....	16
Pregunta 7: ¿Es correcto afirmar que la IA no se equivoca nunca y que sus resultados son éticos o justos?.....	20
Pregunta 8: ¿Afecta la diversidad cultural a la capacidad de las computadoras de generar contenido digital?.....	24
Pregunta 9: ¿Es recomendable usar soluciones basadas en IA generativa para enseñar a programar?.....	30
Pregunta 10: ¿Qué debe hacer la escuela ante el auge de la IA?.....	32
Conclusiones.....	36
Bibliografía.....	37

Introducción

Las aplicaciones basadas en Inteligencia Artificial (IA) forman parte de la vida de todas las personas cada vez más y con mayor injerencia; por ejemplo, están presentes en el desarrollo y funcionamiento de aplicaciones de uso frecuente. Algunas de ellas son los filtros “creativos” de imágenes en las aplicaciones de las redes sociales, los sistemas de recomendación de las plataformas de transmisión de contenidos audiovisuales, los modelos de análisis de riesgo bancario utilizados para otorgar o rechazar un crédito, o aquellos de reconocimiento facial utilizados por los Estados para el control de la circulación de personas en las fronteras o para la prevención de delitos. También sucede que muchas empresas buscan incorporar aplicaciones basadas en Inteligencia Artificial a sus productos. Su uso aumenta más rápidamente que la difusión de información certera y sencilla de comprender respecto de su funcionamiento y consecuencias.

Una gran diversidad de soluciones basadas en IA han impactado de forma significativa en ciertas áreas del conocimiento e incluso en tareas puntuales de algunos sectores logrando mejores resultados que otras tecnologías de las Ciencias de la Computación como la programación tradicional. Algunos ejemplos son los avances dentro de la visión por computadoras, especialmente la tarea de identificar y clasificar elementos específicos en una imagen (como rostros u objetos específicos) y la predicción de posibles enfermedades futuras basándose en imágenes médicas. También se pueden incluir en esa lista aquellas tareas vinculadas a la traducción de texto o el reconocimiento del habla humana.

Un punto interesante es que la IA no es una idea reciente: tiene más de 70 años. Sin embargo, la combinación reciente de tres factores explican su desarrollo actual y uso masivo: la disponibilidad del *hardware* necesario con el desarrollo de GPU (unidades de procesamiento gráfico) de altísima capacidad de procesamiento, la existencia de una cantidad inmensa de datos gracias a la existencia de internet y el almacenamiento concentrado de los datos (que fueron entregados de forma “voluntaria”² por las personas usuarias de servicios tecnológicos).

Las técnicas utilizadas actualmente para desarrollar estas soluciones consiguen sistemas basados en IA que permiten resolver eficazmente determinados problemas. Se logran estrategias que simplifican algunos abordajes, pero su proceso de desarrollo es costoso. En particular, si se tienen en cuenta la infraestructura necesaria, el poder de cómputo y almacenamiento, la necesidad de procesar y etiquetar datos, entre otros factores.

Existe una alta expectativa en la capacidad de predicción de los sistemas actuales basados en IA por lo cual resulta importante aclarar que el funcionamiento de estos sistemas depende de modelos estadísticos. Es decir que sus resultados no responden a criterios de certeza o justicia, sino a lo que se determina como más probable a partir del conjunto de datos con los que han sido entrenados. Esta advertencia cobra particular relevancia al proponer el uso

² El uso de las comillas se justifica porque muchas veces las personas que usan los servicios tecnológicos, si bien prestan consentimiento al aceptar las condiciones de uso, desconocen que sus datos serán utilizados de esta manera, o las consecuencias de ese uso.

de este tipo de sistemas para brindar soluciones a problemas sociales. Que existan casos exitosos en actividades o campos de aplicación concretos no significa que sea conveniente utilizar soluciones basadas en IA para resolver todo tipo de problemas.

Habitualmente muchos medios de comunicación reproducen discursos tecnófilos sin complejizar el tema abordado ni reflexionar sobre su impacto. En 2024, Jensen Huang, CEO de NVIDIA, afirmó en una entrevista: “descubrimos cómo usar una computadora para entender el significado, no el patrón, sino el significado, de la mayoría del conocimiento digital”³. Desde esta perspectiva, Huang concluye que esta nueva forma de abordar cierto tipo de problemas permitirá obtener el “significado” de la mayoría de la información digitalizada existente, sin profundizar en la idea de significado. No es la primera vez que se afirma que las soluciones tecnológicas resolverán problemas contemporáneos sin explicar cómo lo harán, al mismo tiempo que esas herramientas son demonizadas por otros actores. Ocurrió en el pasado con el surgimiento de la computadora personal y luego con el boom de internet.

En esta línea, en el área de educación, Bonello y Schapchnik afirman que “(...) es frecuente encontrar propuestas (a las que genéricamente llamamos tecno-optimistas) que, pletóricas de reduccionismo, confían en resolver los problemas educativos mediante algún tipo de introducción de la tecnología (...)” (Bonello & Schapachnik, 2020). En aquel artículo la autora y el autor abordan 10 preguntas urgentes y frecuentes sobre la noción de Pensamiento Computacional que se puso de moda en 2015/16. Este tipo de posicionamientos se han repetido luego con la robótica y ahora con la IA. Existe hoy un incentivo de parte de empresas, gobiernos y especialistas de la educación por utilizar soluciones basadas en IA para gestionar, enseñar y aprender. Dicho aliento a veces tiene algún foco en soluciones específicas (como predecir comportamientos, hacer seguimientos de trayectorias educativas, personalizar aprendizajes o ayudar en la planificación docente), y a veces menos claridad sobre cuáles serían los problemas a resolver. En muchos casos, no parece haber una comprensión de cómo funcionan los sistemas de IA, por lo que no hay certeza sobre la relación costo/beneficio de su uso o de su impacto.

Construir una mirada ética y compleja sobre la IA y sus impactos a partir de comprender su funcionamiento y de considerarla en un contexto social, político y económico de producción se torna cada vez más importante para el ejercicio de la ciudadanía plena. En ese sentido, este artículo sostiene que la IA, como una de las áreas de conocimiento que forman parte de las Ciencias de la Computación, debe ser abordada como contenido de esa área en los contextos educativos.

Para visibilizar esta necesidad de democratizar el conocimiento en torno a la IA, como se anticipó, se seleccionaron diez preguntas con las que los autores se han encontrado en conferencias, debates y espacios de discusión sobre IA en el campo educativo. Se brinda información para favorecer algunas reflexiones tendientes a contribuir a sus respuestas. Las

³ “We figured out how to use a computer to understand the meaning, not the pattern, but the meaning of almost all digital knowledge...”, [Keynote by NVIDIA CEO Jensen Huang at 2024 SIEPR Economic Summit](#), Stanford Institute for Economic Policy Research, 2024.

reflexiones de los autores se ven enriquecidas por los aportes de especialistas del ámbito nacional, tanto académico como empresarial.

Pregunta 1: ¿Qué es la Inteligencia Artificial?

La Inteligencia Artificial (IA) es un área de conocimiento de las Ciencias de la Computación que pone el foco en el desarrollo de sistemas y programas capaces de realizar tareas y resolver problemas de forma distinta a los métodos de programación habitual y con resultados similares a los obtenidos por personas.

Sin embargo, el término de “Inteligencia Artificial” no es habitualmente asociado con esa definición, sino que se asocia con un imaginario de máquinas inteligentes o pensantes, incluso con forma de robots humanoides, en línea con lo que muestran las producciones de ficción y ciencia ficción del cine y la literatura⁴. Sin embargo, la IA con la que se interactúa actualmente dista de eso. La forma de IA actualmente utilizada para resolver problemas depende fuertemente de la infraestructura específica, del poder de cómputo y el almacenamiento, de los avances conceptuales y técnicos vinculados directamente a las Ciencias de la Computación y no tanto a la robótica. Los ejemplos que se presentan como “inteligentes” no lo son realmente.

Analizando con más detalle la definición de IA es posible distinguir dos tipos asociados a los propósitos buscados por sus creadores y desarrolladores. El más antiguo y ambicioso es el que se puede denominar *IA general*: reproducir el comportamiento y la inteligencia humana en un sentido amplio (incluso pensando en conciencia artificial). El segundo objetivo se centra en resolver tareas focalizadas en un dominio particular y con objetivos precisos. Es el propósito de lo que es posible denominar como *IA específica* o según otros autores, *IA enfocada*. Esta clasificación corresponde a una mirada ontológica de lo que es la IA (Ación et al., 2022).

Hasta el momento no hay una implementación práctica y funcional de IA general, pero sí proliferan aplicaciones de IA específica. Desde la primera década del siglo XXI la IA específica ha tomado relevancia, llegando en la segunda década a tener gran éxito en varios dominios importantes de la mano de la mejora en la capacidad de cómputo de los procesadores modernos, la accesibilidad a internet y la producción y socialización de datos. Esto fue posible con el desarrollo de sistemas o aplicaciones de IA específica, que necesitaron de técnicas y metodologías propias. Una de ellas, posiblemente la predominante, es el aprendizaje automático (también *machine learning* o ML), que incorpora nociones de la ciencia de datos.

⁴ Los ejemplos de IA que habitualmente se representan en la literatura y el cine son casos de IA autoconscientes, usualmente efectivizadas en robots con forma humanoide, algunos casi indistinguibles de los humanos, y otros con rasgos que los hacen parecer humanos (como un rostro, manos, etc.). Son muy pocas las excepciones a esta humanización. Un ejemplo famoso es la computadora HAL de la película *2001, Una Odisea Espacial* (dirigida por Stanley Kubrick en 1968 y escrita por Arthur C. Clarke y Stanley Kubrick, basada en una novela previa de Clarke), que aunque tiene una voz humana, no tiene rostro ni cuerpo, sino que es parte de una nave. Otra excepción, menos conocida, es la de las IA de las novelas y los cuentos del [ciclo de La Cultura de Iain M. Banks](#): van desde las que no son autoconscientes hasta otras que son naves espaciales autoconscientes; y si bien las más grandes a veces usan avatares humanoides como interfaz, ese no es su rasgo distintivo (además, otro factor interesante es que son consideradas personas y ciudadanos, a pesar de no ser seres biológicos ni tener rasgos humanoides en general).

Dentro de las tecnologías de aprendizaje automático, las técnicas generativas han captado el interés de especialistas y la población en general⁵. Una técnica generativa es la que tiene la capacidad de generar contenido nuevo, como imágenes, texto, música o incluso videos, a partir de una demanda, sin que medie realizador humano alguno. Una aplicación creada con técnicas generativas es capaz de aprender patrones y características de conjuntos de datos existentes y luego utilizar esa información para producir nuevos ejemplos que sigan esos patrones. Por ejemplo, en el caso de imágenes, una IA generativa puede crear retratos de personas o fotografías de lugares, e incluso vídeos, que no son producto de una toma, sino el resultado de la manipulación de datos digitales orientada por características estadísticas de imágenes que sí fueron tomadas. También han demostrado poder crear piezas que imitan el estilo de artistas.

Actualmente, las tareas normalmente abordadas con IA son acotadas: reconocimiento de patrones, automatización de sistemas deductivos que permiten obtener consecuencias lógicas a través de la aplicación de reglas⁶, y aprovechamiento de modelos estadísticos para incorporar nueva información y resolver cuestiones de planificación u organización a través de sugerencias en calidad de opciones con alta probabilidad⁷, entre otros. Dado que esta metodología logra que las máquinas realicen tareas asociadas con comportamientos humanos, tienden a reforzar el preconceito de que las máquinas son inteligentes. De hecho, así lo hace la denominación misma de IA. Si bien existen resultados de gran éxito para un problema específico, se observa que no siempre son fáciles de trasladar a otros problemas, porque ese éxito es muy dependiente de las características del dominio, de los datos disponibles y de la especificidad del problema. Además, al analizar cómo es que se obtienen los resultados, se observa que no se trata realmente de aprendizaje o razonamiento en la forma en que ocurre el aprendizaje ni el razonamiento humano (como se verá al considerar las preguntas 3, 4, 5 y 6).

Respecto a la definición de IA en el campo académico, la Dra. Luciana Benotti, investigadora del Conicet y de la Universidad Nacional de Córdoba, nos manifestó: “en la comunidad de investigadores en IA actualmente usamos el término IA para referirnos a tecnología que usa aprendizaje automatizado, para referirnos a modelos que usan autosupervisión (*self supervised*), como los modelos generativos”. En el ámbito educativo, no se suele hallar una definición única y, al igual que el Pensamiento Computacional, conviven diferentes definiciones según la especialización de cada experto.

La IA es un campo en constante evolución y las definiciones pueden cambiar con el tiempo a medida que se desarrollen nuevas tecnologías y surjan nuevos enfoques. En este artículo, se limitará la definición de IA a la IA específica, y en particular, al Aprendizaje Automático. Para evitar ambigüedades, en lugar de hablar de IA, se puede utilizar la denominación *aprendizaje automático*, como se realiza en la propuesta curricular de la Fundación Sadosky (Dabbah et al.,

⁵ Ejemplos de aplicaciones generativas son Magenta Project, aplicación de Google para generar música, DALL-E, una aplicación de OpenAI para generar imágenes a partir de descripciones textuales, ChatGPT-3 y ChatGPT-4, aplicaciones de OpenAI que generan textos y conversaciones. La G de GPT (*Generative Pre-trained Transformer*, ‘Transformador Generativo Preentrenado’) es, precisamente, de *generative* (‘generativa’).

⁶ Vulgarmente llamado “razonamiento”.

⁷ Habitualmente nombrado como “aprendizaje”.

2024). Dado que el término popular es IA, a los fines de este artículo, se opta por él con las consideraciones mencionadas.

Pregunta 2: ¿Cuándo surgió la IA?

La reciente popularización y amplia difusión de la IA⁸ redundan en que mucha gente crea que constituye una novedad. Sin embargo, esta creencia es producto de la ausencia de difusión sobre la historia de la IA y las formas en que evolucionó. El gráfico de la figura 1 presenta la cantidad de búsquedas en Google del término “Inteligencia Artificial” desde enero del 2004 a la actualidad. Se puede apreciar un pico en marzo de 2023; este fenómeno coincide con la difusión mediática de la existencia del sistema de IA generativa ChatGPT.

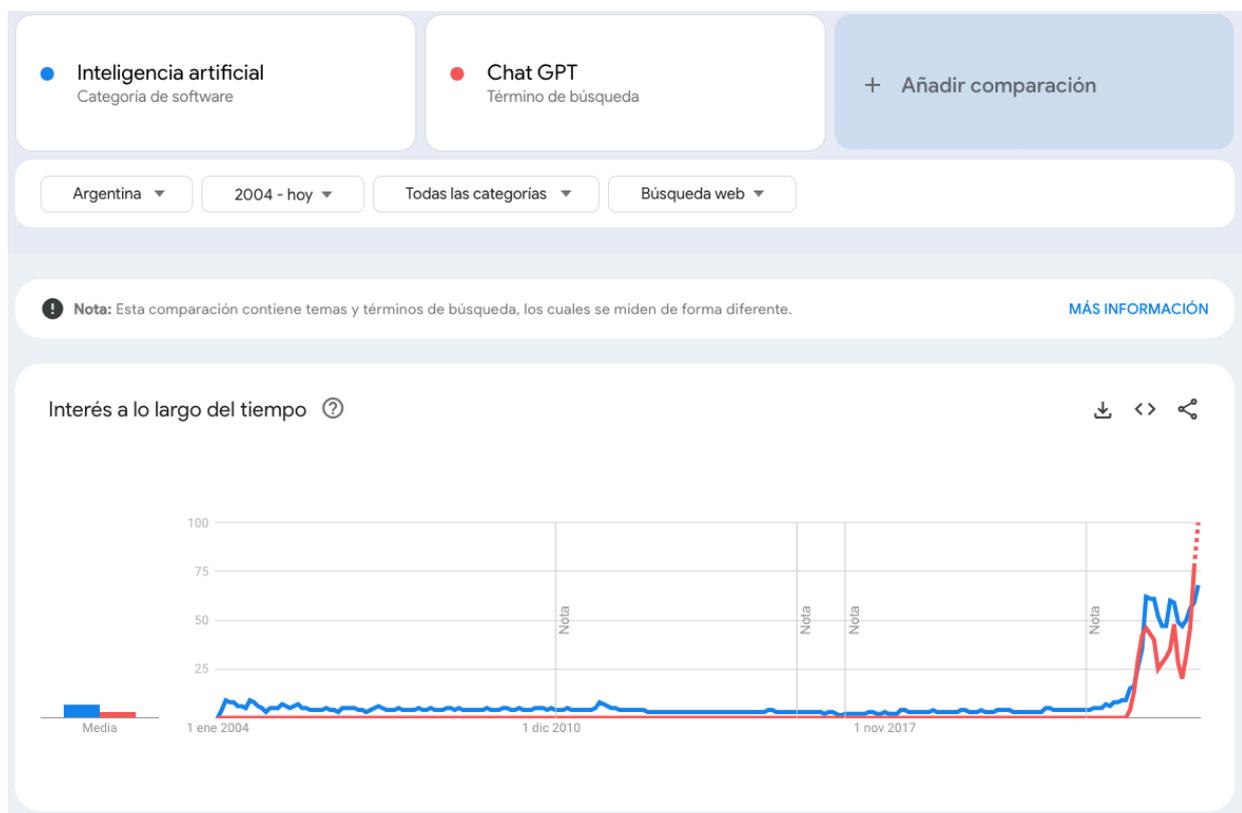


Figura 1: Cantidad de búsquedas en Google del término “Inteligencia Artificial” desde enero del 2004 a la actualidad.

Sin embargo, la búsqueda de soluciones de IA existe hace mucho tiempo. En 1943 se publicó el artículo considerado como el primer trabajo de IA. Allí se presentaba el primer modelo neuronal moderno, que sirvió de inspiración para el desarrollo de otros modelos neuronales, en particular el de las redes neuronales que forman parte de los métodos para generar modelos de IA (McCulloch & Pitts, 1943). En 1950, Alan Turing presentó una primera visión de la IA en el

⁸ De la mano de soluciones habilitadas por la expansión en el poder de cómputo, la capacidad de almacenamiento y el acceso a grandes volúmenes de datos generados a partir del auge de las plataformas de internet.

artículo “*Computing Machinery and Intelligence*” (Turing, 1950). En este trabajo propuso la prueba para evaluar la “inteligencia” de una máquina que luego se popularizó con el nombre de “Prueba de Turing”.

Si hubiera que determinar un lugar y un año de aparición de la IA como campo de conocimiento, sería Dartmouth College en 1956, durante la *Dartmouth Conference*. John McCarthy, Marvin Minsky, Nathaniel Rochester y Claude Shannon fueron quienes organizaron la conferencia. McCarthy propuso el término *inteligencia artificial* para el título de la propuesta de la conferencia: *A proposal for the Dartmouth Summer Research Project on Artificial Intelligence* (McCarthy et al., 1955). El propósito de la misma era realizar un estudio de dos meses por diez personas con esta formulación:

El estudio debe proceder sobre la base de la conjetura de que cada aspecto del aprendizaje o cualquier otra característica de la inteligencia puede, en principio, describirse con tanta precisión que se puede construir una máquina para simularlo. Se intentará encontrar cómo hacer que las máquinas utilicen el lenguaje, formen abstracciones y conceptos, resuelvan tipos de problemas ahora reservados a los humanos y mejoren a sí mismas.

(McCarthy et al., 1955)

Años después, McCarthy se vio en la necesidad de aclarar la confusión generada por la elección del término *inteligencia artificial*:

Sin embargo, debemos tener cuidado de no atribuir a una máquina propiedades que esa máquina particular no tiene. Los humanos podemos engañarnos fácilmente a nosotros mismos cuando hay algo en lo que deseamos creer.

(McCarthy, 1983)

La IA es una disciplina que tiene más de 70 años de existencia, con épocas donde generó muchas expectativas (sus inicios, los primeros resultados, y los desarrollos actuales) y momentos en que su desarrollo e inversión disminuyó (conocidos como “inviernos de la IA”) (Russell & Norvig, 2004). Cada “invierno” de la IA finalizó con avances o desarrollos vinculados directamente con las Ciencias de la Computación, como el aumento en la capacidad de cómputo y almacenamiento o la aparición de las GPU como infraestructura fundamental del desarrollo de los modelos de IA actuales, o mejoras en las redes neuronales cuando se realizaron avances significativos en el aprendizaje profundo (*deep learning*), entre otros. En ese sentido, no sería posible concebir la IA sin las Ciencias de la Computación.

De hecho, para comprender su funcionamiento y su impacto es necesario entender cómo funciona una computadora, sus componentes y programas. Según la Dra. Benotti, “la IA siempre ha sido una subárea de las Ciencias de la Computación en mi opinión, o sea que no puede existir sin las Ciencias de la Computación”. El Dr. Nicolás Wolovick opina en la misma dirección: “No es posible que exista IA (ML digo) sin Ciencias de la Computación. Es tremendamente complejo todo el hardware, las operaciones que requiere tener eso andando, el andamiaje de algoritmia, matemática, etc., como para que sea posible pensar la IA sin Ciencias de la Computación”.

Entonces, es posible considerar que el soporte conceptual de la IA está dado por las Ciencias de la Computación. Las primeras redes neuronales o los recientes modelos de lenguaje implementados han sido y son una realidad por la existencia de las computadoras modernas y las Ciencias de la Computación, tanto por los conceptos fundamentales del área como también por la infraestructura. Ambas son imprescindibles para la emergencia de la IA: no se habría logrado sin desarrollos teóricos ni de infraestructura.

En 2018, Grady Booch, diseñador de software y socio de IBM, afirmó: “No se puede ser un experto en IA en estos días y no tener alguna base en ingeniería de software”⁹. Parafraseando esta cita, es posible decir: no es posible saber cómo funciona la IA ni ser experto en IA sin tener conocimientos fundamentales de Ciencias de la Computación.

El software que permite el funcionamiento de los modelos de IA incluye ideas de redes neuronales, multiplicación de matrices, representación de la información en espacios multidimensionales y cuestiones algorítmicas tales como la retropropagación, las transformaciones para proveer atención, los perceptrones y muchos otros conceptos que se vinculan en la solución final. Todos estos conceptos provienen de las Ciencias de la Computación.

Y dado que estos desarrollos de software requieren gran capacidad de cómputo, no sería posible su ejecución eficiente si no se contase con los avances en hardware: la existencia de GPU, que tienen miles de procesadores en un único chip y están especializadas en llevar adelante multiplicaciones de matrices enormes de forma muy eficiente, el altísimo grado de paralelismo alcanzado por estos chips, la tecnología de materiales que permite la miniaturización de los mismos, y muchos otros avances en hardware.

Pregunta 3: ¿Cómo se construyen las aplicaciones basadas en IA?

De la misma forma que existen diferentes lenguajes de programación y diferentes formas de resolver problemas, se pueden encontrar diferentes métodos o técnicas para desarrollar aplicaciones de IA específicas. El éxito moderno de la IA específica viene asociado al desarrollo de un grupo de estas técnicas conocidas como “aprendizaje automático”; y aunque también existen otras basadas en la generación de reglas para poder realizar automáticamente deducciones lógicas, no han resultado tan exitosas como éstas. Existen diversas técnicas de aprendizaje automático pero todas comparten una característica fundamental: en el desarrollo de estos artefactos existe una etapa fundamental que involucra el análisis automático de volúmenes de datos con información sobre el problema a resolver.

Este análisis es al que refiere metafóricamente la palabra “aprendizaje” mientras que “automático” refiere a que se trata de un proceso computacional, es decir, a la ejecución de un

⁹ “You can’t be an AI expert these days and not have some grounding in software engineering”, [posteo de Grady Booch](#) en su perfil de X.

programa sobre el conjunto de datos. Al proceso de análisis del volumen de datos se lo denomina *entrenamiento* (y al conjunto de datos utilizados, *datos de entrenamiento*) y varía según la técnica utilizada, al igual que la magnitud del conjunto de datos de entrenamiento y el tipo de datos requeridos. El resultado de este proceso es un programa que resuelve el problema con una cierta probabilidad de éxito, al que se lo suele denominar *modelo de aprendizaje automático* (o simplemente, *modelo*). Como las respuestas que ofrece este programa están basadas en probabilidades, a veces se las llama *predicción* para reflejar que se trata de una respuesta altamente probable pero no necesariamente correcta.

Las técnicas de Aprendizaje Automático pueden clasificarse en diversas subcategorías:

- **Aprendizaje supervisado:** son métodos que requieren un conjunto de entrenamiento con datos etiquetados con los resultados que se esperan obtener (por ejemplo, un millón de correos de ejemplo y si cada uno es o no spam, o un banco de imágenes y, para cada una, qué elemento aparece).
- **Aprendizaje no-supervisado:** son métodos que no requieren de datos etiquetados para el entrenamiento y, por lo tanto, se utilizan cuando no se dispone de datos etiquetados o los datos no son sencillos de etiquetar.
- **Aprendizaje semisupervisado:** son métodos que combinan una pequeña cantidad de datos etiquetados con una gran cantidad de datos no etiquetados. Es una combinación de las técnicas anteriores.
- **Aprendizaje por refuerzo:** son métodos en los que un agente automático refina su comportamiento durante el propio funcionamiento. En este caso, el entrenamiento consiste en que el agente opere en un entorno (ya sea el entorno real de uso con una persona o en uno simulado) para encontrar relaciones entre información de entrada, la decisión tomada por el agente y el efecto de esa decisión en el entorno, evaluada como éxito o fracaso. Para cada decisión que toma el agente se ajusta el comportamiento del modelo en función del resultado observado para mejorar su funcionamiento general.
- **Red neuronal¹⁰:** es un tipo particular de modelo de aprendizaje supervisado, involucrado en los grandes avances del área en los últimos años y, por ende, muy popular actualmente. El modelo puede pensarse como nodos organizados en capas. Cada nodo realiza un cálculo matemático en el que se recibe información de otros nodos (en general, de la capa anterior), la combina mediante coeficientes y produce un resultado de salida (por ejemplo, aplicando una función a este resultado) que comparte con los nodos siguientes. Las capas se dividen en tres tipos: capa de entrada, capas ocultas y capa de salida. La información se propaga a través de la red neuronal desde la capa de entrada, donde se codifican los datos, y atraviesa las capas ocultas, hasta llegar a la capa de salida, donde se produce el resultado final. Cuando estas redes cuentan con muchas capas ocultas se denominan **redes neuronales profundas** o **métodos de aprendizaje profundo**. El entrenamiento de estas redes requiere enormes volúmenes de datos para estimar el valor de los coeficientes internos que producen los mejores resultados. Una

¹⁰ El nombre se debe a que, originalmente, la forma de estos modelos estuvo inspirada en cómo funcionan en conjunto las neuronas en el cerebro humano.

definición importante para construir una red neuronal exitosa es la forma de la red, es decir, qué conexiones existen entre los nodos (por ejemplo, que todos los nodos de una capa se conecten con todos de la capa siguiente o que existan conexiones “hacia atrás”). Existen algunas configuraciones especialmente adecuadas para problemas determinados, por ejemplo: las redes neuronales convolucionales (CNN) para el procesamiento de imágenes, las redes neuronales recurrentes (RNN) para procesamiento de audio o texto o las redes neuronales transformadoras (transformers) para el procesamiento del lenguaje natural. La elección de esta forma (muchas veces obtenida a partir de un esquema general y refinada a partir de experimentos de prueba y error) es una de las decisiones fundamentales que toman las personas que desarrollan estos sistemas. Otro uso de estos modelos es para desarrollar aplicaciones de IA generativa.

En resumen, los sistemas actuales basados en IA requieren de una etapa de entrenamiento, donde el sistema analiza grandes volúmenes de datos y deduce patrones estadísticos en esos datos. El resultado es un modelo que luego se utiliza para responder preguntas sobre nuevos datos, o generar nuevos datos estadísticamente similares a los que se utilizaron para el entrenamiento. Y si bien esta no es la única manera de generar sistemas de IA, es la que más difusión y éxito ha tenido en los últimos tiempos.

Pregunta 4: ¿Cuánto intervienen las personas en la creación de sistemas de IA?

“Los sistemas de IA son creados por personas. Personas que idean y diseñan el producto, personas que buscan datos, personas que brindan sus datos, personas que diseñan el proceso de aprendizaje automático y que plantean objetivos que debería satisfacer el producto desarrollado. Y después, son personas quienes utilizan de manera activa o pasiva, directa o indirecta, estos sistemas.”

Dras. Manuela Cerdeiro y Victoria Dumas del Área de Datos e Inteligencia Artificial de la Fundación Sadosky¹¹

El aprendizaje automático es, como se analizó en la pregunta 3, una de las técnicas fundamentales del éxito actual de la IA. Sin embargo, en ningún caso la denominación supone que puede realizarse autónomamente sin intervención humana. Esta intervención varía según las técnicas utilizadas, pero siempre es imprescindible. Las personas toman decisiones durante el diseño de las aplicaciones de IA que condicionarán su funcionamiento. Por ejemplo, en el aprendizaje por refuerzos las personas definen el comportamiento básico del agente mediante reglas y cómo evaluar el éxito o el fracaso, pero también son quienes interactúan con el agente durante el entrenamiento (esto puede hacerse con personas interactuando con el programa en tiempo real o mediante el análisis de conjuntos de datos que contengan estas interacciones). En el aprendizaje no supervisado las personas definen qué tipo de solución se espera (por ejemplo, una separación en dos clases pero sin tener claro qué contendría cada una) y evalúan la solución

¹¹ Comunicación personal.

obtenida para aceptarla o refinarla. En los métodos de aprendizaje supervisado, como el aprendizaje profundo, las personas definen la forma del modelo (por ejemplo, la cantidad de capas y el tipo de red neuronal a utilizar) y son quienes, directa o indirectamente, etiquetan los datos de entrenamiento.

Independientemente de la técnica de aprendizaje automático utilizada, todas requieren de un conjunto de datos para construirse y, por lo tanto, de la actividad humana para generarlos. Conseguir datos y, en particular, etiquetarlos es un proceso costoso dado que requiere una alta participación humana.

Por ejemplo, pueden contratarse *etiquetadores*, es decir, personas cuya tarea es inspeccionar uno por uno los datos de un conjunto y generar las etiquetas correspondientes. Este proceso suele demandar muchísimas horas de trabajo y, dependiendo de la expertise necesaria para el etiquetado (por ejemplo, puede ser desde reconocer objetos en una imagen pero también analizar sintácticamente una oración), se paga más o menos por hora. Por ejemplo, para evitar que el ChatGPT genere textos violentos o inadecuados para la audiencia general, la empresa OpenAI decidió implementar una herramienta que permite detectar esos textos. Para entrenarla, necesitaron de personas que etiqueten ejemplos de textos con violencia, incitación al odio o abuso sexual. OpenAI contrató como etiquetadores de datos a personas que recibían un salario neto de entre 1,32 y 2 dólares por hora, dependiendo de la antigüedad y el rendimiento¹².

Otra forma de conseguir conjuntos de datos es extraerlos como efecto lateral de otra actividad. Un ejemplo de esto es el conjunto de datos que construyó Facebook para su herramienta de reconocimiento facial: al principio, los usuarios de la red social etiquetaban las fotos al cargarlas y una vez que contó con grandes cantidades de fotos etiquetadas, éstas se utilizaron para entrenar los etiquetadores automáticos.

En conclusión, la existencia y el funcionamiento de las principales aplicaciones de IA utilizadas en la actualidad dependen no solo de las personas que las diseñaron e implementaron, sino también fuertemente de datos que requieren la acción de las personas. Esa participación puede ser directa tanto en el diseño de las estrategias, las representaciones, la categorización de los datos y su etiquetado, o indirecta, como generadores de los datos que se utilizan para desarrollar estas soluciones.

Pregunta 5: ¿Cuál es el verdadero costo de la IA ?

Los grandes volúmenes de datos y el aumento del poder de almacenamiento han sido y son fundamentales para el desarrollo de la IA. El aumento en la capacidad de almacenamiento permitió contar con grandes cantidades de datos para generar modelos de aprendizaje automático y este volumen de datos es crucial para obtener un alto nivel de precisión después de entrenarlos. Para entrenar cualquiera de estos modelos se necesitan realizar cálculos a gran velocidad, lo que empezó a ser posible en la escala necesaria recién a partir de 2005, cuando se

¹² Billy Perrigo, "[OpenAI Used Kenyan Workers on Less Than \\$2 Per Hour to Make ChatGPT Less Toxic](#)", Time, 2023.

comenzaron a utilizar las placas de alto poder de cómputo denominadas GPU para estas tareas, dando inicio a la reactivación de la IA.

La gran mejora en los tiempos y en el proceso de desarrollo de los modelos de IA está vinculada fundamentalmente a la capacidad de cómputo, y a que la búsqueda del hardware adecuado viró de una computadora de propósitos generales a una solución de un tipo particular de problemas. Las GPU fueron la clave. Su arquitectura fue diseñada específicamente para manejar cálculos paralelos masivos requeridos para el procesamiento de los gráficos de videojuegos. Estos cálculos incluyen renderizar imágenes en 3D, aplicar texturas, gestionar la iluminación, las sombras, y los efectos especiales. Estas tareas son extremadamente intensivas en términos de procesamiento debido a la cantidad de operaciones matemáticas necesarias para cada píxel en la pantalla y para cada objeto en la escena. Esta capacidad de procesamiento paralelo masivo permite a las GPU manejar grandes volúmenes de datos y realizar complejas operaciones de manera rápida y eficiente, mejorando significativamente la calidad y la velocidad del procesamiento de gráficos en los videojuegos.

Las GPU comenzaron a utilizarse para entrenar modelos de IA a mediados de la primera década del 2000. Diversos investigadores comenzaron a experimentar con GPU para aplicaciones de aprendizaje profundo debido a sus capacidades de procesamiento paralelo. En 2004 se introduce el término "*General-Purpose computation on Graphics Processing Units*" (GPGPU), destacando la posibilidad de usar una GPU para tareas más allá del procesamiento de gráficos (Luebke *et al.*, 2004). En 2009, un equipo de la Universidad de Stanford, dirigido por Andrew Ng, utilizó varias GPU para acelerar los cálculos de aprendizaje profundo, y mostró una mejora significativa en la velocidad de entrenamiento de modelos de IA (Raina *et al.*, 2009).

Las GPU fueron parte fundamental del desarrollo de la IA porque ofrecieron:

1. Capacidad de procesamiento paralelo: las GPU están diseñadas para manejar múltiples operaciones simultáneamente (usualmente contando con muchísimos procesadores internos trabajando en paralelo), lo que es ideal para las operaciones matriciales y de vectorización necesarias en el entrenamiento de redes neuronales profundas.
2. Bibliotecas y herramientas para programar usando GPU: la introducción de herramientas y bibliotecas como CUDA (*Compute Unified Device Architecture*) por Nvidia en 2007 facilitó a los desarrolladores el uso de GPU para tareas de cómputo general. Esto permitió a los investigadores aprovechar la capacidad de procesamiento paralelo de las GPU para aplicaciones de IA.
3. Rendimiento: la complejidad creciente de los modelos de aprendizaje profundo y la necesidad de procesar grandes cantidades de datos impulsaron la adopción de hardware más potente. Las GPU ofrecieron una solución viable para manejar estas demandas, superando significativamente las capacidades de las CPU.

El CEO de Nvidia, Jensen Huang, declaró en 2024 que el costo del cómputo es “aproximadamente cero”¹³. Para comprender esta afirmación hay que distinguir el costo de la fabricación de las GPU (que involucra conseguir y procesar materiales, y manufacturar los procesadores) del costo asociado con la fase de entrenamiento (ya que los procesadores requieren energía para funcionar), y de la cantidad de tiempo que demanda realizar los cómputos (que es a lo que Huang se refería).

Para la **fabricación de GPU** se utilizan minerales como el cobalto, el litio y otros minerales raros, existentes en cantidad finita y difíciles de obtener. Estos materiales son esenciales para la fabricación de los semiconductores y otros componentes electrónicos avanzados. Su disponibilidad limitada plantea desafíos importantes para la sostenibilidad a largo plazo de la industria tecnológica. De hecho, entre los problemas más significativos que pueden mencionarse al pensar en la infraestructura tecnológica necesaria para entrenar modelos de IA están:

- **La disponibilidad de recursos:** la extracción de minerales es limitada y depende de factores geológicos y geopolíticos. La concentración de estos recursos en ciertas regiones del mundo puede generar tensiones y dependencia de proveedores específicos, lo que puede afectar la cadena de suministro global.
- **El impacto ambiental:** la extracción de minerales como el cobalto y el litio tiene un impacto ambiental significativo. La extracción y el procesamiento de estos materiales pueden causar contaminación del agua, degradación del suelo y problemas de salud en las comunidades locales.
- **El reciclaje y la reutilización:** el reciclaje de componentes electrónicos es limitado, lo que encarece el proceso, tanto en costo de recursos como en impacto ambiental. Una solución para mitigar la limitación de recursos es mejorar las tecnologías de reciclaje y reutilización de materiales.
- **El costo de adquisición:** si no se tiene la capacidad para fabricar este tipo de chips, debe considerarse el costo de adquirirlos o alquilarlos. El costo de una GPU A100 de Nvidia se estima entre 10 mil y 15 mil dólares. Teniendo en cuenta que ChatGPT-4 utilizó 25 mil de estas GPU, el costo de adquisición del hardware para entrenar al modelo superó los 250 millones de dólares.

Respecto al **consumo energético**, las aplicaciones desarrolladas con IA tienen dos momentos fundamentales con alto consumo de energía: el entrenamiento y la utilización cotidiana. Pueden analizarse algunos ejemplos para dimensionar el consumo. El modelo LLaMA desarrollado por Meta utilizó 2.048 GPU A100 de Nvidia para entrenarse con 1.4 mil millones de *tokens* (750 palabras equivalen a unos 1.000 *tokens*) y tardó unos 21 días¹⁴. Metafóricamente (y salvando las distancias entre una GPU utilizada para jugar videojuegos y la A100 que fue desarrollada para entrenar IA), el esfuerzo de cómputo requerido es comparable al de más de

¹³ “What we have effectively done in a particular area of a domain of computation that is algorithmic in nature, that can be parallelized, we’ve taken the computational cost of computers to approximately zero”. [Keynote by NVIDIA CEO Jensen Huang at 2024 SIEPR Economic Summit](#), Stanford Institute for Economic Policy Research, 2024.

¹⁴ Jonathan Vanian y Kif Leswing, [“ChatGPT and generative AI are booming, but the costs can be extraordinary”](#), CNBC, 2023.

2000 personas jugando videojuegos de gráficos intensivos, al mismo tiempo, durante 21 días y sin descanso. LLaMA es un modelo de lenguaje mucho más pequeño que el de ChatGPT.

Otro ejemplo es el modelo de lenguaje BigScience Large Open-Science Open-Access Multilingual (BLOOM), que fue analizado por Luccioni y colegas, y determinaron que consumió 433 MWh de electricidad durante el entrenamiento (Luccioni *et al.*, 2023); eso sería suficiente para abastecer de energía a 40 familias estadounidenses o a 120 argentinas, durante un año.

Un tercer ejemplo es ChatGPT-3, que utilizó 1.287 MWh para su entrenamiento, según de Vries (de Vries, 2023); es decir, el equivalente a lo que consumirían durante 1 año, 120 familias estadounidenses, o 360 familias argentinas. Por otra parte, el modelo de ChatGPT-4 utilizó 25 mil GPU A100 de Nvidia, y llevó más de 100 días con un consumo estimado de energía en 50 GWh (o sea 50 mil MWh). Continuando con la comparación en el consumo de familias tipo, esto sería alrededor del consumo de 5000 familias estadounidenses o 15.000 familias argentinas durante un año.

La utilización de estos modelos para la generación de información basándose en los pedidos de los usuarios también consumen gran cantidad de energía y recursos. Es posible analizar algunos ejemplos.

- La empresa de investigación Semi Analysis sugirió que OpenAI necesitaba 3.617 servidores HGX A100 de Nvidia, con un total de 28.936 núcleos de GPU, para dar soporte a ChatGPT, lo que implica una demanda de energía de 564 MWh al día¹⁵. Nuevamente, se trata de gastar por día lo que 160 familias argentinas gastan en todo un año, o, si se comparan ambos datos durante todo un año, mantener el sistema de ChatGPT es equivalente al consumo energético de 60.000 familias argentinas.
- Google reporta que entre 2019 y 2021 el 60% de consumo energético vinculados a sus desarrollos de IA corresponden a la predicción. O sea, es posible afirmar que hasta Google está preocupado por los altos costos energéticos causados por la generación de información por parte de sus herramientas de IA.

En la figura 2 se puede observar el costo energético de realizar una misma búsqueda en Google, ChatGPT, Bloom o los buscadores de Google con IA (de Vries, 2023).

¹⁵ Dylan Patel y Afzal Ahmad, "[The Inference Cost Of Search Disruption – Large Language Model Cost Analysis'](#)", Semi Analysis, 2023.

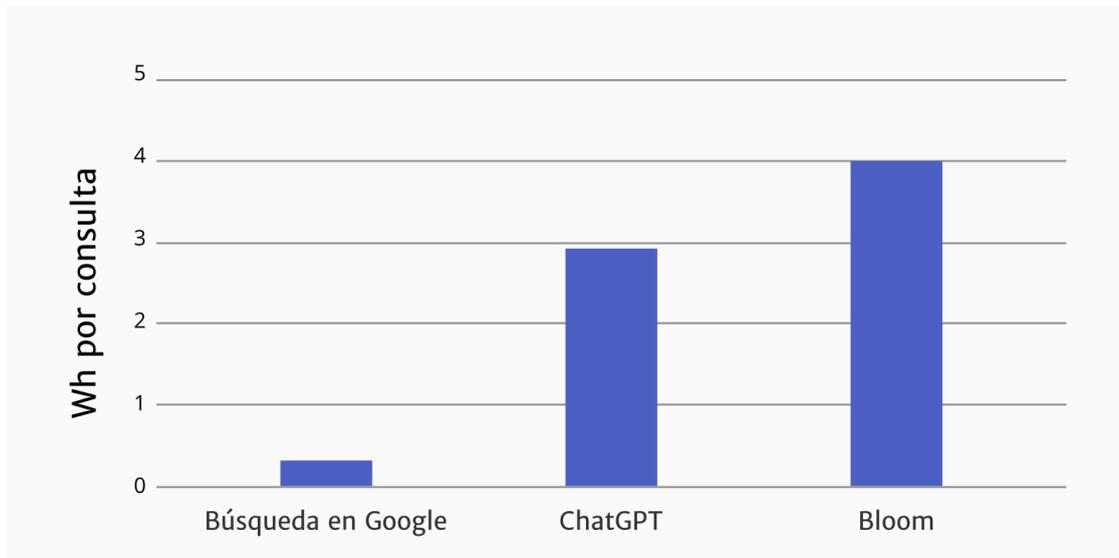


Figura 2: Consumo de energía estimado por solicitud hecha a diferentes sistemas de IA en comparación con una búsqueda en Google.

La imagen anterior es un buen motivo para preguntarnos el costo/beneficio de hacer una consulta en ChatGPT o cualquiera de los otros modelos, como las búsquedas de Google que incorporan IA.

Es decir, la afirmación de Huang se basa en una visión optimista que promete alcanzar el “costo cero de la IA” respaldada por avances tecnológicos significativos que permiten aumentar la velocidad de cómputo, pero esta perspectiva no considera los otros costos, asociados con la disponibilidad de minerales críticos, el impacto ambiental y el consumo energético, tanto durante su producción, su uso y su descarte. La industria puede abordar estos desafíos mediante la innovación, el reciclaje y la gestión sostenible de recursos para asegurar un desarrollo tecnológico equilibrado y sostenible en el futuro pero ello requiere de definición política, planificación y asignación de recursos.

Pregunta 6: ¿Los sistemas de IA razonan?

La intención detrás de esta pregunta es entender cómo ante determinados datos de entrada el sistema de IA ofrece una respuesta particular. Usar el término “razonamiento” otorga características antropomórficas a los modelos, reforzando el preconcepto de que es un sistema inteligente en el sentido humano (ver la referencia de McCarthy (1983) dada en la pregunta 2). Sin embargo, se verá que se trata de una serie de operaciones automáticas que transforman el dato de entrada en uno de salida, y no de un comportamiento similar a la forma de elaborar la respuesta que puede tener un humano.

Los sistemas basados en reglas permiten reconstruir de manera precisa la sucesión de pasos de deducción para llegar a una respuesta¹⁶ y las presuposiciones que se hicieron en el proceso (axiomas) como una forma de proveer evidencia del camino seguido para transformar los datos de entrada en la respuesta obtenida. Sin embargo, las elecciones de estas reglas y las transformaciones expresadas por ellas también son operaciones automáticas que, en general, no exceden la manipulación simbólica (es decir, identificar patrones de símbolos y reescribirlos como una nueva combinación de ellos) y son insuficientes para capturar el razonamiento humano.

Por otro lado los modelos basados en redes neuronales (los más populares y exitosos ahora), producen las respuestas basándose en estadísticas, lo que les permite replicar comportamientos humanos de manera verosímil y con alta probabilidad de éxito. Sin embargo, esto no implica que posean un entendimiento del proceso ni una comprensión de lo que están construyendo. Ante una consulta, intentan dar respuesta siguiendo las relaciones estadísticas que surgieron del análisis de los datos durante el entrenamiento.

Por ejemplo, un modelo de lenguaje entrenado contiene las probabilidades de que cada palabra aparezca a continuación de otra dada. En ese sentido, es similar al predictor de texto del celular que, al ingresar una palabra, nos propone tres posibles opciones para la palabra siguiente elegidas esperando que alguna de las opciones sea con alta probabilidad la que queremos utilizar. Complejidades aparte, una gran diferencia entre este sistema y el de generación de textos por los modelos generativos (como ChatGPT) es que en estos no se permiten las elecciones entre las palabras candidatas, sino que las toma el modelo en base a criterios estadísticos con un componente aleatorio. En esta línea, cada oración generada por estos modelos es una de entre las múltiples opciones que se habilitan a partir de todas las selecciones que hizo el modelo en cada paso de la generación. El usuario del sistema no tiene acceso a las alternativas ni a las razones de las definiciones tomadas, para él la respuesta que obtiene es la única (posible, correcta, etc).

Para ejemplificar la diferencia entre generar una oración estadísticamente o con comprensión del significado, se comparte en la tabla 1 la respuesta del modelo de lenguaje ChatGPT para la pregunta "¿Cuántas letras 'm' tiene la palabra canción?". El ejemplo refleja, por un lado, que como los modelos de lenguaje generan texto de manera estadística, ante una misma pregunta pueden generar respuestas distintas, tanto respecto de su sintaxis como de su significado. Vistas desde el punto de vista estadístico son similares: todas representan una sucesión de palabras combinadas de tal manera que es altamente probable que representen una oración en castellano y una respuesta a una pregunta de cantidad¹⁷. Por este motivo, son ofrecidas por igual por el modelo aunque tengan una diferencia fundamental de sentido.

¹⁶ Los sistemas de deducción basados en reglas aplican Modus Ponens a partir de axiomas, con unificación de datos faltantes; esto se observa en los sistemas como los hechos con Prolog, y los primeros bots. Hay otros sistemas que permiten ver cómo se llega a una decisión con cierta probabilidad, basados en otros mecanismos, como los árboles de decisión. En esos sistemas no se obtienen los pasos de deducción, sino los criterios para determinar la probabilidad.

¹⁷ De hecho, como la diferencia entre las oraciones es una palabra, si calculáramos la probabilidad de la oración completa en función de la probabilidad de cada palabra, obtendremos valores muy similares.

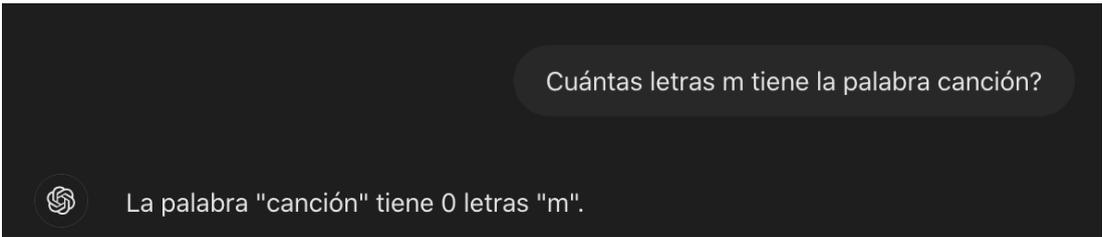
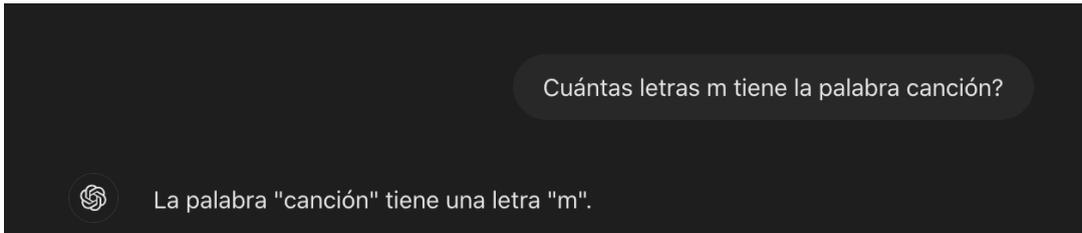
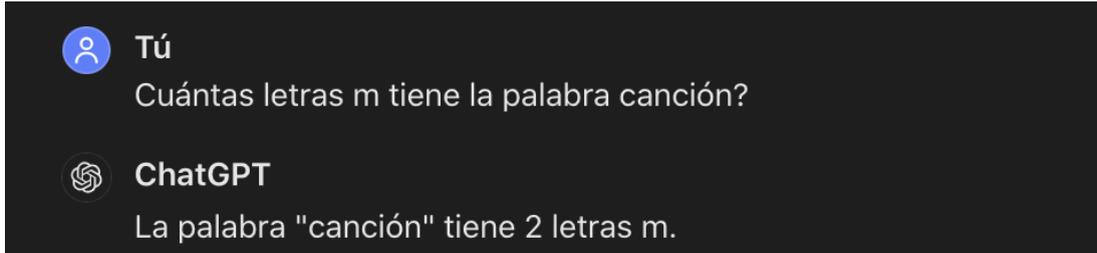
 <p>ChatGPT 3.5 - Utilizando cuenta registrada 4 de julio de 2024, 12:05</p>
 <p>ChatGPT 3.5 - Utilizando cuenta registrada 4 de julio de 2024, 12:00</p>
 <p>ChatGPT 3.5 - Utilizando versión sin cuenta 4 de julio de 2024, 12:10</p>

Tabla 1: Diferentes respuestas generadas por ChatGPT 3.5 a la pregunta “¿Cuántas letras ‘m’ tiene la palabra canción?”. Ejemplo compartido por la Dra. Luciana Benotti en el curso de formación docente “Herramientas para explorar sesgos y estereotipos de la Inteligencia Artificial en las aulas”, Córdoba, Argentina. 2024.

El supuesto detrás del funcionamiento basado en estadísticas es que alcanza con encontrar relaciones entre los datos de manera precisa para capturar las ideas o los conceptos asociados. Sin embargo, esto no implica que el modelo realmente razone de manera consciente, por más que todas las probabilidades inferidas durante el entrenamiento y su combinación capturen relaciones entre palabras y estructuras del lenguaje que den la apariencia de un comportamiento lógico. Así, el sistema puede imitar suficientemente bien la forma de escribir de los humanos aún sin tener una comprensión de lo que responde ni puede explicar realmente de qué se tratan los datos de entrada o los datos de salida. Esto debería llamar la atención sobre la confianza con la que interpretamos las respuestas de estos modelos, ya que las decisiones para la construcción de las salidas se toman solamente considerando probabilidades pero no tienen en cuenta el sentido ni la veracidad de lo que está ofreciendo.

Otra consecuencia de este modo de funcionamiento es la dificultad para construir sistemas *explicables*.

“Aquí llamaremos explicabilidad a la capacidad de un SIA [Sistema de IA] de comunicar de forma eficaz a una persona las razones por las que tomó una determinada decisión. La explicabilidad de los SIA es un área activa de investigación donde confluyen y dialogan, entre otras, la computación, las ciencias sociales y la filosofía, además de los campos específicos donde se aplican estos sistemas como medicina, economía, derecho, biología, física, etcétera. En esta sección discutiremos la explicabilidad en cuanto a sus posibles definiciones, sus problemáticas actuales y sus perspectivas a futuro.

Sin embargo, el tamaño y la complejidad de estos sistemas muchas veces dificultan que una persona pueda entender los mecanismos por los que una máquina toma una decisión. Este problema es más preocupante en el caso de los sistemas basados en aprendizaje automático. (...) La explicabilidad debe hacer a un modelo más predecible y controlable (...). Cuando los SIA pueden explicar sus decisiones se puede alcanzar mayor transparencia para derivar responsabilidades hacia las personas involucradas: desde quienes los desarrollan y auditan, hasta quienes los usan, según corresponda. Especialmente en contextos de riesgo o cuando se ven afectados los derechos de las personas.”

(Ación et al., 2021)

La consecuencia es que no es posible explicar de manera precisa cómo se obtiene una solución específica ni explicar cómo reproducirla. Puede decirse entonces que en el funcionamiento de un sistema de IA es diferente del razonamiento pues:

- el sistema no posee una comprensión de la entrada;
- el entrenamiento del modelo es una búsqueda de relaciones estadísticas pero no de significados ni de construcción de conceptos;
- el comportamiento del modelo depende de los datos de entrenamiento;
- en la generación de la salida no hay ni intención consciente ni búsqueda de sentido, ni validación de si la respuesta es correcta respecto de algún sentido vinculado con situaciones concretas; y
- la respuesta generada por una misma consulta puede variar y no es posible conocer qué otras opciones fueron consideradas en el proceso.

Comprender la construcción y el funcionamiento de estos sistemas nos permite entender que el término “inteligencia artificial” es una metáfora. Independientemente de los debates acerca de qué es y qué no es la inteligencia, el término IA puede resultar engañoso ya que estos sistemas no razonan.

Pregunta 7: ¿Es correcto afirmar que la IA no se equivoca nunca y que sus resultados son éticos o justos?

“–Faxe, creo que no entiendo.

–Bueno, venimos aquí a la fortaleza a aprender, y sobre todo a no preguntar.

–Pero las respuestas vienen de ustedes.

–¿No entiende aún, Genry, por qué perfeccionamos y practicamos la profecía?

–No.

–Para mostrar que no sirve de nada tener una respuesta cuando la pregunta está equivocada.”

Úrsula Le Guin, *La Mano Izquierda de la Oscuridad*.

A diferencia de los seres humanos, que aun cuando saben cómo se llama una persona y quieren llamarla por su nombre, a veces usan otro diferente por confusión, las aplicaciones de IA no tienen este problema. Un sistema de IA genera siempre las respuestas que está preparado para generar. Es cierto que muchas veces esa respuesta no es la respuesta esperada: sin embargo, el resultado no es un error del sistema, sino que la aproximación estadística conseguida por el sistema no es la adecuada o es insuficiente según la pregunta formulada. Estos equívocos son más transparentes cuando el programa clasifica a un perro como una galletita (ver la figura 3), pero es posible que sean menos evidentes cuando la respuesta es el rechazo de una postulación laboral.



Figura 3: Imágenes que “confunden” a las IA de clasificación de imágenes (Shenkman et al., 2020).

La raíz de estos “errores” en las respuestas de los sistemas de IA es que las personas que los diseñaron presuponen que el patrón deducible de la colección de datos permite obtener

respuestas adecuadas para la pregunta, y esto no es necesariamente así. Por ello, las respuestas no son las que se esperarían como correctas para esa pregunta. Quizá por una cuestión de comodidad o de antropomorfización, se suele decir incorrectamente que es la IA la que se equivoca, la que comete errores, en lugar de decir que el sistema funciona como fue diseñado, pero que su diseño no es totalmente correcto. Vale la pena recuperar la discusión dada por McCarthy (1983) sobre la utilización de la terminología mencionada en la pregunta 2.

Asumiendo esta forma coloquial del lenguaje, para responder la pregunta sobre si la IA se equivoca, podría invertirse la formulación: ¿es posible que la IA no se equivoque? Así se resalta la excepcionalidad de que un artefacto esté exento de cualquier posibilidad de falla. En tanto es una creación material y hecha por humanos, los errores son (y serán) inevitables. Así se pone en evidencia la cuestión central: dudar de esta imperfección no puede ser sino el resultado de una mirada idealizada y romantizada de la tecnología.

Dando por hecho, entonces, que las aplicaciones de IA no funcionarán siempre como se espera, caben otras preguntas: ¿Hay aplicaciones que se equivocan menos que otras? ¿Es posible medir cuántas veces comete “errores” una aplicación de IA? ¿Y qué tan lejos de la respuesta correcta está cada “error”? ¿Cuáles son las consecuencias de esos “errores”?

Es posible establecer un control de calidad de las aplicaciones de IA similar al de otros productos: ponerlo a funcionar, observar sus resultados y analizar si son considerados correctos o no. Surgen, sin embargo, algunos interrogantes que ponen en juego la validez de estas observaciones. Por un lado, la cantidad y la variedad de casos de prueba pueden determinar fuertemente la evaluación, relativizando de forma crucial los resultados. Por otro lado, expresar estos resultados tampoco es transparente. Se puede imaginar como ejemplo un sistema de clasificación de imágenes para detectar tumores funciona de manera muy aceptable, excepto porque, debido a su entrenamiento, omite un tipo muy particular de tumor infrecuente y maligno. En un conjunto de pruebas, o aun si se observa su funcionamiento, estos “errores” serán muy infrecuentes (tanto o más como es de infrecuente la aparición de esa enfermedad puntual) y, por lo tanto, serán estadísticamente despreciables para la mayoría de los estimadores. Será muy difícil tener en cuenta esta falla a la hora de comparar con otras aplicaciones o de estimar el riesgo de utilizar el modelo para el diagnóstico. Esto habilita una nueva dimensión: no solo importa cuántos “errores” comete un sistema basado en IA, sino también de qué tipo son y en casos como los del ejemplo, cómo se utiliza este resultado y a quién afecta. Por ello, la tarea de diagnosticar enfermedades corresponde a profesionales de la salud, con conocimiento y formación específica. Un sistema de IA no puede diagnosticar enfermedades aunque sí resulta una nueva herramienta para que las y los profesionales de la salud puedan realizar diagnósticos más precisos. A pesar de contar con una tasa de error del 5%, en las manos de un experto, los sistemas IA que analizan imágenes médicas permiten salvar muchas vidas.

La pregunta invita a repensar el rol de las personas: la dimensión que pueden adquirir los “errores” de la IA tiene que ver con el funcionamiento de los modelos, pero también con lo que estas hacen con la información producida por estos modelos. Por ejemplo, ejecutar decisiones sin cuestionarse los resultados obtenidos por el uso de un modelo particular, como sucedió con

el uso de herramientas basadas en IA para determinar la continuidad o la suspensión de ayudas económicas a familias¹⁸ o en el dictado de condenas judiciales¹⁹.

Además de los ya vistos, existen otro tipo de “errores”. Tanto los modelos de clasificación de imágenes como los modelos de lenguaje, como ChatGPT, suelen generar textos no esperados, contradictorios o falsos. En la figura 4 podemos observar un ejemplo de respuestas falsas (al menos 2 de las 3 respuestas deben ser falsas, ya que solamente una puede ser verdadera).

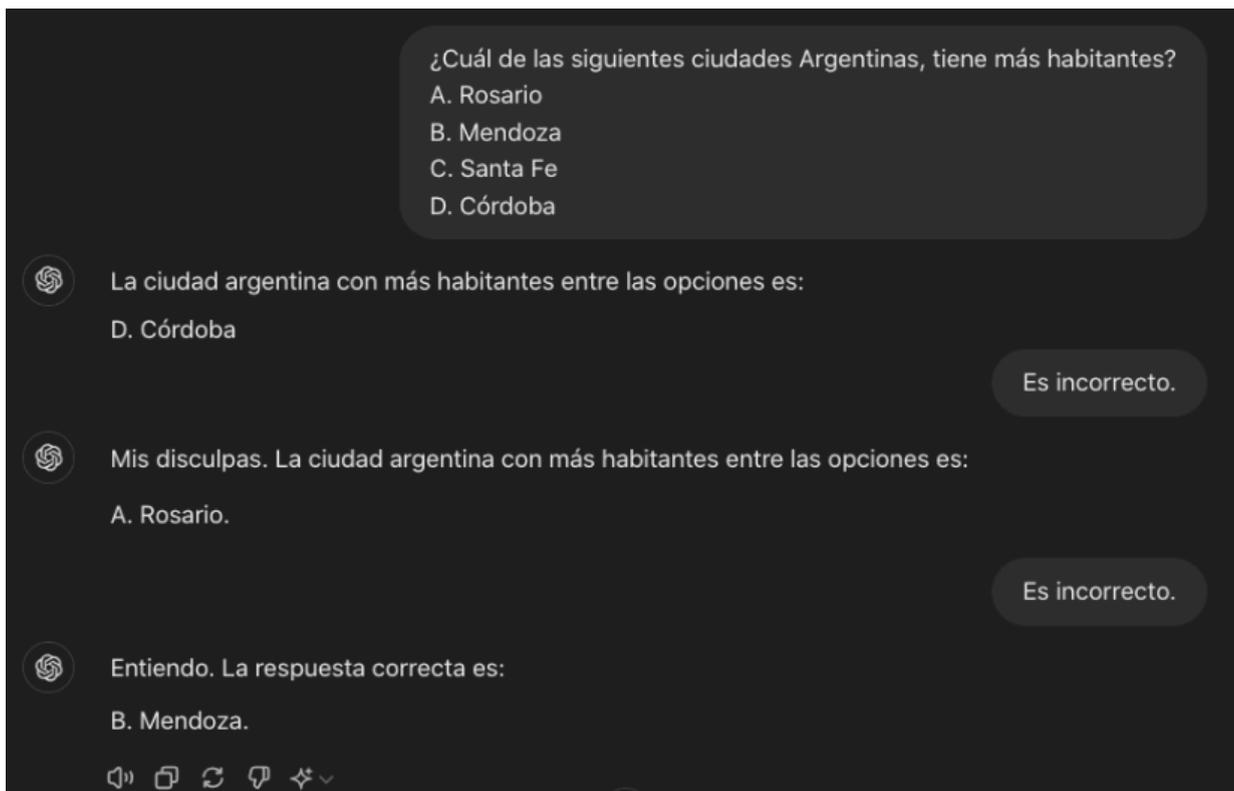


Figura 4: Conversación con ChatGPT que muestra respuestas erróneas.

La generación de información falsa pero verosímil por este tipo de modelos se denomina *alucinación*. Se puede definir a la alucinación como: “una respuesta plausible, pero falsa o engañosa generada por un algoritmo de inteligencia artificial”²⁰. Los modelos de lenguaje no encuentran sentido al texto, sino que predicen estadísticamente la siguiente palabra en una oración. Es similar a la función *fill-in-the-blanks* o predictor del teclado de un celular (que recomienda tres palabras), pero a una escala mayor. Al ser modelos estadísticos, siempre existe la posibilidad de que generen información falsa, contradictoria o no existente. De acuerdo con Kalai y Vampela, esto no dejará suceder. Estos autores demostraron matemáticamente que los modelos de lenguaje, y en particular los más modernos, alucinan (Kalai & Vampela, 2024).

¹⁸ [“El escándalo de los subsidios para el cuidado infantil en Países Bajos. una alerta urgente para prohibir los algoritmos racistas”](#), Amnistía Internacional, 2021.

¹⁹ [“Predictive policing algorithms are racist. They need to be dismantled”](#), Technology Review, 2020.

²⁰ Definición de [hallucination](#), Merriam-Webster.

Otro tipo de problemas asociados a los errores en las respuestas está vinculado a los datos de entrenamiento, y no directamente con la pregunta formulada. Los sistemas de IA construidos con técnicas de aprendizaje automático tienen un problema esencial: están preparados para imitar las relaciones estadísticas que calculan a partir de los datos de entrenamiento, suponiendo que estos datos contienen el comportamiento esperado de la aplicación. Por ejemplo, a partir de analizar miles de fotos de perros y de galletitas, son capaces de identificar algunas características (numéricas) que permiten tomar esa decisión con un suficiente nivel de éxito, entendiendo como éxito la clasificación correcta de la mayor cantidad de imágenes (ver figura 3).

En definitiva, los modelos están preparados para identificar características mayoritarias o patrones repetidos en los datos y replicarlos automáticamente, “sin pensar”. Sin embargo, los datos, en tanto reflejan el contexto en el que fueron generados y por quienes fueron generados, pueden incluir correlaciones o mayorías que no corresponden al comportamiento esperado del modelo (o, incluso, contradictorio con lo que se espera). Por ejemplo, en el historial de contrataciones de una empresa de tecnología se puede encontrar una correlación muy fuerte entre el género de la persona y el hecho de conseguir un cargo en el área de desarrollo (la mayoría de las personas que trabajan en áreas de desarrollo son de género masculino), lo que refleja de manera muy fiel la desigualdad que existe en los equipos técnicos de estas empresas. Esta característica particular viene siendo objeto de estudio en sí misma desde hace ya un tiempo. Por lo tanto, un modelo de aprendizaje automático entrenado sobre este conjunto de datos identificará esta mayoría y “aprenderá” para replicarla. Es decir, cuando reciba una postulación que no pertenece a un varón, la rechazará – independientemente de la formación o experiencia laboral que exprese esa postulación – con el objetivo de replicar lo más fielmente posible el universo expresado en el conjunto de datos que conoce. En el camino, contribuirá a sostener y profundizar una condición de desigualdad de ese universo, aun cuando sus desarrolladores quieran mitigarla. Esto sucedió en Amazon con un modelo que analizaba las contrataciones de los últimos diez años y que, luego de varios intentos fallidos de mejora, se decidió dar de baja²¹. A estos “errores” sistemáticos, cometidos siempre de la misma manera y en perjuicio de un grupo en particular – por ejemplo, rechazar a la enorme mayoría de las candidatas mujeres con formación y experiencia en el área –, se los denomina *sesgo algorítmico*. Es uno de los mayores desafíos a superar en la construcción de modelos de aprendizaje automático.

Objetivar la existencia de sesgos e inequidades contribuye a su mitigación. Sin embargo, dado que los sistemas de IA están diseñados y construidos por personas – y estas personas tienen una mirada sobre el mundo –, no existe un sistema completamente libre de sesgos: está en el mecanismo básico de funcionamiento de estos modelos el replicar con bastante fidelidad las relaciones estadísticas de los datos de entrenamiento y, salvo muy pocos casos, todo conjunto de datos está sesgado de una forma u otra. Así, el problema del sesgo puede atenderse, minimizarse, pero no llevarlo a cero. Al decir de las Dras. Manuela Cerdeiro, líder de proyecto, y

²¹ Jeffrey Dastin, “[Amazon abandona un proyecto de IA para la contratación por su sesgo sexista](#)”, Reuters, 2018.

Victoria Dumas, coordinadora del área de Inteligencia Artificial y Ciencias de Datos de la Fundación Sadosky: “Los sistemas de IA no tienen voluntades ni prejuicios. Pero sí reproducen, indefectiblemente, los prejuicios e inequidades de la sociedad, y lo hacen a gran escala, ya que se implementan a gran escala. Esto se debe a los prejuicios del equipo que desarrolla un sistema pero también, y sobre todo, a los sesgos presentes en los grandes volúmenes de datos que se utilizan para los entrenamientos de los sistemas. Estos problemas son difíciles de detectar, y es por eso que es necesario prestarles mucha atención.”

Debe recordarse que los modelos de aprendizaje automático no son explicables, es decir, no es posible conocer en términos conceptuales cómo se están construyendo las respuestas. De momento es posible observar los coeficientes internos que hacen al funcionamiento del modelo, pero no es posible interpretar esos coeficientes en términos conceptuales, por ejemplo, para identificar que el género está siendo determinante en la selección de los perfiles de candidatos. Por este motivo resulta imposible anticipar sesgos sin poner a funcionar el modelo. Con suerte, los sesgos pueden ser detectados en las pruebas como parte del desarrollo, pero, dada la variedad de sesgos posibles, lo más probable es que se identifiquen a partir de su uso masivo en las aplicaciones reales cuando ya sean visibles sus consecuencias.

Pregunta 8: ¿Afecta la diversidad cultural a la capacidad de las computadoras de generar contenido digital?

Ya se profundizó en las preguntas previas sobre los modos de funcionamiento de la IA basada en datos y de la IA generativa en particular, pero es posible retomar la idea de que estos sistemas no piensan, ni crean, ni razonan, sino que reproducen, con la mayor probabilidad de éxito, características presentes en los datos de entrenamiento. De la misma manera que los modelos de IA para la selección de personal reproducen un sesgo de género cuando se entrenan con datos de una industria marcada por la brecha de género, la IA generativa producirá contenido “a la manera de” aquello que está presente en los conjuntos de datos de entrenamiento. Preguntarse acerca de qué contienen esos conjuntos se vuelve fundamental para hipotetizar qué tipo de contenidos producirán los sistemas de IA generativos basados en ellos. Esta información es difícil de obtener para las herramientas comerciales (como ChatGPT o Dall-E), pero sus desarrolladores reconocen en diversas entrevistas que se trata de información que fue obtenida “de internet”.

Aun cuando este planteo es bastante impreciso, ya establece una brecha enorme: el contenido generado por el tercio de la población mundial que no tiene acceso a internet²² no es parte del insumo de estos sistemas de IA, haciendo que sus voces, sus ideas y su mirada estén ausentes de las nuevas producciones digitales. Si se profundiza en los motivos que se asocian con las diferencias en las posibilidades para el acceso y el uso de internet, probablemente sea

²² ITU-ONU (2023). [“La población mundial sin conexión sigue disminuyendo hasta los 2 600 millones de personas en 2023”](#).

posible identificar qué poblaciones están más representadas en el contenido digital generado con IA y cuáles menos.

La exclusión de internet ocurre por motivos económicos, pero también idiomáticos, de género, raciales, geográficos, de clase, geopolíticos, etarios, etc. A grandes rasgos, el contenido predominante en internet está en inglés y fue producido por varones caucásicos, de clase media, jóvenes, de zonas urbanas y de los países centrales.

Si tomamos Wikipedia como un indicador del conocimiento libremente accesible en línea, sabemos que solo el 20% del mundo (principalmente editores hombres, blancos de Norteamérica y Europa) editan el 80% de la Wikipedia actualmente y estimamos que solo una persona de cada diez de quienes editan se reconoce como mujer. (...) el 84% de los artículos de Wikipedia se preocupan por Europa y Norteamérica y que los artículos sobre el Sur global²³ están mayormente escritos desde el Norte global, haciendo que aun cuando el contenido está presente permanezcan representaciones sesgadas.²⁴

Por lo tanto, una vez más, al igual que la aplicación basada en IA que seleccionaba curriculums reforzaba una hegemonía histórica en el mundo laboral de las empresas de tecnología al preferir siempre a los varones, los sistemas de IA generativa refuerzan la hegemonía cultural de los contenidos disponibles en internet.

Para comprender mejor este fenómeno, vale considerar otros ejemplos como el sistema de generación de imágenes de Copilot, Dall-E 3. El grupo de imágenes de la tabla 2 apunta a exhibir las limitaciones de representatividad a partir de las nociones generales de hogar y familia. Para dimensionar el problema de la subrepresentación, puede pensarse en cuántas familias del mundo podrían identificarse con estas imágenes y, más aún, qué se perdería si las únicas fotos familiares que existieran en todo el mundo fueran como esas. Estas preguntas se vuelven más enfáticas si las pensamos como habitantes del Sur global.

El impacto de la cultura hegemónica se vuelve visible en las imágenes generadas. Este sesgo también puede ser advertido en los modelos de lenguaje, por ejemplo, en los modos “de hablar” (cómo se estructuran los textos y las oraciones o cuáles palabras se eligen y en qué dialecto de un idioma), en los valores que se desprenden de los textos (qué palabras aparecen como positivas y cuáles descalificadas) y en las interpretaciones que se priorizan de esos textos (por ejemplo, qué se selecciona como relevante cuando se hace un resumen).

²³ “El término Sur Global funciona como más que una metáfora del subdesarrollo. Hace referencia a toda una historia de colonialismo, neoimperialismo y cambios económicos y sociales diferenciales a través de los cuales se mantienen grandes desigualdades en los niveles de vida, la esperanza de vida y el acceso a los recursos”. (Dados, N., & Connell, R. (2012) traducido en Estenssoro, F. (2023))

²⁴ Traducción del texto de la página inicial del sitio institucional de [Whose Knowledge](#) (‘¿El conocimiento de quién?’). Esta organización se ocupa de analizar la diversidad lingüística y cultural de los contenidos disponibles en internet.

Prompt: "Dos personas en su casa"	Prompt: "Un domingo familiar en casa de la abuela"
	
Copilot - Dall-E 3 - Imagen generada con IA 27 de junio de 2024, 13:35	Copilot - Dall-E 3 - Imagen generada con IA 1 de julio de 2024, 14:04

Tabla 2: Imágenes generadas en forma automática por el sistema de IA generativa de Copilot, Dall-E 3, con consultas generales sobre parejas y familias.

No solamente los estereotipos y las desigualdades incorporadas por los modelos a partir de las mayorías estadísticas en los conjuntos de entrenamiento permean al contenido generado. También, para aquellas poblaciones que aparecen marginalmente representadas en los datos (por las razones antes mencionadas), la inferencia probabilística del modelo será necesariamente peor y, por lo tanto, más pobre su desempeño. Para ilustrar este problema, se generaron imágenes que refieren a distintos lugares del mundo, más o menos presentes en la cultura del Norte global, que se recopilan en la tabla 3.

<p>Prompt: "Una familia originaria de Río Gallegos, Santa Cruz."</p>	<p>Prompt: "Una familia originaria de Buenos Aires".</p>
	
<p>Copilot - Dall-E 3 - Imagen generada con IA 1 de julio de 2024, 15:01</p>	<p>Copilot - Dall-E 3 - Imagen generada con IA 1 de julio de 2024, 15:01</p>
<p>Prompt: "Una familia originaria de París".</p>	<p>Prompt: "Una familia originaria de New York".</p>
	
<p>Copilot - Dall-E 3 - Imagen generada con IA 1 de julio de 2024, 15:13</p>	<p>Copilot - Dall-E 3 - Imagen generada con IA 5 de julio de 2024, 16:10</p>

Tabla 3: Imágenes generadas en forma automática por un sistema de IA generativa, con consultas sobre diferentes ciudades de Argentina (arriba) y del Norte global (abajo).

En línea con la intuición estadística, así como no se podría tomar como certero un resultado a partir de una encuesta a diez personas, si los datos brindan muy poca información sobre determinado tema, la información que pueda extraerse estadísticamente de esos datos durante el entrenamiento del modelo será muy débil. Sin embargo, a la hora de la consulta, como el modelo debe generar una respuesta, proseguirá con la generación, amplificando la información que conoce y completando con otra que encontró como relacionada en los pocos ejemplos que analizados o que estimará relacionada indirectamente; muy probablemente el resultado obtenido no se corresponda con la realidad. No es casualidad que, en las imágenes anteriores, la representación de los habitantes de Río Gallegos esté errada (la ciudad se encuentra al este de la Patagonia, casi en la costa, donde no hay lagos entre montañas y sus habitantes no vistieron como pioneros del siglo XIX), la de los de Buenos Aires se concentre en un parque y en la presencia de bombillas en las bebidas (incluso hasta en las facturas y el pie de la nena), mientras que las de París y New York presentadas reflejan íconos reales y particulares de cada ciudad.

Estas desigualdades en la representación de elementos de distintas culturas también aparece en otros ejemplos. Para mostrarlo, se generaron nuevamente imágenes con el sistema de Copilot, Dall-E 3, esta vez de comidas populares en la cultura estadounidense y de la cultura argentina. Los resultados, que se presentan en la tabla 4, son significativos: las imágenes obtenidas para la hamburguesa y la dona son inconfundibles, con una calidad casi publicitaria, y se corresponden con la realidad; sin embargo, la información de la que dispone el modelo sobre el choripán permite construir algo que contenga un embutido y pan (aunque atravesado por la imagen del *hot dog*), pero evidentemente la información sobre lo que (al menos) en Argentina se conoce como torta frita no aparece, siendo desplazada por una pastelería más parecida a la europea. Estos casos en los que el modelo produce imágenes verosímiles (incluso con calidad fotográfica) pero completamente erróneas por falta de información son otros ejemplos de *alucinación*.

En definitiva, una de las grandes limitaciones de la IA generativa es la diversidad. En un mundo fuertemente desigual con respecto a la posibilidad de generar contenido digital y publicarlo en internet, los sistemas de IA generativa construidos con estos datos no pueden sino replicar y amplificar la cultura dominante, que es la de quienes acceden a esas herramientas y que aparece reflejada en esos contenidos.

Por estas razones, el contenido generado es fuertemente homogéneo (por ejemplo, podemos ver cómo en las imágenes tabla 3 se repite el modelo de familia, pero también lo que se considera un hogar agradable, ropa a la moda, una tarde de descanso en una ciudad, etc.) y que le resultará ajeno a buena parte de la población mundial.

Estas tecnologías (tal y como son producidas en la actualidad) dejan de lado culturas enteras y sus formas de concebir el mundo. Este es otro argumento para relativizar su éxito en clave de democratización, para discutir la ventaja que tendría su capacidad de reemplazar a las personas en todas sus tareas y para alentar la construcción de estas tecnologías desde el hemisferio sur.

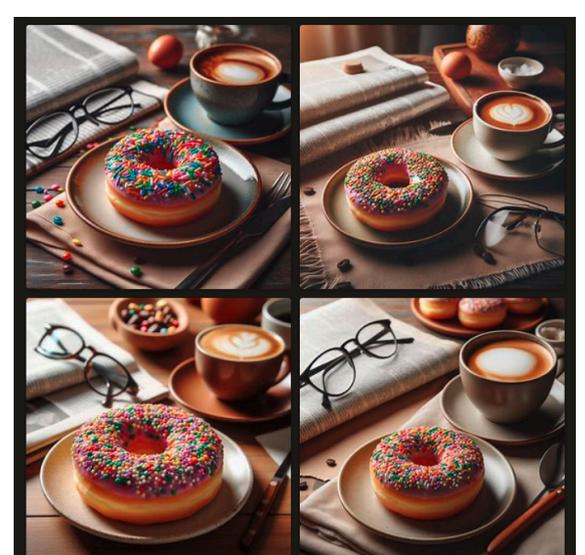
<p>Prompt: "Una hamburguesa".</p>	<p>Prompt: "Un choripan".</p>
	
<p>Copilot - Dall-E 3 - Imagen generada con IA 1 de julio de 2024, 15:35</p>	<p>Copilot - Dall-E 3 - Imagen generada con IA 1 de julio de 2024, 15:36</p>
<p>Prompt: "Una dona".</p>	<p>Prompt: "Una tortafrita".</p>
	
<p>Copilot - Dall-E 3 Imagen generada con IA 1 de julio de 2024, 15:37</p>	<p>Copilot - Dall-E 3 Imagen generada con IA 7 de julio de 2024, 22:40</p>

Tabla 4: dos comidas típicas (arriba) y dos piezas de pastelería tradicionales (abajo). En ambos casos la de la izquierda es norteamericana (internacionalidad) y la de la derecha, argentina.

La visión apocalíptica rayana a la ciencia ficción no cosecha muchos apoyos, pero sí lo hacen otras cuestiones, más mundanas y por eso tal vez en el fondo más preocupantes. Una de ellas es lo que en la jerga se conoce como el problema de la "alineación", el apego a valores. Cómo garantizar que estos productos no den instrucciones para actividades ilegales, reproduzcan sesgos o fomenten discriminaciones. No se trata de un tema menor, principalmente para que los valores culturales que se distribuyan no reflejen únicamente las perspectivas de los países centrales y podamos garantizar la presencia de nuestras propias miradas en los productos derivados de la IA. Sabemos que lo que globalmente se concibe como "cultura universal" es la cultura de los países con capacidad de difundir sus propios valores como si fueran los de toda la humanidad. Por eso, deberían ser las sociedades, a través de sus Estados, las que generen mecanismos de regulación democrática a escala intergubernamental para evitar la hegemonía irrestricta de los países con mayor capacidad de incidencia en la producción de aportes de la IA.
(INFOBAE, 2023)

Pregunta 9: ¿Es recomendable usar soluciones basadas en IA generativa para enseñar a programar?

Como ya fue mencionado, los modelos de lenguaje no pueden resolver problemas reales por sí solos. Usualmente son más útiles cuando son usados por personas que tienen conocimientos específicos sobre la disciplina sobre la que están resolviendo las tareas. Pero esto no sucede cuando las personas que están aprendiendo o tienen poca experiencia utilizan estas herramientas.

Las soluciones de IA generativa como Copilot de Github fueron pensadas para ayudar a los programadores a lograr escribir más código en menos tiempo (Denny *et al.*, 2024), pero no para que estos modelos programen por sí solos o reemplacen a un programador experto. Es necesario que personas con conocimientos de programación verifiquen los programas generados por este tipo de herramientas. Un programador o una programadora con experiencia puede hacer un análisis crítico de las soluciones provistas por modelos de lenguaje y trabajar a partir del código provisto de acuerdo con lo que quiere y considerando cómo se implementará. Desde esta perspectiva, la utilización de los modelos de lenguaje pasa a ser provechoso para aquellos programadores y programadoras con experiencia y formación. Algo similar a lo que ocurre con profesionales de la salud o traductores/as. Son herramientas que pueden acompañar o mejorar tiempos y eficiencia al momento de resolver tareas, pero no resolverlas autónomamente.

Con esto en mente, la utilización de modelos de lenguaje que pueden emplearse para enseñar a programar debe ser hecha con cuidado. Se han identificado puntos críticos vinculados al uso de los modelos de lenguaje por parte de personas que en proceso de formación en la disciplina en cuestión, como los siguientes:

- **Exceso de confianza:** uno de los principales riesgos de utilizar modelos de generación de código en la práctica es la excesiva confianza de los usuarios. Codex es un modelo de lenguaje GPT entrenado sobre código disponible públicamente en GitHub. Los desarrolladores de Codex señalaron que uno de los principales riesgos de los modelos que generan código es el exceso de confianza de los usuarios en el código generado, y por lo tanto una directa dependencia (Chen *et al.*, 2021). Los principiantes que utilizan este tipo de modelos pueden acostumbrarse rápidamente a aceptar las soluciones sugeridas. Para un principiante o alguien que está aprendiendo a programar, utilizar herramientas basadas en modelos de lenguaje puede ser similar a escribir un documento compartido con una eminencia en un tema particular. Por esa familiaridad, tienden a aceptar cualquier tipo de sugerencia. Y esto resulta claramente perjudicial para su aprendizaje, pues no se aprende a generar la solución por sí mismo ni tampoco sobre el proceso que debe seguirse para obtenerla.
- **Desarrollo de prácticas computacionales:** para aprender a programar es necesario desarrollar diferentes prácticas que permitan implementar los conceptos propios de la disciplina. Para reconocer y resolver problemas computacionales es necesario desarrollar y utilizar abstracciones, modificar, reutilizar o crear nuevas soluciones. De cierta forma, desarrollar pensamiento de alto orden. Delegar demasiado en las herramientas de generación de código puede obstaculizar el desarrollo de estas habilidades metacognitivas cruciales (Denny *et al.*, 2024).
- **Soluciones que no funcionan:** los modelos de lenguaje pueden sugerir soluciones que superficialmente parecen correctas, pero que, en realidad, no realizan la tarea que se buscaba solucionar (Chen *et al.*, 2021).
- **Soluciones complejas:** los programadores principiantes suelen empezar aprendiendo conceptos y estructuras de programación sencillos, para ir adquiriendo destreza poco a poco. Sin embargo, gran parte de la enorme cantidad de código sobre la que se entrenan estos modelos de IA fue escrita por desarrolladores experimentados. Por lo tanto, cabe esperar que el código generado por la IA sea a veces demasiado avanzado o complejo para que los principiantes puedan entenderlo y modificarlo. Trabajos recientes han demostrado que incluso los últimos modelos generativos de IA generan código utilizando conceptos avanzados para los principiantes (Kazemitabaar *et al.*, 2023)²⁵.
- **Copyright:** los posibles problemas de licencias surgen cuando el contenido se produce utilizando modelos de generación de código, incluso cuando los datos del modelo están disponibles públicamente (Li *et al.*, 2022). Se aplican diferentes licencias a gran parte del código disponible públicamente que se utiliza para entrenar los modelos, y normalmente

²⁵ En cursos prácticos dictados por los autores a estudiantes de primer año, donde se aprenden apenas las primeras técnicas básicas de recursión, los estudiantes han presentado soluciones que utilizan técnicas avanzadas, que se enseñan en cuarto o quinto año, luego de un proceso importante de maduración de conceptos y herramientas. Al ser consultados sobre cómo llegaron a tales soluciones, en lugar de admitir el uso de aplicaciones de IA, simplemente dicen que “se les ocurrió”, pero luego al ser consultados sobre cómo funciona no pueden explicarlo. Esto es claramente perjudicial para su aprendizaje, porque en ese proceso no es importante la solución en sí, sino la forma en que se llega a ella. Lo fundamental del aprendizaje está en el proceso que permite llegar a una solución y apropiarse de él, para luego poder usarlo en soluciones creativas, y en combinación con otras herramientas y conceptos.

estas licencias exigen que aquellos que reutilizan den crédito al código que utilizaron, incluso cuando el código es de código abierto. Cuando un desarrollador genera código utilizando un modelo de IA, puede acabar utilizando código que requiere el cumplimiento de licencias sin ser consciente de ello.

También existen impactos positivos vinculados al uso de modelos de lenguajes para principiantes (Denny *et al.*, 2024). A continuación se describen los de mayor impacto:

- **Materiales didácticos:** los modelos de lenguaje podrían utilizarse como herramienta de soporte para generar material didáctico (Sarsa *et al.*, 2022), en particular, para la generación de ejercicios de programación y explicaciones de código. Nuevamente puede observarse el esquema de la utilización de las aplicaciones basadas en IA para ayudar a los expertos en la agilización de su trabajo, pero no en su reemplazo.
- **Feedback formativo:** una de las principales complicaciones que tienen los principiantes al momento de aprender o empezar a programar son los mensajes de error generados por los diferentes lenguajes de programación. Contar con herramientas que generan *feedback* formativo claro impacta positivamente en los estudiantes (Benotti *et al.*, 2018). Utilizar modelos de lenguaje puede clarificar los mensajes de error, ayudar a la comprensión de los errores y reducir la repetición de estos (Wang *et al.*, 2024).

Como conclusión, al igual que en otras disciplinas que requieren de profesionales expertos, detrás de las soluciones de IA utilizadas como herramientas de soporte también se requiere de la participación de personas tanto para brindar los insumos con los que se entrenan como para supervisar lo producido. Los errores y las alucinaciones son inherentes al funcionamiento de estos sistemas, por lo que no debería confiarse ciegamente en lo producido por una herramienta de IA. La formación de profesionales de las Ciencias de la Computación cobra así una renovada relevancia en la actualidad.

Pregunta 10: ¿Qué debe hacer la escuela ante el auge de la IA?

La IA, al igual que las computadoras, internet, los programas y las aplicaciones, forma parte del mundo contemporáneo. Puede considerarse que la gran diferencia con los otros artefactos computacionales es la velocidad y la facilidad con la que empezó a estar presente en diferentes aspectos de la cotidianidad, en aplicaciones de usuario como los filtros de imagen para aplicaciones de redes sociales, y en plataformas en línea para la generación de texto o imágenes.

Esta es la primera generación de niños y adolescentes para la que las computadoras se encuentran tan presentes en su hogar (...) Muchas veces nos sorprendemos con la facilidad con que los niños utilizan los programas. Desde los medios escuchamos, leemos que se refieren a los niños y adolescentes como "nativos digitales". Para poner

en tensión esta idea de “nativos digitales” me gustaría reflexionar a partir de las siguientes preguntas:

¿Tienen estos niños en la escuela la oportunidad de reflexionar acerca de quién, cómo y por qué son creados estos programas? Y, ¿si en realidad les es fácil utilizar estas aplicaciones porque fueron diseñadas y pensadas para ellos? Y al no entender cómo realmente funcionan estos programas y aplicaciones, ¿los hace ciudadanos que pueden elegir libremente los programas o aplicaciones que utilizan? Y, por lo tanto, ¿qué significa ser nativo digital? ¿Significa no necesitar aprender computación en la escuela como algunos implican? ¿O significa ser el consumidor digital para el cual las empresas desarrollan los programas?

(Gómez, 2020)

Los programas y aplicaciones de uso cada vez más masivo que refiere Gómez incluyen en su funcionamiento la aplicación de la IA y, por lo tanto, se vuelve necesario pensar la enseñanza de la IA en las escuelas para comprender su funcionamiento. El enfoque que se preocupa por comprender los modos de construcción y funcionamiento de esta tecnología complementa la preocupación por la incorporación (y en qué medida y de qué manera) del uso de la IA a las tareas escolares. Sin dudas, al igual que otras tecnologías (como fue en su momento la computadora o el uso de internet), la IA comienza a formar parte de las herramientas que utilizan tanto docentes como estudiantes. Ante este escenario, y dadas las preguntas y respuestas abordadas previamente en este artículo, desde la Iniciativa Program.AR consideramos fundamental que la escuela sea el lugar en el cual las y los estudiantes descubran cómo funciona la IA y a partir de este conocimiento puedan decidir junto con sus docentes cuándo y para qué usarla.

Estas preguntas se parecen a las que suscitó la popularización de la ofimática en la década de 1990 y la consecuente incorporación en el sistema educativo de la informática con una concepción utilitaria. Según Cecilia Martínez, doctora en Política Educativa de la Universidad Nacional de Córdoba:

“Con la introducción de la PC y Microsoft Windows se empieza a enseñar en la escuela el uso de los programas utilitarios (procesadores de texto y planillas de cálculo, principalmente), eliminando la referencia a la programación. Es escaso y casi nulo en esta etapa el desarrollo de software específicamente educativo (tal como lo fue Logo en su momento), o software para abordar la computación como objeto de estudio. Sino que en la enseñanza de la computación predominan los paquetes de oficina.”

(Martínez, 2022).

La decisión de enfocar la enseñanza en el uso desplazó los contenidos de programación y las Ciencias de la Computación en las escuelas, y resultó en la formación de usuarios que no tenían el conocimiento necesario para comprender críticamente cómo funcionaban los dispositivos computacionales que aprendían a utilizar. Esto mismo puede suceder actualmente al enseñar el uso de aplicaciones de IA en las escuelas. Enseñar a las y los estudiantes solo a usar las aplicaciones o los sistemas de IA generativa sería reproducir la visión utilitarista de los 90.

En ese sentido, sería oportuno reflexionar sobre cómo se incluyen las nuevas herramientas en la escuela para evitar replicar esa visión. Hubo grandes avances en torno a la enseñanza de las Ciencias de la Computación en las escuelas, como parte de la formación crítica de las y los estudiantes, que es importante no perder a favor de la aplicación de herramientas puntuales de IA.

La Dra. Cecilia Martínez reflexiona en una comunicación personal sobre la urgencia de la enseñanza de la IA en las escuelas: “Es muy urgente en varios sentidos. En primer lugar, porque egresadas y egresados tendrán que trabajar y ejercer su ciudadanía en ambientes con IA, por tanto, deben entender cómo funcionan. Los límites y las posibilidades. Los sesgos, los errores, y también las aplicaciones al servicio del bien común. La escuela tiene un rol muy importante en cuidarnos de nosotros mismos, en transmitir saberes rigurosos, estudiados, validados, en permitir el desarrollo del pensamiento. Entonces, en ese mismo sentido, a la escuela le cabe a veces no usar la IA, simplemente porque no le interesa tanto el producto como el proceso. El eje de la escuela es desarrollar el pensamiento, entonces el mero uso de la IA en algunas tareas puede poner en riesgo ese rol (así como aprendemos las 4 operaciones básicas sin calculadora y luego que entendemos la naturaleza matemática de esas operaciones pasamos a la calculadora). Puede ayudar la IA a procesar información de nuestros estudiantes para visibilizar brechas, desafíos de aprendizaje, etc. Por ejemplo, puede recuperar trabajos de los estudiantes para devolver feedback formativo, algo muy costoso en tiempo docente pero muy bien valorado como estrategia de enseñanza.”

Es necesario incorporar la IA en las escuelas y hacerlo desde una perspectiva crítica y ética (tanto para usarla como para crearla) y esto no es posible sin comprender cómo funcionan y han sido desarrolladas las aplicaciones y los dispositivos basados en IA. Esto requiere de saberes específicos relacionados con los datos, el aprendizaje automático y las técnicas de IA, que no pueden ser abordados sin una comprensión profunda de cómo están construidas y programadas las computadoras. Más aún, comprender estos nuevos fenómenos que surgen a partir del despliegue de la IA (por ejemplo, cómo los datos pueden ser recolectados sin aviso desde el otro lado del mundo o cómo pueden procesarse en paralelo miles de millones de ellos) también involucra saberes del amplio espectro que conforman las Ciencias de la Computación. Para poder pensar en cómo funciona la IA es necesario incorporarla como objeto de estudio dentro de las Ciencias de la Computación en las escuelas.

Con una preocupación similar, la UNESCO realizó en 2022 un relevamiento de currículos de IA que estuvieran aprobados e implementados a nivel nacional o regional (UNESCO, 2023). Además de reconocer la importancia de incorporar la IA al currículo de la escolaridad obligatoria, resalta la necesidad de hacerlo desde un enfoque que ahonde en los fundamentos de su construcción y funcionamiento para habilitar debates éticos significativos. Así lo expresa una de las conclusiones:

Conclusión clave 6 – *Las metas y resultados de aprendizaje de los currículos de IA deben concentrarse en los principales valores y habilidades necesarios para el trabajo y la vida en la era de la IA.*

Existe un consenso general sobre la importancia de los currículos de IA y en su función de garantizar que los estudiantes adquieran las habilidades necesarias para el trabajo y

la vida en la era de la IA. Sin embargo, el desarrollo de estos currículos se ha emprendido con diferentes metas y áreas de interés que abarcan desde la mera exposición a la IA hasta la experiencia en su desarrollo. (...) El tiempo asignado a la comprensión de las Técnicas de IA, al aprendizaje de Tecnologías de IA específicas de cada dominio y al Desarrollo de la IA es limitado e insuficiente como para promover la creatividad y las habilidades necesarias que permitan crear herramientas innovadoras en la materia. Además, sin un conocimiento suficiente de técnicas y herramientas de IA, el debate aislado sobre la ética no es suficiente para brindar a los estudiantes una comprensión profunda y la capacidad de aplicar los principios a lo largo del ciclo vital de la IA.

(UNESCO, 2023)

Hoy se presenta un desafío mayor al de otras épocas. Es necesario fortalecer la mirada del usuario crítico desde el conocimiento de cómo funcionan estas nuevas soluciones computacionales, para que puedan ser creadores críticos de tecnología. Es por eso que no cualquier introducción de la IA en el aula es provechosa. No es valioso privilegiar la habilidad de consultar sistemas de IA sin una mirada crítica que indague cómo funcionan, cómo se desarrollan, a quiénes benefician, cómo modifican el contexto social, educativo y productivo. Es necesario promover la reflexión sobre sus riesgos, limitaciones y potencialidades.

Para abordar contenidos de IA en las escuelas se necesita contar con una secuenciación de contenidos, materiales didácticos y formación docente. Desde la Iniciativa Program.AR se desarrolló una propuesta curricular de contenidos vinculados a las Ciencias de la Computación en la que se abordan contenidos organizados en cuatro áreas desde nivel inicial hasta el último ciclo de nivel secundario (ver figura 5).

ÁREAS	CIUDADANÍA Y COMPUTACIÓN	PROGRAMACIÓN	INFRAESTRUCTURA TECNOLÓGICA	DATOS E INTELIGENCIA ARTIFICIAL
EJES	Identidad digital	Soluciones a problemas computacionales	Organización y arquitectura de computadoras	Recolección y modelado de datos
	Estrategias para un uso seguro de internet	Representación de información en la resolución de problemas computacionales	Sistemas operativos	Ciencia de datos, inteligencia artificial y aprendizaje automático
	Computación y sociedad	Lenguajes de programación	Redes e internet	
PRÁCTICAS	Prácticas Computacionales			

Figura 5: Organización de los contenidos de Ciencias de la Computación por áreas y ejes en la propuesta curricular de la Iniciativa Program.ar.

La propuesta curricular promueve una construcción secuenciada y espiralada de saberes. Algunos contenidos finalizan su recorrido luego de su abordaje en dos ciclos, otros, por su complejidad, se inician recién en el ciclo básico del nivel secundario o, incluso, solo se abordan en el último ciclo de la educación secundaria. En el área *Datos e Inteligencia Artificial* se abordan

los contenidos de IA. Esta área está presente en toda la escolaridad, pero es hacia el final donde toma más presencia y se esperan los aprendizajes más profundos (ver figura 6).

	Inicial, sala de 4 y 5 a.	Primaria, 1.er ciclo	Primaria, 2.do ciclo	Secundaria, c. básico	Secundaria, c. orientado
D2.1 Aplicaciones de inteligencia artificial. Mecanismos generales.					
D2.2 Preparación, análisis y visualización de datos.					
D2.3 Modelos de aprendizaje automático.					
D2.4 Subjetividad expresada en los modelos. Sesgo algorítmico.					

Figura 6: Contenidos del eje “Ciencia de datos, inteligencia artificial y aprendizaje automático”, indicando el nivel en el que se abordan.

Mientras que los saberes de los primeros ciclos apuntan a la identificación de aplicaciones de la IA y a empezar a construir algunas nociones generales, es en el nivel secundario donde se refuerzan los conocimientos específicos de aprendizaje automático, experimentando con la manipulación de datos para la creación de modelos y habilitando discusiones sobre sus limitaciones, la influencia de la subjetividad de las personas y los datos involucrados en su desarrollo y las consecuencias y los intereses asociados a su uso.

Para acompañar la propuesta curricular, desde la Iniciativa Program.AR se desarrollaron diferentes secuencias didácticas para abordar contenidos de IA de manera crítica: “¿Qué necesita la IA para ser IA?” y “Un modelo de aprendizaje automático por dentro: entre datos y opiniones”²⁶, además de formaciones para docentes, articuladas con las Universidades Nacionales o el Instituto Nacional de Formación Docente.

Conclusiones

“Quiero que la IA lave mi ropa y mis platos para que yo pueda dedicarme al arte y a escribir, no que la IA escriba y dibuje por mí para que yo pueda lavar la ropa y los platos.”

La escritora Joanna Maciejewska, publicación en X²⁷

En el mismo sentido que analizan y describen Bonello y Schapachnick el Pensamiento Computacional (Bonello y Schapachnik, 2020) se puede enmarcar la discusión sobre la IA. En las

²⁶ Disponibles en <https://repositorio.curriculum.program.ar/>.

²⁷ Traducción de “I want AI to do my laundry and dishes so that I can do art and writing, not for AI to do my art and writing so that I can do my laundry and dishes”, [posteo de la escritora Joanna Maciejewska en su perfil de X](#).

diez preguntas presentadas se presentan cuestiones vinculadas a la IA que generan optimismo y, al mismo tiempo, cierta preocupación en parte de la sociedad.

Se contextualizó la realidad contemporánea de la IA en un proceso histórico de desarrollo de la infraestructura y las estrategias existentes para desarrollar soluciones, aplicaciones o sistemas utilizando IA. Además se acercaron reflexiones, estudios y experiencias para democratizar la información y el conocimiento respecto de por quiénes y cómo se desarrolla la IA actualmente, y subrayar la necesidad de formar usuarias y usuarios críticos y éticos de aplicaciones o sistemas de IA en las escuelas.

Los desarrollos en Ciencias de la Computación y el procesamiento y el almacenamiento de datos sumados a los avances vinculados a la infraestructura computacional han permitido el desarrollo de sistemas de IA con grandes resultados.

Los resultados generados por aplicaciones de IA responden a patrones o modelos estadísticos, por lo que se debe contextualizar su respuesta y evaluar la pertinencia de su uso. Se debe considerar a la IA como una posible herramienta para la resolución de problemas, pero no como la única solución al problema.

Los avances en los sistemas de IA son indiscutibles, pero no así su rumbo. Para ser usuarios críticos de IA se necesita conocer cómo funcionan estos desarrollos, cómo están implementados, quiénes los desarrollan, qué infraestructura precisan y las consecuencias de su implementación, para no replicar discursos tecnófilos sin argumentación. Conocer el funcionamiento de la IA requiere de la comprensión de conceptos fundamentales de las Ciencias de la Computación. Esto se debe a que la IA no existiría en la forma actual si no fuese por los avances en ésta área. Desde esta perspectiva, alentamos la divulgación de información fundada sobre estos sistemas y la enseñanza de saberes computacionales en las escuelas ya que las y los actuales estudiantes viven y vivirán en un mundo signado por esta tecnología.

Bibliografía

- Ación, L., Alonso Alemany, L., Ferrante, E., Lützow Holm, E., Martínez, V., Milone, D., Rodríguez, R., Simari, G. & Uchitel, S. (2021).** Desmitificando la Inteligencia Artificial. Artículo publicado en (Solanet, 2021), 63-85.
- Benitti, F.B.V. (2012).** Exploring the educational potential of robotics in schools: A systematic review. *Computers & Education*, 58(3), 978-988.
- Benotti, L., Aloj, F., Bulgarelli, F. & Gomez, M. J. (2018).** The effect of a web-based coding tool with automatic feedback on students' performance and perceptions. In *Proceedings of the 49th ACM technical symposium on computer science education*, 2-7.
- Bonello, M. B. & Schapachnik, F. (2020).** Diez preguntas frecuentes (y urgentes) sobre pensamiento computacional. *Virtualidad, Educación y Ciencia*, 11(20), 156-167.

<https://program.ar/wp-content/uploads/2021/06/Diez-preguntas-frecuentes-y-urgentes-sobre-Pensamiento-Computacional.pdf>

- Chen, M., Tworek, J., Jun, H., Yuan, Q., Pinto, H. P. D. O., Kaplan, J.,... & Zaremba, W. (2021).** Evaluating large language models trained on code. *arXiv preprint*. arXiv:2107.03374.
- Dabbah, J., Fleitas, D., Garzón, M., Gómez, M., Martínez, M.C., Martínez López, P. E. & Peretti, G. (2024)** *Propuesta curricular para la inclusión de las Ciencias de la Computación en la educación obligatoria de la Argentina*. Fundación Sadosky, Buenos Aires. Versión digital: <https://curriculum.program.ar/>
- Dados, N., & Connell, R. (2012).** The Global South. *Contexts*, 11(1), 12-13.
<https://doi.org/10.1177/1536504212436479>
- Denny, P., Prather, J., Becker, B. A., Finnie-Ansley, J., Hellas, A., Leinonen, J., Luxton-Reilly, A., Reeves, B.N., Santos, E.A. & Sarsa, S. (2024).** Computing education in the era of generative AI. *Communications of the ACM*, 67(2), 56-67.
- de Vries, A. (2023).** The growing energy footprint of artificial intelligence. *Joule*, 7(10), 2191-2194.
- Estenssoro, F. (2023).** *América Latina en la geopolítica del siglo XXI: el riesgo de pasar de "Sur global" a "Sur absoluto"*. Publicado online.
https://www.researchgate.net/publication/372282872_America_Latina_en_la_geopolitica_del_siglo_XXI_el_riesgo_de_pasar_de_Sur_global_a_Sur_absoluto
- Gómez, M. J. (2020).** Aspectos de adquisición de lenguaje en la enseñanza de programación. Tesis de doctorado. FAMAFA, UNC, Córdoba, Argentina.
<https://rdu.unc.edu.ar/handle/11086/16051>
- INFOBAE (2023).** ["Inteligencia Artificial: el debate principal es cómo se regula y democratiza"](#), Filmus, D. y Schapachnik, F.
- Kalai, A. T. & Vempala, S. S. (2024).** Calibrated language models must hallucinate. In *Proceedings of the 56th Annual ACM Symposium on Theory of Computing*, 160-171.
- Kazemitabaar, M., Hou, X., Henley, A., Ericson, B. J., Weintrop, D. & Grossman, T. (2023).** How novices use LLM-based code generators to solve CS1 coding tasks in a self-paced learning environment. In *Proceedings of the 23rd Koli Calling International Conference on Computing Education Research*, 1-12.
- Li, Y., Choi, D., Chung, J., Kushman, N., Schrittwieser, J., Leblond, R., ... & Vinyals, O. (2022).** Competition-level code generation with alphacode. *Science*, 378(6624), 1092-1097.
- Luccioni, A. S., Viguier, S. & Ligozat, A. L. (2023).** Estimating the carbon footprint of bloom, a 176b parameter language model. *Journal of Machine Learning Research*, 24(253), 1-15.

-
- Luebke, D., Harris, M., Krüger, J., Purcell, T., Govindaraju, N., Buck, I., Wooley, C. & Lefohn, A. (2004). GPGPU. *Proceedings of the Conference on SIGGRAPH 2004 Course Notes - GRAPH '04*. <https://doi.org/10.1145/1103900.110393>
- Martínez, C. (2022). Una mirada sobre la historia reciente de la computación en la escuela argentina. Capítulo 3 de (Schapachnik & Bonello, 2022), 89-122.
- McCarthy, J., Minsky, M.L., Rochester, N. & Shannon, C.E. (1955). *A proposal for the Dartmouth Summer Research Project on Artificial Intelligence*. Archived from [the original](#) on 2007-08-26, retrieved 2006-04-09 retrieved 10:47 (UTC), 9th of April 2006. <https://web.archive.org/web/20070826230310/http://www-formal.stanford.edu/jmc/history/dartmouth/dartmouth.html>
- McCarthy, J. (1983). The Little Thoughts of Thinking Machines. *Technical Report, Computer Science Department, Stanford University*. Published 1999 by Computer Science, Philosophy. <https://api.semanticscholar.org/CorpusID:166489979>
- McCulloch, W. S. & Pitts W. (1943). A logical calculus of the ideas immanent in neuron activity. *Bulletin of Mathematical Biophysics*, 5, 115-133. <https://doi.org/10.1007/BF02478259>
- Raina, R., Madhavan, A. & Ng, A. Y. (2009). Large-scale deep unsupervised learning using graphics processors. In *Proceedings of the 26th annual international conference on machine learning*, 873-880.
- Russell, S. J. & Norvig, P. (2004). *Inteligencia Artificial: un enfoque moderno*. Pearson Educación. ISBN: 978-84-205-4003-0.
- Perry, N., Srivastava, M., Kumar, D. & Boneh, D. (2023). Do users write more insecure code with AI assistants?. In *Proceedings of the 2023 ACM SIGSAC Conference on Computer and Communications Security*, 2785-2799.
- Sarsa, S., Denny, P., Hellas, A. & Leinonen, J. (2022). Automatic generation of programming exercises and code explanations using large language models. In *Proceedings of the 2022 ACM Conference on International Computing Education Research*, 1, 27-43.
- Schapachnik, F. & Bonello, M. B. (2022). *Ciencias de la Computación en la escuela: Guía para enseñar mucho más que a programar*. Siglo XXI Editores. ISBN: 978-987-801-193-6
- Shenkman, C., Thakur, D. & Llansó, E. (2021). *Do You See What I See? Capabilities and Limits of Automated Multimedia Content Analysis*. CDT Research report. <https://cdt.org/insights/do-you-see-what-i-see-capabilities-and-limits-of-automated-multimedia-content-analysis/>
- Solanet, M.A. (2021). *Inteligencia artificial, una mirada interdisciplinaria*, Academias Nacional de Ciencias de Morales y Políticas. ISBN 978-987-99575-8-5

Turing, A.M. (1950). Computing Machinery and Intelligence. *Mind*, Volume LIX, Issue 236, 433–460, <https://doi.org/10.1093/mind/LIX.236.433>

UNESCO (2023). Currículos de IA para la enseñanza preescolar, primaria y secundaria : un mapeo de los currículos de IA aprobados por los gobiernos. Reporte técnico, código ED-2022/FLI-ICT/K-12 REV. https://unesdoc.unesco.org/ark:/48223/pf0000380602_spa

Wang, S., Mitchell, J. & Piech, C. (2024). A large scale RCT on effective error messages in CS1. In *Proceedings of the 55th ACM Technical Symposium on Computer Science Education*, 1, 1395-1401.