

45
2025
N°1
45
ENERO - JUNIO
EPISTEME NS
U.C.V

EPISTEME NS

Revista del Instituto de Filosofía

ISSN: 07984324

Dep. Leg.: pp198102DF423

Consideraciones analíticas

Margaret Masterman, su crítica a la noción de paradigma y su invisibilización

MARÍA CAROLINA ÁLVAREZ PUERTA

El papel de las heurísticas en el desarrollo de una Inteligencia Artificial General

LUZ MARINA BARRETO

El Dilema de Benacerraf como una objeción al Platonismo tradicional en matemática

RICARDO DA SILVA Y SOLANI RIVAS

Fundamentos técnicos y dilemas éticos en los modelos de lenguaje natural en las Inteligencias Artificiales

NAHIR HURTADO

An interpretation of the I/Me distinction in G. H. Mead's "The definition of the psychical" and H. Joas' criticism

LEOPOLDO MÁRQUEZ VELASCO

¿Hay matemática más allá de la lógica?

EDGAR VALENZUELA

45

Enero - Junio

N°1

2025



Universidad Central de Venezuela
FACULTAD DE HUMANIDADES Y EDUCACIÓN

Revista Semestral

ISSN: 0798-4324

Depósito Legal: pp198102DF423

EPISTEME NS

Revista del Instituto de Filosofía

45

Enero - Junio

Nº 1

2025

Universidad Central de Venezuela

FACULTAD DE HUMANIDADES Y EDUCACIÓN

© by Instituto de Filosofía
Facultad de Humanidades y Educación
Universidad Central de Venezuela
Publicado en línea en abril 2026
Caracas—Venezuela

EPISTEME NS

EPISTEME NS es una revista de crítica e investigación en filosofía, abierta a todas las corrientes y estilos de pensamiento y a la reflexión en todos los ámbitos del saber filosófico, sin más requisitos que la originalidad, seriedad y rigor argumentativos. Los números salen alternativamente dedicados a temas de Filosofía, Historia de la Filosofía y Filosofía Social (Serie Azul) y de Lógica, Análisis del Lenguaje y Filosofía de las Ciencias (Serie Roja), además de las ediciones especiales (Serie Gris). La Revista, de periodicidad semestral y arbitrada según los procedimientos *ad usum*, es una publicación de circulación internacional indexada en *The Philosopher Index*, *Latindex* y *Clase*, y registrada en *Ulrich's International Periodical Directory*; también aparece en versión electrónica en las plataformas SciELO (www.scielo.org.ve) y Saber UCV (www.saber.ucv.ve). Todas las colaboraciones deben ser inéditas y ser enviadas por triplicado para ser sometidas a la consideración del Comité de arbitraje. Los trabajos podrán ser de tres tipos: Artículos (no deben exceder de 10.000 palabras), Notas y Discusiones (5.000 palabras) y Reseñas. Los Artículos y las Notas y Discusiones deben ser presentados en cuartilla tamaño carta a doble espacio con notas a pie de página (también pueden ser enviados en *attachment* a las siguientes direcciones electrónicas: revista.epistemens@gmail.com y/o nnunezm@gmail.com y deberán estar acompañados de sendos resúmenes, en inglés y en castellano, de una extensión no mayor de diez (10) líneas, acompañados con tres palabras claves. Para cualquier uso posterior de los materiales publicados en EPISTEME NS, es necesaria la autorización escrita del Editor de la Revista.

EPISTEME NS/Revista del Instituto de Filosofía, Vol. 45, (enero-junio, 2025). Caracas: Ediciones de la Facultad de Humanidades y Educación, 1981.

V./ 22 cm.

2 veces al año.

Continúa: Episteme NS

Título de cubierta.

Director –fundador Anuario Episteme: 1957 – Juan David García Bacca

Director–fundador Episteme NS: 1981 – Juan Nuño

Directora: diciembre 2021 – María Carolina Álvarez

ISSN: 0798–4324

1. Filosofía – Publicaciones periódicas.

Depósito Legal: pp198102DF423

EPISTEME NS

Semestral

<i>Director Fundador:</i>	Juan David García Bacca†
<i>Directora:</i>	María Carolina Álvarez Puerta
<i>Comité Editorial:</i>	
<i>Editora:</i>	Nancy Núñez
<i>Secretaría de Redacción:</i>	María Carolina Álvarez Puerta
<i>Secretaría de Administración:</i>	Benjamín Sánchez
<i>Consejo Consultivo:</i>	Omar Astorga (UCV), Francisco Bravo (UCV) †, Luz M. Barreto (UCV), Sabine Knabenschuh de Porta (LUZ), Carlos Kohn (UCV), Vincenzo P. Lo Monaco, (UCV), Nancy Núñez (UCV), Benjamín Sánchez (UCV), Gustavo Sarmiento (USB), Fabiola Vethencourt (UCV), Corina Yoris (UCAB)
<i>Consejeros Internacionales:</i>	J. J. Acero (UGR), Francisco Ayala (UCLA), Mario Bunge (McGILL) †, Hugo Calello (UBA), Marisa Kohn de Beker (UCV) †, Pedro Lluberes (USB), M. Reyes Mate (CSIC), Jesús Mosterín (CSIC) †, Ulises Moulines (LMU), José M. Rosales (UMA), Giulio F. Pagallo (Unipd) †.
<i>Responsable de este número:</i>	Ricardo Da Silva
<i>Asistentes:</i>	María Alejandra Pérez y Sebastián Bogoljubskij
<i>Diseño de Portada:</i>	Lic. José Bifano
<i>Arte Final:</i>	María Alejandra Pérez

EPISTEME NS

Volumen 45, N° 1, 2025

Contenido Consideraciones analíticas

Presentación

DA SILVA, R.:	Presentación.....	<i>i</i>
---------------	-------------------	----------

Artículos

Págs.

ÁLVAREZ, M.:	Margaret Masterman, su crítica a la noción de paradigma y su invisibilización (<i>Margaret Masterman, her criticism of the notion of paradigm and its invisibility</i>).....	01
BARRETO, L.:	El papel de las heurísticas en el desarrollo de una inteligencia artificial general (<i>Heuristic's role on the development of a general artificial intelligence</i>).....	25
DA SILVA, R. Y RIVAS, S.:	El Dilema de Benacerraf como una objeción al Platonismo tradicional en matemática (<i>Benacerraf's Dilemma as an objection to traditional Platonism in mathematics</i>).....	49
HURTADO, N.:	Fundamentos técnicos y dilemas éticos en los modelos de lenguaje natural en las Inteligencias Artificiales (<i>Technical foundatinos and ethical dilemmas of natural language in Artificial Intelligence</i>).....	81

Marquez, L.:	An interpretation of the I/Me distinction in G. H. Mead's "The definition of the psychical" and H. Joas' criticism (<i>Una interpretación de la distinción Yo/Mí en "The definition of the psychical" de G. H. Mead y la crítica de H. Joas</i>).....	95
Valenzuela, E.:	¿Hay matemática más allá de la lógica? (<i>Is there mathematics beyond logic?</i>).....	121
<i>Notas y discusiones</i>		
Blanco, F.:	Notas introductorias al concepto de «Lógica» (<i>Introductory notes on the concept of «Logic»</i>).....	137
Chávez, M.:	¿Es la <i>existencia</i> una propiedad, o una <i>modalidad</i> ? (<i>Is existence a property, or a modality?</i>).....	153
<i>Recensiones</i>		
Bogoljubskij, S.:	Posy, C - Intuicionismo matemático (<i>Mathematical Intuitionism</i>).....	159
<i>miscelánea</i>		165

RICARDO J. DA SILVA ARAUJO

PRESENTACIÓN

Bajo la premisa de que la unidad filosófica no exige uniformidad temática, este número ofrece un panorama diverso donde la multiplicidad de problemas abordados cobra sentido bajo el rigor que el análisis filosófico confiere a la investigación académica actual.

Aunque es factible trazar conexiones entre ciertos núcleos temáticos que se comentarán más adelante, el verdadero denominador común de esta entrega es el compromiso de presentar los problemas desde una visión argumentativa que los sustente. De esta manera, cada texto aquí reunido destaca por la claridad conceptual empleada para desarrollar sus hipótesis, ofreciendo al lector un ejercicio de pensamiento crítico y estructurado.

Aunque el índice obedece a un orden lexicográfico, podemos reagrupar los artículos según ciertas similitudes. Los primeros serían aquellos que reflexionan directamente sobre la obra de un autor. Así, María Carolina Álvarez se propone revalorizar y visibilizar las críticas que Margaret Masterman le hiciese a la noción de “paradigma” de Thomas Kuhn en la edición de 1962 de *La estructura de las revoluciones científicas*; por su parte, Leopoldo Márquez propone una nueva lectura de la subjetividad en el trabajo temprano de George H. Mead, heredada luego por su obra madura, tratando así de sustentar una exégesis continuista sobre el pensamiento del autor.

En otra dirección, encontramos trabajos que giran en torno al tema de la inteligencia artificial, tales como su capacidad para capturar la semántica del lenguaje y los desafíos éticos que presentan estas nuevas herramientas. Luz Marina Barreto busca examinar el desarrollo de algunos métodos algorítmicos de procesamiento de lenguaje y explicar, especialmente, las limitaciones de los grandes modelos de lenguaje (LLM) para actuar como modelos de razonamiento sin la retroalimentación continua de anotadores humanos. Nahir Hurtado, por su parte, hace un recorrido elucidatorio por los mecanismos que han hecho

posible el surgimiento de modelos como GPT, pasando así revista a las redes neuronales y las arquitecturas transformers. Ambas autoras concluyen remarcando el énfasis en las consideraciones éticas y en las exigencias de legislación sobre el uso de estas nuevas tecnologías.

El último par de artículos trata sobre cuestiones inherentes a la filosofía de la matemática. Edgar Valenzuela revisa la persistencia de la lógica en el quehacer matemático a través del estudio de sistemas lógicos no clásicos, defendiendo, así, una tesis pluralista en contraposición al monismo lógico tradicional. A su vez, Solani Rivas y Ricardo Da Silva realizan una reconstrucción argumentativa del dilema de Benacerraf en contra del platonismo matemático clásico y presentan la propuesta de Mark Balaguer, el platonismo pleno, como un intento de solución.

El presente número se ve, además, enriquecido con dos notas ensayísticas y una reseña bibliográfica. La primera de las notas, escrita por Francisco Blanco, busca indagar en un concepto de lógica que se reconstruya desde sus compatibilidades e incompatibilidades semánticas con otros conceptos. En este sentido, el material puede servir de apoyo para las primeras sesiones de un curso de lógica en cualquier pregrado de humanidades. La segunda nota, de Marcel Chávez, revisa críticamente la idea de que la existencia es una propiedad de los individuos, es por ello que esboza una interpretación desde la ontología de Duns Escoto, desde la cual se concluye que la existencia no es una propiedad, sino una forma de modalidad. Finalmente, Sebastián Bogoljubskij reseña el texto *Mathematical Intuitionism* de Carl Posy. La obra reseñada forma parte de la colección *Cambridge Elements: The Philosophy of Mathematics* y tiene por objetivo introducir a los lectores en la filosofía intuicionista de la matemática desde consideraciones filosóficas y formales. Bogoljubskij reconoce la claridad analítica en la exposición propuesta por el autor; sin embargo, identifica que algunos capítulos, debido a su reconstrucción formal, no resultarían accesibles para estudiantes que no estén familiarizados con ciertos contenidos de lógica matemática.

Podemos concluir afirmando que la presente edición contribuye a la idea de que la *buena filosofía* —sea lo que sea— se sustenta en argumentos. Que resulte de provecho para todos.

LUZ MARINA BARRETO

EL PAPEL DE LAS HEURÍSTICAS EN EL DESARROLLO DE UNA INTELIGENCIA ARTIFICIAL GENERAL

Resumen: Este artículo examina el desarrollo de algunos métodos matemáticos de procesamiento de lenguaje natural a la luz de su capacidad para capturar la semántica del lenguaje. Menciono brevemente las máquinas de soporte vectorial de Vapnik y Chervonenkis, así como el cálculo de la adjudicación latente de Dirichlet de Blei y Lafferty, para compararlas a ambas con los métodos de los transformadores generativos de propósito general o GPTs, que definen los grandes modelos de lenguaje actuales. Seguidamente, explico las limitaciones de los Llms para actuar como grandes modelos de razonamiento sin la retroalimentación continuada de anotadores humanos. La dependencia de estos modelos del input humano, que se evidencia en particular en los métodos de optimización que se conocen como “*fine tuning*” o ajuste fino de estos modelos, así como su dependencia general de la investigación en métodos de optimización de funciones que existen ya desde mediados del siglo XX, apuntan a las posibles limitaciones de los modelos grandes de lenguaje, e incluso de los modelos grandes de razonamiento, para alcanzar los estándares de una inteligencia artificial real. En el párrafo 4 hipotetizo que estas limitaciones en la implementación de métodos de optimización en modelos grandes de lenguaje es un problema profundo en el modelado geométrico-analítico de grandes conjuntos de datos, cuyo tratamiento estadístico no puede evitar problemas de sesgo y los causados por la multidimensionalidad de los atributos de un modelo. Aquí sugiero entonces que la AGI o inteligencia artificial general, si quiere parecerse a los modos como los seres humanos generalizamos inferencias a través de distintos dominios de objetos, debería poder crear heurísticas y no limitarse a optimizar funciones sobre un conjunto de datos dentro de un

dominio dado. Finalmente, reflexiono sobre lo que todo esto significa para la presunción de que las actuales máquinas de inteligencia artificial puedan sustituir el trabajo humano o convertirse en interlocutores válidos de un entorno humano caracterizado por una comunicación ética.

Palabras clave: Análisis semántico automatizado, modelos grandes de lenguaje y razonamiento, alcances y límites de la IA.

HEURISTIC'S ROLE ON THE DEVELOPMENT OF A GENERAL ARTIFICIAL INTELLIGENCE

Abstract: This article examines the development of certain mathematical methods in natural language processing regarding their capacity to capture language semantics. I briefly mention Vapnik and Chervonenkis's support vector machines and Blei and Lafferty's latent Dirichlet allocation to compare them with generative pre-trained transformers (GPTs), which define today's large language models. Subsequently, I explain the limitations of LLMs in serving as large reasoning models without ongoing feedback from human annotators. The reliance of these models on human input—seen especially in "fine-tuning" optimization methods—and their general dependence on function optimization research from the mid-20th century highlights potential barriers for these models to achieve true artificial intelligence. In paragraph 4, I hypothesize that these implementation limitations are a deep issue in the geometric-analytical modeling of large datasets, where statistical treatment cannot escape problems of bias and multidimensionality. I suggest that for AGI (Artificial General Intelligence) to mimic human-like inference generalization across different domains, it must be capable of creating heuristics rather than just optimizing functions within a single domain. Finally, I reflect on whether current AI can truly replace human labor or act as valid interlocutors in environments defined by ethical communication.

Keywords: Automated semantic analysis, large language and reasoning models, scope and limitations of AI.

1. Antecedentes teóricos de las máquinas actuales de IA entendidas como herramientas de procesamiento de lenguaje natural.

Los modelos grandes de lenguaje o Llms, *large language models*, son una innovación de las herramientas de procesamiento del lenguaje natural y, por lo tanto, no puede decirse que sean modelos de razonamiento en sentido estricto. Si nos preguntamos qué pueden o no pueden hacer los modelos de lenguaje grandes como ChatGPT, Claude, Gemini, Deepseek, Llama de Meta, etc., la gran pregunta es, entonces, si pueden imitar nuestros modos de razonamiento humano, es decir, si son inteligentes de verdad, si el término “inteligencia artificial” en relación con ellos está justificado.

Como lo escribí en mi tesis de maestría sobre métodos matemáticos y algorítmicos de análisis de la semántica latente, que puede consultarse en el repositorio de la Universidad Central de Venezuela “Saber UCV”,¹ desde el siglo pasado se ha trabajado mucho en la búsqueda de métodos para detectar la semántica latente en corpus de textos digitalizados, con el fin de emular en lo posible la comprensión semántica humana con la ayuda de modelos de aprendizaje automático o *machine learning*. Todos estos métodos requieren, en principio, “tokenizar” o dividir oraciones en largas listas de palabras y procesar estas listas en matrices susceptibles de tratamiento con diversos métodos de álgebra lineal, para encontrar las palabras dentro de oraciones que son más significativas para entender la semántica de la oración o del texto que se somete a este tipo de preprocesamiento. Se dice, entonces, que estos métodos buscan la semántica “latente”, en el sentido de que un largo texto puede tener un significado susceptible de ser resumido en una serie de palabras o tópicos clave, lo que lo volvería más manejable para ser analizado o transmitido a través de sistemas de información.

Muchos matemáticos a mediados del siglo XX pensaron en algoritmos ingeniosos para develar la semántica latente de una matriz de términos o palabras. A mediados del siglo XX, el matemático ruso Vladimir Vapnik y Alexey Chervonenkis, por ejemplo,

¹ Barreto, L., *Estudio de dos paradigmas de modelado de tópicos en un corpus de documentos tomados de una red social*, 2018. Disponible en: <https://saber.ucv.ve/handle/10872/20186>.

desarrolló un método famoso llamado “máquinas (o funciones) de soporte vectorial” o SVM, que consistía en calcular lo que se conoce en álgebra lineal como la “forma espectral de una matriz” de listas de palabras, o dicho de otro modo los autovectores o autovalores de estas matrices, a fin de calcular una suerte de valor “medio” o “línea media” que diera cuenta de las palabras más importantes o centrales en una oración.

Otro método ingenioso que se desarrolló a finales del siglo pasado y comienzos de este consistía en calcular con una aplicación del teorema de Bayes la probabilidad de que una palabra fuera la realmente importante dentro de un corpus. Este método fue desarrollado por David Blei y John Lafferty como la “Adjudicación o Atribución Latente de Dirichlet”. Esta calcula la probabilidad posterior de que una palabra sea importante dentro de un corpus dado. Blei corrió este algoritmo para un corpus conformado por 100 años de la revista *Science* y así pudo determinar qué palabras, tópicos o conceptos fueron importantes durante cada década durante ese lapso de 100 años.² Este experimento fue realizado en la primera década de este siglo y con el poder computacional disponible Blei y colaboradores tardaron unos cuatro días en hacer converger el algoritmo, a un costo considerable: 40.000\$ en AWS, si mal no recuerdo. Los resultados son muy impresionantes y sugerentes: en el año en que mi mamá se graduó de Bioanalista en la Universidad Central de Venezuela los estudios hematológicos eran el tópico más importante de la ciencia de la época y durante los años en que mi papá estudió Ingeniería Eléctrica en la UCV, el tópico de moda eran las aplicaciones de teoría de potencia eléctrica. De esta manera, los tópicos o conceptos más importantes en los artículos de la revista *Science* a lo largo de los años reflejaban lo que la comunidad académica y universitaria de la época iba considerando año tras año como los temas más importantes de una investigación científica actualizada.

Los actuales modelos de lenguaje grandes o Llms son un paso adelante en esta búsqueda para mejorar el procesamiento automatizado del lenguaje natural. Las innovaciones actuales se apoyan en el aumento de nuestro poder computacional. Uno de estos

2 Blei, D y Lafferty, J., *Modeling Science*, 2008. Disponible en: <https://es.slideshare.net/slideshow/modeling-science/10727180>.

algoritmos, desarrollado por Geoffrey Hinton y colaboradores, también durante el siglo pasado, son las redes neuronales. Se trata de un algoritmo que va corrigiendo los errores en la salida a través de la continua supervisión de una función matemática de pérdida o error, que se autocorrije iterativamente con un método de retro-propagación o “backpropagation” de los parámetros correctos, que se “devuelven” por el gradiente para hacerlos coincidir con los correctos si se desvían demasiado de las funciones objetivo. Dicho en lenguaje palatino: el modelo sabe cuál es la salida correcta o esperada y si en su búsqueda automatizada se equivoca, se “devuelve” para buscar de nuevo el camino correcto, que en su caso son los parámetros correctos.

En la medida en que los modelos de lenguaje grandes son ahora una implementación de las redes neuronales para el procesamiento de lenguaje natural, es una instancia de métodos del aprendizaje llamados “supervisado” y “autoreforzado”, en donde ya el modelo sabe cómo se usa correctamente el lenguaje de modo que parezca inteligible, y de lo que se trata es que esa estructura, que sabemos cómo es cuando está bien hecha, se reproduzca correctamente cuando la máquina de IA vaya construyendo una respuesta en lenguaje natural. Las redes neuronales son un método de aprendizaje supervisado porque el objetivo del modelo o su función objetivo ya se conoce (se trata de una oración correctamente construida) y lo único que hace falta es que el modelo encuentre el camino. Los Llms, o los grandes modelos de lenguaje que se consideran hoy en día “modelos de IA”, son un paso más en una aplicación de una red neuronal que implementa un método de aprendizaje auto-supervisado o por refuerzo (RL), en el sentido de que el modelo por sí solo va aprendiendo, viendo muchos ejemplos, cómo se estructura una oración correcta en cualquier idioma.

Matemáticamente este tipo de modelo calcula fundamentalmente la probabilidad de que una palabra siga a otra en una oración correctamente construida y, por lo tanto, es un poco exagerado llamarlos métodos de IA. El algoritmo de David Blei que vimos arriba y que calcula la adjudicación latente de Dirichlet para la distribución de probabilidad de la semántica latente de un corpus de texto también calcula probabilidades y no se consideraba un análogo de la inteligencia humana, por la razón sencilla de que nosotros no nos limitamos a calcular

probabilidades cuando razonamos sobre un problema, aunque podemos hacerlo perfectamente. Ahora bien, calcular la probabilidad de que una palabra siga a otra cuando se conoce la estructura de un lenguaje bien construido es incluso más fácil de hacer que encontrar la semántica latente de un corpus de texto, y ambos cálculos dependen de que el lenguaje tenga una estructura que se pueda discretizar o “tokenizar”, como se dice técnicamente, en sus elementos individuales o factores. De este modo, es siempre posible calcular la probabilidad de las secuencias de estos símbolos discretos. Pero, como dice Yann LeCun,³ ingeniero y científico de datos que trabajó durante muchos años en Meta como investigador, esto puede hacerse porque la estructura de una oración o de un conjunto de oraciones bien hechas es relativamente fácil de reconstruir con suficientes datos de entrenamiento y poder computacional. Esto último es importante, porque la generación de esta estructura es costosa computacionalmente y requiere un consumo importante de energía: un típico modelo de lenguaje puede llegar a tener 20 billones o millones de millones de tokens o 20×10^{12} , si se toma en cuenta más o menos la información que ha sido digitalizada hasta la fecha. Para recrearla, lo que estos modelos hacen, de acuerdo con LeCun, es generar millones de salidas aleatorias y las van probando, hasta que todas convergen en la función matemática objetivo, que es la oración correcta. Estos modelos son modelos autoregresivos o GPTs: *general purpose transformers*, transformadores de propósito general. En la medida en que se trata de GPTs o transformadores generativos de propósito general, los Llms generan millones de ejemplos de posibles oraciones correctas hasta que el algoritmo converge en las estructuras que tienen una mayor probabilidad de parecerse a las que producen los seres humanos reales.

Sin embargo, es claro que no es así como razona un ser humano. Nosotros obviamente no producimos millones de ejemplos de oraciones posibles hasta que nuestro cerebro converge en las que más probablemente se parecen a las correctas. Por otro lado, los Llms o modelos grandes de lenguaje son inútiles sin pasar por un proceso de

3 "Father of AI: AI Needs PHYSICS to EVOLVE | prof. Yann LeCun." YouTube, uploaded by This Is World, 22 May 2024. Disponible en: <https://www.youtube.com/watch?v=L2Gj3hJ8y14>.

optimización que acerque sus respuestas a los modos como razonan los seres humanos. Esta tarea se conoce como “refinamiento o *fine tuning*” del modelo de base y no puede hacerse sin que se deba informar o “instruir” continuamente al modelo con respuestas dadas por seres humanos reales.

2. Un GPT no es automáticamente un modelo de razonamiento humano.

En efecto, en la medida que el modelo de base tiene que ser ajustado o refinado para que use el lenguaje *como si estuviera razonando* como un ser humano, estas máquinas de inteligencia artificial, que se apoyan solamente en la capacidad de reproducir la estructura correcta de un lenguaje, no pueden hacer nada por sí solas si no poseen grandes cantidades de información que representen de modo correcto o preciso el mundo. Para que esto suceda, los Llms tienen que funcionar como LRMs o *large reasoning models*, “grandes modelos de razonamiento”, los cuales deberían poseer una arquitectura diferente de los modelos grandes de lenguaje de los que disponemos actualmente. La diferencia entre uno y otro es que los grandes modelos de razonamiento, como los Llms “tuneados” o refinados, también requieren mucha información producida, etiquetada o anotada por seres humanos reales. De este modo, detrás de los modelos de lenguaje grande que ofrecen respuestas correctas a una pregunta hay un batallón de seres humanos reales etiquetando imágenes o corpus de texto, produciendo conocimiento a través de blogs y páginas web, calificando respuestas correctas, instruyendo a los modelos sobre como eludir respuestas tóxicas o que pongan en peligro la seguridad del usuario y, en general, puliendo las respuestas que dan los Llms.

Todos los modelos de lenguaje y razonamiento grande pagan a miles de usuarios reales, por lo general gente joven en países fuera de la regulación laboral que obliga a las grandes corporaciones digitales a pagar un salario mínimo a los contribuyentes fiscales dentro de EEUU, para etiquetar o anotar los datos que luego usa el modelo para entrenar el algoritmo y producir la salida u output deseado. Esta práctica se conoce como *crowdsourcing* y yo diría que conforma el 50% de lo que hace posible que esos modelos existan (el otro 50% son las

soluciones de implementación de algoritmos que crean los científicos e ingenieros que trabajan para esas compañías y la existencia de una infraestructura de hardware adecuada). Estos etiquetadores incrementan la eficacia de un método de refinamiento o ajuste de los modelos que se llama RLHF, *reinforcement learning from human feedback*, aprendizaje reforzado por retroalimentación humana. En la máquina de IA de Google, por ejemplo, Gemini, a ellos se les agrega la ventaja que ofrece poseer la mayor base de datos de la industria.

Todos estos modelos son, pues, en el fondo, parasitarios de la propiedad del lenguaje humano para ser portador de conceptos generales que simbolizan semánticamente nuestro mundo. La estructura del lenguaje natural no representa por sí sola el mundo sin estos conceptos y por eso, por sí solos, los Llms no pueden decir nada que tenga sentido para un usuario real. Necesitan información, inmensas bases de datos, gente real diciendo cómo es el mundo. Las oraciones, textos y contenidos que generan los Llms son por ello dependientes de anotaciones humanas, es decir, de que miles de personas le enseñen a un modelo de lenguaje no a representar el mundo, porque es un tema abierto de la filosofía entender cómo es que nosotros podemos hacer esto, sino a construir oraciones que *parezcan* que hacen esto o se acerquen un poquito a este estándar.

Esto explica, en parte, por qué los grandes modelos de lenguaje arrojan resultados extrañamente confusos y muchas veces incorrectos. Dependen de la calidad del ajuste fino basado en el aprendizaje auto-reforzado con retroalimentación humana, de la amplitud de su base de datos, y también de la capacidad que tenga el modelo para hacer inferencias válidas que capturen la riqueza del contexto desde donde un usuario dado hace una pregunta. Esto le permite escalar sus salidas al nivel de un LRM, un *large reasoning model*, un modelo grande de razonamiento. Modelos de este tipo tienen enormes bases de datos y por ello hay muy pocos en el mundo que alcancen estos estándares. Es tal la sed de datos digitalizados que necesitan esos modelos que nos hackearon hace poco la base de datos de la plataforma de publicaciones académicas de acceso libre de la Universidad Central de Venezuela, “Saber UCV”, para poder acceder a nuestras publicaciones científicas digitalizadas.

Gemini es uno que cuenta con toda la base de datos de Google y muchísimo poder computacional para generar el contexto que resulta de la base de información que ya se tiene sobre el usuario, además de otros datos de tipo demográfico. Esos algoritmos son un secreto corporativo, pero es posible imaginar que, con mucho poder computacional, la máquina o función de IA puede simultáneamente actualizar de forma paralela toda esa información para generar inferencias válidas sobre la pregunta que hace un usuario y qué es lo que éste quiere realmente saber. Por eso es que yo creo que el presidente de los EEUU, Donald Trump, quiere crear un consorcio para la inversión de 500 mil millones de dólares en sistemas de IA. Tanta computación no es para entrenar un Llm, que no la necesitaría en todo caso. Es para hacer *inferencias* sobre cada uno de los usuarios que interactúa con una IA. Yo me imagino que para eso es también necesario poseer Groenlandia, un lugar helado en donde, supongo, no se incendiarán todo el tiempo, como sucede ahora, los enormes repositorios de servidores que generan calor y consumen energía en tareas de computación.

Pero, ¿qué sé yo?

3. Los LRMs existen desde hace décadas y se apoyan en métodos de optimización de datos.

¿Qué distingue exactamente un modelo de lenguaje grande y un modelo grande de razonamiento? Todos los grandes modelos de lenguaje entrenan sus algoritmos de métodos de aprendizaje automático supervisado en base a posts “salvajes” o desordenados que usuarios de redes sociales han subido voluntariamente a la red. Piénsese, por ejemplo, en un post de Facebook o en un comentario en X. Muchos de estos contenidos no se caracterizan precisamente por ser cuidadosos, estar bien escritos o ser incluso realmente “inteligentes”. Por si fuera poco, es probable que los usuarios que suben contenido sean una minoría, la cual, por la ley de Pareto, no superaría el 20% de los usuarios de la red, mientras que el resto, un 80% no escribe mucho o casi nada, para no mencionar que la inmensa mayoría de la información humana no ha sido digitalizada todavía. Este es un punto que algunos científicos han observado antes, pero que no se suele mencionar mucho en voz alta, dado que estas plataformas viven del valor de sus acciones, del

ingreso prospectivo en publicidad y de la posibilidad de vender datos de usuarios para estimaciones estadísticas de distinto tipo (con fines de manipulación o propaganda política, investigación social, etc.).

Por esta razón, todos estos modelos de base o “salvajes” necesitan ser refinados, “*fine-tuned*” o ajustados de modo fino, como decía más arriba. Este trabajo de refinamiento es el verdadero esfuerzo que deben hacer todas estas empresas para que sus modelos grandes de lenguaje sean realmente útiles para un usuario, en el sentido de que más o menos razonen como un humano cuando un usuario interactúa con él.

Y aquí viene el núcleo de mi argumento en relación a las capacidades reales de razonamiento de una máquina de inteligencia artificial. Refinar uno de estos modelos para que “hable bien”, o parezca que razona como un experto cuando uno interactúa con ellos, requiere un proceso matemático que se conoce como optimización de funciones matemáticas para el procesamiento de datos. Este campo de estudio es el resultado de décadas de trabajo sistemático de profesores e investigadores en los campos teóricos de los fundamentos de la matemática, la geometría analítica y la ciencia de la computación. Por esta razón, no tiene mucho sentido decir que la IA va a sustituir el trabajo humano y eliminar empleos, dado que detrás de todas estas tecnologías existe un batallón de científicos, investigadores e ingenieros cuyas intuiciones y experimentos, la mayoría de los cuales requieren una inversión enorme en educación e infraestructura para la investigación, son requeridos continuamente para mantener las arquitecturas de software y la infraestructura de hardware que posibilitan la existencia de estos modelos y sistemas.

La lógica subyacente en el análisis de funciones de optimización es la de una función objetivo que queremos optimizar para una serie de recursos o conjuntos de datos x . Los métodos de optimización son una rama enorme y muy estudiada de la matemática que se usa desde hace muchas décadas en finanzas, adjudicación de recursos para industrias o políticas de estado y en el diseño de sistemas que requieren control autónomo. Métodos de optimización como la programación lineal y el método Simplex, desarrollados por el matemático del siglo XX George Dantzig, al inicio no tenían nada que ver con computadoras, sino que simplemente ayudaban a adjudicar de modo eficiente distintos recursos

escasos para obtener de ellos el máximo rendimiento posible. Estos métodos de optimización, que han venido evolucionando, hacen uso ahora del cada vez mayor poder computacional disponible para resolver problemas prácticos tales como el diseño de la ruta más económica y eficiente para la flota de repartidores de una empresa de transporte como FedEx o Amazon, para distribuir de modo eficiente en una ciudad y en un país los recursos de energía eléctrica o hidrológicos con los que se cuentan, en la restitución eficiente de inventarios para empresas comerciales, en la asignación de pilotos y aviones en rutas de aviación comercial, en la maximización de cirugías, cuartos y camas disponibles en clínicas y hospitales, y de modo muy importante, en conflictos bélicos. El estudio de estos métodos de optimización forma una rama de análisis de procesos cuyas aplicaciones florecieron durante la Segunda Guerra Mundial para resolver problemas de táctica y estrategia bélica.

En el ámbito de análisis de conflictos bélicos, por ejemplo, no sólo es importante saber dónde colocar suministros, municiones, combustible y comida, de modo eficiente y minimizando costos, a un ejército desplegado en un país extranjero, sino también cómo planificar misiones y seleccionar objetivos causando el mayor impacto con el menor riesgo y costo para quienes atacan. Viene a la mente el reciente ataque a Caracas del 3 de enero, una acción que, comparada con otras acciones parecidas realizadas por los EEUU en suelo extranjero, se caracterizó por su óptima elección de objetivos, minimización del riesgo de pérdida de combatientes y eficiente neutralización de los recursos defensivos del enemigo, que en ese caso eran nuestros soldados y los equipos defensivos con los que contaba nuestra fuerza nacional bolivariana. Pero ya durante la Guerra del Golfo, en 1991, se usó una herramienta de IA llamada DART (*Dynamic Analysis and Replanning Tool*)⁴ en lo que se considera uno de los mayores ejercicios de optimización bélica de la historia. Esto pone de relieve que la IA no se reduce a lo que puede hacer un Llm, sino que expresa nuestra capacidad para aplicar algoritmos de optimización en contextos diferentes. No es una propiedad de la estructura sintáctica del lenguaje como tal, sino de nuestra capacidad para aplicar herramientas de análisis matemático en procesos de optimización de funciones matemáticas para el mejor uso, comprensión y

4 Véase https://en.wikipedia.org/wiki/Dynamic_Analysis_and_Replanning_Tool.

distribución de los recursos con los que contamos, incluyendo conjuntos de datos, como veremos enseguida.

Las tareas de refinamiento de un modelo de lenguaje grande, por su parte, se optimizan a través de su input humano. Hay que enseñar a estos modelos, con ejemplos, como es que razona un ser humano, porque por su cuenta no lo pueden hacer. En este sentido, la eficacia real de estos modelos es parasitaria de la inteligencia de millones de *crowdworkers*, en su mayoría jóvenes, que, por una remuneración mísera, enseñan a los modelos a responder de forma humana o adecuada a las preguntas de los usuarios. Este refinamiento o ajuste de los modelos se hace con un método de aprendizaje reforzado con retroalimentación humana conocido como “optimización de políticas” o de “estrategias” que maximizan el resultado esperado o deseable. Los métodos pioneros para la optimización de Llms son los creados por los investigadores de OpenAI John Schulman y colaboradores como el *Trust Region Policy Optimization*,⁵ y una variante del mismo, llamada *Proximal Policy Optimization*.⁶ Ambos consisten en ir aproximando la estrategia deseada acercándola poco a poco al ejemplo que se le ha dado.

Este tipo de ajuste fino es especialmente relevante cuando se quiere evitar que el Llms dé respuestas tóxicas o que pongan en peligro la seguridad del usuario o la responsabilidad que tiene la empresa frente al usuario. Las grandes corporaciones tecnológicas que gestionan grandes modelos de lenguaje hacen todos los días esfuerzos para evitar que atacantes ingresen al sistema para generar contenido tóxico o irresponsable que ponga en peligro la vida de un usuario vulnerable o sugestionable. Estos esfuerzos requieren que se le dé al modelo miles de ejemplos de respuestas adecuadas, las que se conocen como “*ground truths*” o verdades de base, las cuales tienen que redactar muchas veces expertos o seres humanos reales. Esto es difícil de hacer porque para que un Llm de base aprenda a responder correctamente, lo que lo

5 Schulman, J., Levine, S., Abbeel, P., Jordan, M., & Moritz, P. 2015. Trust Region Policy Optimization. *Proceedings of the 32nd International Conference on Machine Learning*, 37, 1889–1897. Disponible en: <https://proceedings.mlr.press/v37/schulman15.html>.

6 Schulman, J., Wolski, F., Dhariwal, P., Radford, A., & Klimov, O. 2017. Proximal policy optimization algorithms. *arXiv preprint arXiv:1707.06347*. Disponible en: <https://arxiv.org/abs/1707.06347>.

eleva a lo que en la industria se conoce como un modelo “instruido”, se requiere, por una parte, entornos computacionales de alto desempeño y, por la otra, trabajadores que redacten las respuestas deseadas, lo cual tampoco es barato (si se les paga a anotadores, a través de plataformas diseñadas para ello como Prolific, unos pocos dólares para redactar o anotar respuestas correctas, para 1200 trabajadores la cuenta pudiera ser elevada).⁷

En todo caso, los métodos de optimización en sistemas informáticos se investigan y existen desde hace mucho tiempo. Gracias a ellos, se puede pagar un servicio público a través de un botón dentro de la interfaz de un banco, un botón que vincula la página del Banco con un API que ofrece la gestora del servicio público. Lo que es nuevo ahora es que se pueda hacer eso usando lenguaje natural, pudiendo decir a la página del banco: “paga por favor el agua desde alguna de mis cuentas”. Hay muchas cosas que se pueden hacer de este tenor, como pedirle a un agente de IA que reserve y compre un pasaje de avión, que redacte un email corporativo, que escriba un reporte de investigación con la información que uno tiene en la computadora, que llene un formulario, que te corrija un script de programación, que envíe distintos posts a redes sociales, etc., pero todo ello se puede hacer desde hace décadas interactuando con interfaces que ofrecen acceso a bases de datos a través de APIs, es decir, sin usar el lenguaje natural. Poder interactuar ahora con esas bases de datos usando lenguaje natural es lo nuevo y esto pudiera ser conveniente porque sin duda ahorramos tiempo, o por si un usuario de edad avanzada tiene dificultades para interactuar con botones en una interfaz de computadora.

Por estas razones, me parece inadecuado llamar a la implementación de métodos de optimización para el uso de sistemas informáticos una forma avanzada de inteligencia automatizada. Esta es una tecnología de aprendizaje automático o *machine learning* que ya existía antes de los GPT o transformadores generativos. Con estas tecnologías se han identificado desde hace tiempo patrones patológicos en

7 Véase por ejemplo, el caso del uso de un benchmark de Llms como EmotionBench: Huang, J.-T., Wang, W., Feng, F., Duan, C., Li, Z., & He, X. 2023. Emotionally numb or empathetic? Evaluating how LLMs feel using EmotionBench. *arXiv preprint arXiv:2308.03656*. Disponible en: <https://doi.org/10.48550/arXiv.2308.03656>.

imágenes médicas, se pueden traducir textos o hacer resúmenes, reconocer imágenes o rostros, hacer recomendaciones a clientes potenciales desde una plataforma de *e-commerce* o construir vehículos autónomos. No obstante, no es una forma de inteligencia artificial en sentido estricto porque todas estas tecnologías se limitan a optimizar una función matemática en un espacio de datos dado.

4. ¿Pueden las máquinas de inteligencia artificial de hoy en día, entonces, sustituir el trabajo humano?

Para responder esta pregunta no sólo es necesario constatar que, antes bien, los actuales modelos de lenguaje y razonamiento requieren crucialmente de retroalimentación humana y de la capacidad de implementar algoritmos de optimización para gestionar información digital y la interacción eficaz entre sí de sistemas de tecnología de la información. También es necesario que nos preguntemos si esas capacidades pueden emular realmente la inteligencia y el razonamiento humano. No es suficiente que nos demos cuenta de cómo estos sistemas usan razonamiento humano para funcionar correctamente porque esto dejaría abierta todavía la pregunta de si alguna vez puede sustituirse del todo la retroalimentación humana para dar paso a sistemas autónomos de inteligencia artificial.

En otras palabras: hemos visto que las capacidades actuales de los así llamados sistemas de IA requieren trabajo humano intensivo. De este modo, la pregunta es realmente: ¿será posible prescindir eventualmente de esos trabajadores cuando se alcance un cierto umbral de eficiencia?

Para responder esta pregunta hay que avanzar un paso más en el análisis de los métodos de estos sistemas de IA. Hemos visto que son eficientes para modelar un ámbito dado de aplicación, cuyas funciones descriptivas optimizan. Pero la pregunta crucial es si pueden hacer lo que nosotros hacemos todo el tiempo, a saber, relacionarnos racionalmente no con un solo dominio de objeto, sino con varios a la vez, trasladando las intuiciones válidas de uno a otro. Tomando en cuenta la visión que he esbozado anteriormente, esta pregunta interroga realmente sobre nuestra capacidad para generalizar nuestros métodos de optimización para que no queden restringidos a un sólo campo de

implementación, como cuando adaptamos una función matemática a un conjunto de datos particular. Es decir, pregunta si es posible desarrollar una tal cosa como una AGI, una inteligencia artificial general, lo que la aproximaría a nuestros modos humanos de razonar sobre el mundo.

Los seres humanos podemos generalizar nuestra comprensión del mundo sin problemas, aplicándola a través de distintos dominios de objeto. El mejor ejemplo que se me ocurre para ello es contarles cómo me parece que el príncipe Louis, el hijo menor de los actuales príncipes de Gales, Guillermo y Katherine, aprendió a saludar a una multitud desde el balcón del Palacio de Buckingham. Cuando Louis tenía unos tres años, vi por televisión cómo saludaba al pueblo inglés desde la balaustrada de piedra de ese balcón. Con los dos codos apoyados firmemente sobre ella, movía los brazos como si fuera el parabrisas de un Range Rover, balanceándolos de un lado para otro y cruzándolos entre sí con las palmas de sus manos hacia el frente. Es posible imaginar que no debe ser fácil enseñar a un principito de tres años a mover adecuadamente los brazos para saludar a su pueblo y es también posible imaginar que alguien pudo haberle comentado que el movimiento que debía hacer se parecía un poco al movimiento de las aspas de un parabrisas cuando despejan la lluvia que cae sobre el vidrio delantero de un vehículo.

Esta relación, entre movimiento de un parabrisas y saludo al pueblo, no la puede hacer espontáneamente ninguna máquina de inteligencia artificial a menos que se la entrene específicamente para ello. Un sistema de aprendizaje automático sin duda puede reconocer cómo mueven los brazos los miembros de la realeza cuando saludan desde un balcón, pero la relación con el parabrisas de un automóvil no la puede por sí sola. Sin embargo, el pequeño príncipe Louis no tiene problemas para establecer este tipo de relaciones entre contextos, para comprender y crear estas heurísticas, y aplicarlas con éxito en la tarea de aprender a saludar apropiadamente a la multitud desde el balcón del palacio. Esto es lo que significa trasladar nuestra comprensión del mundo de un dominio de aplicación a otro. Y eso lo podemos hacer los seres humanos desde muy temprana edad.

Si un sistema de IA pudiera hacer esto significaría que estamos ante un AGI, *artificial general intelligence*, una inteligencia artificial general, el santo grial de todas las corporaciones digitales que andan tras su posible desarrollo.

Nuevamente, la idea de que la inteligencia es una capacidad para trasladar soluciones entre ámbitos diferentes es también antigua, como los algoritmos de optimización. En la investigación sobre por qué una persona puede ser muy inteligente y otra no, que dio lugar a todos los tests que miden inteligencia o IQ, esta capacidad se conoce como el “factor G (o general)”. Definido por el psicólogo británico Charles Spearman en 1904,⁸ el Factor G especifica lo que se considera inteligencia, que es la capacidad para encontrar relaciones entre fenómenos complejos en dominios distintos. Spearman pensaba que ser inteligente no es saber mucho o moverse bien dentro de un ámbito particular, sino poder resolver problemas relacionando soluciones que han sido efectivas en un ámbito y trasladándolas a otro. Esto no lo puede hacer por sí sólo un sistema de aprendizaje automático y, de hecho, no sabemos muy bien cómo lo podemos hacer nosotros. El filósofo y teólogo Bernard Lonergan llamaba en 1957 a esta capacidad de generalización una capacidad para crear *heurísticas*, las que permiten a los seres humanos comprender algo a través de analogías y ejemplos que se toman de muchos ámbitos diferentes, iluminando el entendimiento humano sobre un tema o problema de una manera que no podemos explicar muy bien todavía y que no podemos reproducir en sistemas de aprendizaje automático e IA.⁹

La incapacidad de un sistema de aprendizaje automático para generalizar sus soluciones de modo que se apliquen en muchos ámbitos diferentes, o para crear una heurística, una capacidad que sería relevante si queremos que un robot sustituya a un ser humano en tareas delicadas, es, pues, el problema de fondo en toda esta discusión.

8 Véase Spearman, Ch., “*General Intelligence, Objectively Determined and Measured*”, 1904.

9 Lonergan, B. J. F., *Insight: A Study of Human Understanding* (F. E. Crowe & R. M. Doran, Eds.; Vol. 3). University of Toronto Press, 1992. (Obra original publicada en 1957).

4.1. El problema del intercambio sesgo/varianza

En mi opinión, se trata de un problema profundo que hunde sus raíces en nuestros métodos matemáticos y, en particular, estadísticos, de modelado del mundo en espacios geométrico-analíticos.

La primera dificultad que viene a la mente se conoce como el dilema de la compensación o intercambio sesgo-varianza, que fue introducido por primera vez en 1992 por Stuart Geman y colaboradores del MIT en su artículo seminal: “Neural Networks and the Bias/Variance Dilemma”.¹⁰ El dilema del intercambio sesgo-varianza dice que mientras más datos y atributos hemos mapeado, o sustituido por variables cuantitativas, o parametrizado, en nuestro modelo matemático del mundo, más sesgo tendremos, lo que se conoce también como un “sobreajuste” u *overfitting* del modelo. Mientras más preciso es el modelado en una función de un conjunto de datos, mejor será el modelo, pero, por eso mismo, su capacidad para ser generalizado, es decir, para poder ser aplicado a casos y objetos no vistos, será menor. Por el contrario, si queremos adaptar el modelo para que abarque casos no vistos o se pueda aplicar a conjuntos de datos diferentes, se supone que deberíamos poder mapear los datos con una mayor varianza, pero en ese caso es evidente que el modelo será menos efectivo. Dicho de otro modo: en un esfuerzo para analizar las características de los estudiantes en un salón de clases, para dar un ejemplo, si quisiera abarcar con una función matemática las propiedades de todos ellos tendría que elegir las más generales o las que más se acercan a una distribución normal: edad, sexo, estatura, etc. Pero una función de este tipo no será muy precisa o informativa. Si quisiera hacer más preciso mi mapeo de las propiedades de cada una de las personas que conforman la clase y tomar en cuenta no sólo estas características generales, sino datos más precisos, por ejemplo, sus libros y películas favoritas, en dónde viven en Caracas, cómo se trasladan a la universidad si en bicicleta, en transporte público o en carro, etc., más sesgada será la función resultante, dado que mapeará con precisión los atributos de un grupo de individuos dejando fuera a otros que no comparten las características

10 Geman, S., Bienenstock, E., & Doursat, R., Neural networks and the bias/variance dilemma. *Neural Computation*, 4(1), 1992, 1–58. Disponible en: <https://doi.org/10.1162/neco.1992.4.1.1>.

de los miembros que sí las exhiben. Se trata de un intercambio o compensación inevitable que es un resultado de nuestros modos de tratar estadísticamente datos y que se conoce en la literatura científica como la *compensación sesgo-varianza*.

4.2. La maldición de la multidimensionalidad.

El segundo problema es éste. La compensación sesgo/varianza parece ser un caso especial de un problema mayor que Richard Bellman definió, en 1957, como la “maldición de la dimensionalidad” o, tal vez más claro todavía, de la multidimensionalidad,¹¹ otro problema de optimización según el cual mientras más dimensiones o propiedades del modelo tenemos que definir con una función, más intratable se vuelve el problema computacionalmente. Por lo general, el valor de una función matemática se encuentra buscando sistemáticamente en un plano geométrico cuadrículado. Si calculamos para definirla cien puntos de datos que hay que chequear para cada dimensión o atributo relevante para el modelo, para dos dimensiones tendríamos 10.000 puntos que chequear o 100^2 y para apenas 10 dimensiones, o atributos, tendríamos que chequear 100^{10} o 10^{20} lo que excedería las capacidades de cualquier computadora, por más masivo que sea su sistema de computación paralela, dado que se trata aquí de buscar una función con algoritmos de fuerza bruta. Esto sucede así porque el crecimiento de los datos es exponencial; la complejidad del problema se incrementa por un factor de 2^{90} ,¹² que es un número enorme, dado que hay que calcular también cómo las distintas dimensiones interactúan o afectan los grados de libertad de la función que estamos buscando, lo que vuelve al problema un cálculo factorial. Calcular una tal función tardaría más tiempo que la duración del universo conocido porque es un problema de dificultad NP, es decir, no puede hacerse en tiempo polinomial, lo que significa que no tenemos un algoritmo ni un truco matemático, como los que describí al principio, o de optimización, como los que también describí, para resolverlo.¹³

11 En Bellman, R., *Dynamic Programming*, Princeton University Press, 1957.

12 Según la información dada por Gemini: Google. 2026. *Gemini* (versión del 8 de febrero) [Modelo de lenguaje grande].

13 Por cierto, que hay un premio de un millón de dólares pendiente para quienes pue-

Pero que nuestros mejores trucos matemáticos no sirvan para emular cómo los seres humanos construimos una representación racional del mundo no es, en sí, un problema matemático. Lo que quiere decir es que nosotros no conocemos nuestro mundo y nos familiarizamos con éste como lo hace un sistema computacional en donde usamos distintos métodos matemáticos para representar puntos de datos en un espacio geométrico-analítico. Yann LeCun, refiriéndose a este problema en el video ya citado, que en la literatura de ciencias de la computación se conoce como la paradoja de Hans Moravec, dice que la cantidad de información que puede procesar un ser humano es tan enorme que no tiene nada que ver con la estructura del lenguaje natural. El lenguaje ofrece poca información o ninguna por sí solo. Lo que es informativo no es la sintaxis del lenguaje, sino el tejido semántico del mundo, que no sabemos muy bien cómo lo construimos o lo tenemos. Éste es el gran tema del que nos ocupamos los filósofos profesionales, el del análisis y definición de conceptos básicos. Para usar un ejemplo de LeCun, mientras que cuesta mucho tiempo y esfuerzo programar un vehículo autónomo, que de todos modos habrá que estar ajustando todo el tiempo, un muchacho de 16 años puede aprender a manejar en un día. No sabemos muy bien por qué esto es así, ni cómo lo hace exactamente. Pero sea cual sea la explicación de este fenómeno, lo que es seguro es que nuestros modos de encarar y describir la realidad con funciones matemáticas, con toda la utilidad que han tenido y tienen para viajar a la luna y resolver problemas a través de sistemas de computación, no equivalen a lo que nosotros podemos hacer cuando comprendemos y analizamos nuestro mundo junto con otros interlocutores. Por eso es que esos sistemas no funcionan si no los estamos ajustando continuamente. Por esta razón, los Llms pueden generar contenidos, pero no conocimiento. El conocimiento sólo puede producirlo nuestros cerebros individuales y la razón humana, que, como bien sabía Kant, es una actividad que involucra necesariamente grupos humanos. A través de ella es que podemos idear los algoritmos y trucos

dan demostrar que $P=NP$, es decir, que es posible encontrar una función o algoritmo para los problemas que hoy se consideran intratables, incluso con la mayor computadora del mundo. Fuente: https://elpais.com/sociedad/2013/06/07/actualidad/1370616631_328551.html.

matemáticos que nos permiten optimizar una función, eludiendo en lo posible, con distintas técnicas y estrategias, los problemas de complejidad que acabo de esbozar.

La clave del desarrollo de un *verdadero* sistema de inteligencia artificial no es, pues, la optimización de una función matemática sobre un dominio dado, sino poder automatizar cómo representan los seres humanos el mundo. Este es un problema muy difícil porque no es estrictamente matemático, o de optimización, sino filosófico: el problema de toda la vida del análisis filosófico que atañe a cómo delimitamos un dominio de objeto con conceptos básicos. Es el problema del que se ocupa Aristóteles en su tratado sobre las categorías y aquello que llamaba Kant la “metafísica primera”.¹⁴ Tiene que ver con cómo construimos modelos sobre el mundo y los llenamos de información. Yann LeCun y otros están tratando, desde hace muchos años ya, de construir arquitecturas de software que de algún modo emulen nuestras maneras de comprender y orientarnos en el mundo dentro de los sistemas de representación en el aprendizaje automático. Pero que nadie lo haya conseguido hasta ahora lo evidencia las limitaciones de los actuales sistemas de IA.

Yo creo que esos esfuerzos son muy interesantes y valiosos, pero soy escéptica en relación con la posibilidad de modelar en espacios geométricos con funciones algebraicas nuestra comprensión básica del mundo de modo que un sistema de inteligencia artificial pueda reproducir nuestra capacidad de generalizar esa comprensión y actuar de forma autónoma. Después de la muerte de Kant en 1804 las prestigiosas escuelas de matemática de Königsberg y Göttingen realizaron durante todo el siglo XIX avances muy valiosos en la fundamentación matemática del conocimiento científico, con esfuerzos que desembocaron en nuestra comprensión de la ciencia actual y en los desarrollos en los fundamentos axiomáticos y geométricos de la electrodinámica, la teoría de la relatividad general y la mecánica cuántica. Con todo, como bien lo sabía Heidegger, siempre se producirán crisis en los fundamentos de la ciencia mientras no exista claridad sobre las bases ontológicas de nuestra comprensión del mundo.

14 Véase Kant, I., “Prefacio a la Segunda Edición” de la *Crítica de la razón pura* (P. Ribas, Trad.; 12.^a ed.), Alfaguara, 2017. (Obra original publicada en 1781).

De este modo, lo único que se puede hacer por los momentos es simular nuestros modos básicos de comprender la realidad y afinar todos los modelos de inteligencia artificial con enormes cantidades de informaciones producidas por seres humanos, es decir, con métodos de aprendizaje reforzado por retroalimentación humana.

5. Desafíos éticos a futuro de los sistemas de inteligencia artificial.

Esto abre una serie de nuevos desafíos éticos con los cuales la humanidad en general tendrá que enfrentarse ahora y en el futuro.

El primer desafío tiene que ver con la falsa impresión que las empresas que desarrollan esos modelos ofrecen, de acuerdo a la cual estamos a un paso de construir un sistema de inteligencia artificial con capacidades de comprensión y generalización similares a la humana. Por las razones que acabo de decir, esto es difícil, si no imposible, pero se entiende que corporaciones que invierten miles de millones de dólares en estos desarrollos tengan interés en hacernos creer esto. Todas las redes sociales están llenas de posts sobre posts con promesas de esta índole y esto no es casualidad. Se trata de una carrera por revalorizar continuamente el dinero invertido a través de la promoción masiva del uso y las narrativas sobre las inteligencias artificiales.

Como resultado de este engaño, muchas personas y muchos estudiantes tienen la impresión de que es posible que la IA sustituya el trabajo que tienen que hacer. Pero, en la medida en que los grandes modelos de lenguaje y razonamiento son parasitarios del conocimiento ya producido por seres humanos, es decir, en la medida en que sólo pueden crear contenidos sobre conocimiento ya producido, estas herramientas no los ayudan a aprender realmente. Lo que hacen es hacer accesible de manera más rápida una parte del acervo del conocimiento de la humanidad. En ese sentido, son muy útiles y yo también las uso todo el tiempo. Pero aprender de verdad requiere apropiarse del conocimiento y esto no se puede hacer consumiendo resúmenes un poco ineptos del trabajo que otros han hecho. Hacer una pregunta y recibir una respuesta requiere de parte de quien la hace un proceso de discernimiento y apropiación que apela a una serie de competencias previas que se han adquirido con esfuerzo y atención continuada. El

problema con estas herramientas de IA es que pudieran acentuar las desigualdades que existen ya entre los estudiantes, ayudando a los más inteligentes a estudiar más rápido y con más recursos y hundiendo a los más flojos en un océano de información que entienden cada vez menos y menos, hasta que salen fuera de los estándares mínimos que un estudiante de su nivel debería manejar. Yo lo veo todo el tiempo en mis cursos.

Un segundo desafío ético, relacionado con el primero, es el daño que una continua interacción con herramientas digitales hace a nuestra capacidad de atención y concentración, que es un resultado del estado actual de la difusión electrónica de la información. Porque información no es lo mismo que comunicación. Sin atención ni concentración adecuadas a lo que otros nos comunican no es posible aprender los difíciles tópicos que caracterizan los programas de estudio de la educación formal actual y yo creo que detrás de las recientes políticas en muchos países occidentales para prohibir el uso de las redes sociales a menores de 14 años se encuentra la absoluta alarma que debe estar produciendo la emergencia de una nueva generación de jóvenes que no sólo no saben; no comprenden. Como predijo el filósofo italiano Umberto Eco, vamos hacia una nueva Edad Media, una época en donde un grupito sabía que la tierra era redonda y había calculado su diámetro, mientras se burlaba secretamente de las supersticiones de los demás.

El tercer desafío ético tiene que ver con la toxicidad y la irresponsabilidad frente al individuo y a la sociedad de los contenidos que los usuarios de las herramientas digitales consumen sin parar. Esa toxicidad ha sido no sólo tolerada sino promovida por las grandes corporaciones digitales, que vieron crecer su base de clientes con la creación de ejércitos de bots comprometidos por una parcialidad política. Para entrenar estos modelos se necesita mucha gente generando contenido y, para ello, todas estas corporaciones dependen de bots y ejércitos de usuarios pagados para calentar la discusión. El clima político en todas las democracias del mundo, incluyendo la nuestra, se ha visto enrarecido por la presencia de bots pagados por los actores políticos, que, de manera hasta ahora impune, contratan ejércitos de activistas online para que intoxiquen el diálogo político que debería informar la

construcción de una voluntad política común, destruyendo la paz y la convivencia ciudadana en un país.

La difusión de contenido tóxico en redes sociales y, en general, en las herramientas que interactúan con nosotros como si fueran seres humanos, produce un daño profundo en comunidades y en personas vulnerables, que se ven invadidos por contenidos que no han sido generados por su grupo familiar o su comunidad, enajenando su capacidad para interpretarlos de un modo personalizado y humano. Sin duda alguna detrás de esto se encuentra lo que San Ignacio de Loyola llamaba el “enemigo de natura humana”. También los “espíritus del aire” que mencionaba Evagrio Pónico en sus escritos han sido objetivados de algún modo en todas esas herramientas, volviendo accesibles tentaciones que pueden realmente destruir las vidas y el carácter de miles de millones de personas, al ahogar las inspiraciones del Espíritu Santo en un mar de mensajes despersonalizadores e inhumanos.

No obstante, sin duda alguna también, los seres humanos pueden adquirir y difundir conocimientos muy valiosos a través de esas herramientas. Gracias a ellas, el acervo de conocimiento de la humanidad también se ha enriquecido y extendido a ámbitos a donde no se podía llegar sin egresar antes de una universidad. De este modo, lo que debemos hacer es, me parece, apropiarnos de esas herramientas dentro de un marco legal adecuado y dentro de una comprensión robusta de la educación formal real a la que queremos que nuestros jóvenes y las generaciones futuras se integren, enseñando no sólo su uso adecuado y responsable, sino explicando en términos conceptualmente claros, como he intentado hacerlo aquí, qué es lo que pueden hacer y lo que no.

Luz Marina Barreto
Universidad Central de Venezuela
luzmbarreto@gmail.com