# Revista Aragonesa de Teología

Centro Regional de Estudios Teológicos de Aragón

Universidad Pontificia de Salamanca

Año XXXI – Nº 62 – 2025

**EDITA**

**C.R.E.T.A.**

Centro Regional de Estudios Teológicos de Aragón

**Dirección**

Manuel Fandos Igado

**Subdirección**

Armando Cester Martínez y Bernardino Lumbreras Artigas



**Administración**

C.R.E.T.A

Ronda Hispanidad, 10. 5009. Zaragoza

**Impresión**

COPY CENTER DIGITAL

**ISSN:** 1135-0547

**Depósito Legal:** z-169/95

# Índice de contenidos

EDITORIAL: Entre la fascinación y el discernimiento: nuestra revista ante la inteligencia artificial *(Manuel Fandos Igado)*..... 5

- La no inteligente inteligencia artificial *(Francisco José Génova Omedes)*.....11

- La fe cristiana en la era digital: discernimiento teológico ante los desafíos y oportunidades de la cultura tecnológica *(Manuel Fandos Igado - auxiliado por IA generativa)*..... 39

- Un nuevo areópago en la palma de tu mano *(Pablo Vadillo Costa)*.....57

- Ecología integral y teología de la creación: una relectura de *Laudato Si'* y *Laudate Deum (Manuel Fandos Igado - auxiliado por IA generativa)*..... 79

- Comentario sobre el artículo: IA (2025). "Ecología integral y teología de la creación: una relectura de *Laudato Si'* y *Laudate Deum*" *(David Radoslaw Wroblewski)*..... 95

- La vocación y misión de los laicos en la Iglesia: de *Lumen Gentium* a la sinodalidad contemporánea *(Manuel Fandos Igado - auxiliado por IA generativa)*..... 103

- Análisis crítico sobre el artículo generado por inteligencia artificial (IA): "La vocación y misión de los laicos en la Iglesia, de *Lumen Gentium* a la sinodalidad contemporánea" *(Armando Cester Martínez)*.....121

ISSN: 1135-0547
Año XXXI - Nº 62 - 2025 / julio-diciembre

• La sinodalidad como forma constitutiva de la Iglesia: fundamentos teológicos y desafíos pastorales *(Manuel Fandos Igado - auxiliado por IA generativa)* .....135

• Comentario sobre el artículo: "La sinodalidad como forma constitutiva de la Iglesia: fundamentos teológicos y desafíos pastorales" *(María Dolores Ros de la Iglesia)* .....153

• Teología de la esperanza en tiempos de incertidumbre: una lectura escatológica y pastoral *(Manuel Fandos Igado - auxiliado por IA generativa)* .....161

• Comentario sobre el artículo: "Teología de la esperanza en tiempos de incertidumbre: una lectura escatológica y pastoral". ¿La inteligencia artificial es un lugar de esperanza? *(Galo Pedro Oria de Rueda Molins)* .....177

• El diálogo interreligioso en el siglo XXI: fundamentos teológicos, retos y horizontes *(Manuel Fandos Igado - auxiliado por IA generativa)* ..... 183

• La ética cristiana ante los desafíos de la biotecnología y la inteligencia artificial: discernimiento antropológico y responsabilidad moral *(Manuel Fandos Igado - auxiliado por IA generativa)* ..... 201

• Comentario sobre los artículos: "El diálogo interreligioso en el siglo XXI: fundamentos teológicos, retos y horizontes" - "La ética cristiana ante los desafíos de la biotecnología y la inteligencia artificial: discernimiento antropológico y responsabilidad moral". La investigación teológica y los grandes modelos del lenguaje (LLMs) *(Francisco José Génova Omedes)*.....219

ISSN: 1135-0547
Año XXXI - Nº 62 - 2025 / julio-diciembre

Revista Aragonesa de Teología
Centro Regional de Estudios Teológicos de Aragón

# La no inteligente inteligencia artificial

## The non-intelligent artificial intelligence

Francisco José Génova Omedes
Centro Regional de Estudios Teológicos de Aragón
francisco.genova@cretateologia.es
*https://orcid.org/0000-0001-5760-3337*

## Resumen

Frente al mito de la inteligencia artificial (IA) que hace de ella lo que no es, inteligente, y las promesas de un futuro mejor que nos distraen de la realidad del presente y de la responsabilidad de quienes lo están construyendo; es necesario proyectar una mirada a lo que verdaderamente es la IA y de un modo especial en nuestros días a los grandes modelos de lenguaje. Es necesario, igualmente, no caer en el autoengaño de los lenguajes que nos puede hacer creer que lo que se encuentra en el corazón de las diferentes tecnologías etiquetadas popularmente como IA son más de lo que realmente son. Ni el hardware de los ordenadores es un cerebro, ni el software una mente. Las llamadas redes neuronales ni piensan ni están formadas por neuronas. Y tras esas tecnologías hay un coste humano y ambiental que permanece en la sombra y que es necesario poner sobre la mesa. Si no, corremos el riesgo de caer atrapados en reducir la cuestión ética de esas tecnologías en una cuestión de uso, cuando hay una cuestión previa, la de la ética de su misma concepción y desarrollo.

**Palabras clave:** Grandes modelos de lenguaje, ética de la inteligencia artificial, inteligencia artificial, redes neuronales, sociedad digital.

## Abstract

Faced with the myth of artificial intelligence (AI), which makes it appear intelligent, and with promises of a better future that distract us from the reality of the present and from the responsibility of those building it, it is necessary to examine what AI truly is, especially today, the major language models. It is equally necessary to avoid the self-deception that language can lead us to believe that what lies at the heart of the various technologies popularly labeled as AI is more than it actually is. Neither computer hardware is a

brain, nor software a mind. So-called neural networks neither think nor are they composed of neurons. Behind these technologies lies a human and environmental cost that remains hidden and must be brought to light. Otherwise, we risk reducing the ethical question of these technologies to a matter of use, when there is a prior question: the ethics of their very conception and development.

**Keywords:** Artificial Intelligence, Artificial Intelligence Ethics, Digital Society, Neural Networks, Large Language Models.

## Introducción

En 2017 la investigadora en el campo de la inteligencia artificial (IA) Timnit Gebru, fundadora del grupo sin ánimo de lucro *Black in AI*, tras una de las conferencias promovidas por el grupo, fue abordada por Jeff Dean y Samy Bengin quienes le pidieron que contemplara la posibilidad de unirse al equipo de Google. Algo que haría en 2018 incorporándose como codirectora, junto a Jeff Dean, del *Google's ethical AI team*[1]. Su paso por Google no fue fácil, desde el primer momento manifestó su preocupación por las implicaciones éticas de los grandes modelos de lenguaje (LLMs, *Large Language Models*). Aunque sin encontrar respuesta entre sus colegas. Tras contactar en 2020 con Emily M. Bender, profesora de computación lingüística en la Universidad de Washington, acabaría siendo coautora junto a Bender y otros dos autores más, del artículo: "On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?"[2]. En este artículo Bender y Gebru, junto al resto de autores, no dudan en denominar a los grandes modelos de lenguaje como "loros estocásticos"[3]. Como consecuencia del artículo Gebru fue despedida de Google. Este momento es considerado por Karen Hao como un símbolo de los desafíos que presentan las grandes compañías dedicadas al desarrollo de la IA[4]. Y el artículo en cuestión es una llamada de atención que plantea dos

---

1 Karen Hao, *Empire of AI. Inside the reckless race for total domination* (London: Allen Lane, 2025), 160-174.

2 Emily M. Bender, Timnit Gebru, Angelina McMillan-Major y Shmargaret Shmitchell, "On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?", *FAccT'21: Proceedings of the 2021 ACm Conference on Fairness, Accountability, and Transparency* (March 2021): 610-623.

3 Bender, Gebru, McMillan-Major y Shmitchell, "Stochastic Parrots", 616-618. Gustavo La Fontaine resume así el concepto de loros estocásticos: "El 'loro estocástico' es una metáfora utilizada para describir la naturaleza de los sistemas de inteligencia artificial (IA), especialmente los modelos grandes de lenguaje escala como GPT-4 (Bender et al., 2021). Estos sistemas generan respuestas basándose en la estadística y la probabilidad, modelando cómo las secuencias de palabras tienden a organizarse en el lenguaje humano. Al entrenarse en extensos corpus de texto, aprenden a predecir la siguiente palabra en una secuencia, logrando producir respuestas que superficialmente parecen coherentes y contextualmente apropiadas" (Gustavo La Fontaine, "Sobre loros estocásticos. Una mirada a los modelos grandes de lenguaje", *LÓGOI Revista de Filosofía*, no. 45 (enero-junio 2024): 75-87, 77.

4 Hao, *Empire of AI*, 170.

preguntas fundamentales: qué clase de futuro estamos construyendo con la IA y para quién[5].

Esta historia inicial nos sitúa en el corazón del problema al tratar el tema de la IA: no es inteligente. La denominación IA es en realidad una etiqueta bajo la que se encuentran diferentes tecnologías[6]. Acuñada en 1956, en un encuentro que tuvo lugar en la Universidad de Dartmouth, de la mano de John McCarthy[7]. Una etiqueta que permitía transmitir unidad siendo a la vez una poderosa herramienta de márquetin para la captación de fondos. Una etiqueta rodeada de un halo mítico que ya desde el principio ha contado con la respuesta de quienes no compartían las pretensiones nacidas de este campo. Aunque este artículo se centre más específicamente en los LLMs, es conveniente comenzar con una mirada hacia las críticas clásicas a las pretensiones de inteligencia de la IA, que bien podrían describirse como el relato de una nueva alquimia.

## La nueva alquimia

Cuando el filósofo Hubert Dreyfus comenzó a dar clase en el *Massachusetts Institute of Technology* (MIT) no sospechaba cómo su carrera académica se

---

5 "What kind of future are we building with AI? By and for whom?" (Hao, *Empire of AI*, 170). Traducción al castellano: "¿Qué clase de futuro estamos construyendo con la IA? ¿Por y para quién?".

6 Luc Julia, *IA génératives, pas creatives* (Paris: Le Cherche Midi, 2025), 8. Emily M. Bender and Alex Hanna, *The AI Con. How to Fight Big Tech's Hype and Create the Future We Want* (New York: Harper, 2025), 5-8. Bender y Hanna lo afirman con rotundidad: "To put it bluntly, "AI" is a marketing term. It doesn't refer to a coherent set of technologies" (Bender and Hanna, *The AI Con*, 5). Estos mismos autores consideran clarificador sustituir IA por la palabra automatización, para a continuación preguntarnos qué está siendo automatizado, la respuesta nos acerca a las diferentes tecnologías de se engloban bajo la etiqueta de IA: Tomar decisiones, clasificar, recomendar, traducir y generar textos e imágenes (Bender and Hanna, *The AI Con*, 6-7).

7 Pamela McCorduck, *Machines Who Think* (Natick-Massachusetts: A K Peters, 2004), 111-136. David Levy, *Robots Unlimited. Life in a Virtual Age* (Wellesley-Massachusetts: A K Peters, 2006), 64-68. Daniel Crevier, *AI. The Tumultuous History of the Search for Artificial Intelligence* (New York: Basic Books, 1993), 48-50. Nils J. Nilson, *The Quest for Artificial Intelligence. A History of Ideas and Achievement* (Cambridge: Cambridge University Press, 2010), 52-56.

iba a ver marcada por lo que en los laboratorios de IA del MIT se desarrollaba, aún más si tenemos en cuenta que hasta ese momento ni siquiera había visto en toda su vida un ordenador. Al conocer lo que los investigadores del MIT en IA planteaban, no pudo sino ser consciente de las contradicciones con lo que él mismo enseñaba en sus clases de filosofía. Él no duda en hablar del carácter no-científico del campo de la IA y lo expresa así: "Quienes trabajan en el campo de la inteligencia artificial, a diferencia de la mayoría de los científicos, casi nunca reconocen sus dificultades y son muy sensibles a las críticas."[8] Sería sin embargo un encargo de la Corporación RAND el que le acabaría convirtiendo en un referente de las críticas filosóficas a la IA. El encargo consistía en evaluar el trabajo que en el campo de la simulación cognitiva realizaban investigadores de la corporación. Las conclusiones a las que llegó Dreyfus eran tan negativas que la corporación retraso un año su publicación, la cual finalmente se produjo en diciembre de 1965. El título que eligió Dreyfus para publicar su estudio era ya toda una declaración: *Alchemy and Artificial Intelligence*[9]. De este modo, para él la empresa de la IA es comparable a la de los antiguos alquimistas, siendo su principal objeción el que se haya realizado lo que él denomina la "asunción epistemológica", esto es, partir de que todo el conocimiento humano es reducible a reglas formales.

> Es extremadamente difícil definir lo que es el nivel mental de funcionamiento, y eso es así sea lo que sea la mente, siendo de ningún modo obvio que la mente funcione como una computadora digital. Esto hace prácticamente ininteligibles las afirmaciones de quienes trabajan en Simulación Cognitiva de que la mente puede entenderse como un sistema que procesa información siguiendo unas reglas heurísticas. Tomar el ordenador como modelo resulta insuficiente para explicar qué hacen realmente las personas cuando piensan y perciben, y, a la inversa, el hecho de que las personas piensen y perciban no puede infundir opti-

[8] Hubert L. Dreyfus and Stuart E. Dreyfus, *Mind Over Machine. The Power of Human Intuition and Expertise in the Era of the Computer* (New York: The Free Press, 1986), 8. Original: "Workers in artificial intelligence, unlike most scientist, almost never acknowledge their difficulties and are highly sensitive to criticism". Las citas de bibliografía en otros idiomas se presentarán traducidas en el cuerpo del texto, indicando a pie de página la redacción en la lengua original. Las citas a pie de página no se traducirán.

[9] Hubert L. Dreyfus, *Alchemy and Artificial Intelligence* (Santa Mónica-CA: RAND Corporation, 1965). *http://www.rand.org/content/dam/rand/pubs/papers/2006/P3244.pdf [17/10/2025].*

> mismo a quienes intentan reproducir el funcionamiento humano con computadoras digitales.[10]

Otra importante crítica llega desde el filósofo John Searle, él plantea que lo que se limitan a hacer los ordenadores es manipular símbolos mientras que la mente humana les atribuye significado. Aunque Searle no se opone a la idea de que se pueda reproducir artificialmente el cerebro, se opone a la idea de que la computación por sí misma sea el camino para lograrlo[11]. Searle elaboró el argumento conocido como la habitación china, *Chinese Room Argument*, en 1980[12]: Se parte de que una persona que desconoce el idioma chino se encuentra encerrada en una habitación. En esa habitación la persona tiene a su disposición un libro de instrucciones en su idioma que le da todas las reglas para manipular unos símbolos, correspondientes al idioma chino. De este modo, cuando le entreguen un texto escrito en chino, la persona podrá, siguiendo las instrucciones que posee, dar una respuesta en chino que él mismo no entiende. La relevancia del artículo se refleja tanto en quienes lo critican como entre quienes lo apoyan[13]. Y esta ha llegado hasta

---

[10] Hubert L. Dreyfus, *What Computers Still Can't Do* (Cambridge: The MIT Press, 1994), 189. Original: "It is extremely difficult to define what the mental level of functioning is, and that whatever the mind is, it is by no means obvious that it functions like a digital computer. This makes practically unintelligible the claims of those working in Cognitive Simulation that the mind can be understood as processing information according to heuristic rules. The computer model turns out not to be helpful in explaining what people actually do when they think and perceive, and, conversely, the fact that people do think and perceive can provide no grounds for optimism for those trying to reproduce human performance with digital computers."

[11] "We know, from the Chinese Room Argument, that computation by itself is insufficient to guarantee any such causal powers, because computation is defined entirely in terms of manipulation of abstract formal symbols" (John Searle, "I Married a Computer" in *Are We Spiritual Machines? Ray Kurzweil vs. The Critics of Strong AI*, ed. Jay Richards (Seattle: Discovery Institute, 2002), 67).

[12] John R. Searle, "Minds, brains, and programs", *Behavioral and Brain Sciences* 3, no. 3 (1980): 417-457.

[13] Un claro oponente de este argumento es Ray Kurweil, el cual le dedica un artículo titulado "Locked in His Chinese Room" para rebatirlo, con el argumento principal de que Searle no entiende el funcionamiento del cerebro, que se resume en esta frase de Kurzweil "*I understand English, but none of my neurons do*" (Ray Kurzweil, "Locked in His Chinese Room" in *Are We Spiritual Machines? Ray Kurzweil vs. The Critics of Strong AI*, ed. Jay Richards (Seattle: Discovery Institute, 2002), 131). Ver también: Ray Kurzweil, *The Singularity is Near* (New York:

la actualidad, de modo que seguimos encontrando referencias a él en obras actuales, como es el caso de la propuesta de un humanismo digital por parte de Julian Nida-Rümelin y Nathalie Weidenfeld[14].

Entre los argumentos tradicionales, que niegan la posibilidad de que mediante sistemas computacionales se logre una IA equiparable o superior a la mente humana, nos encontramos también con un físico, Roger Penrose, para quien la mente humana no es reducible a algoritmos. Como físico sus objeciones no provienen de la filosofía sino de la ciencia. Y se fundamentan en la biología neuronal, las matemáticas y la mecánica cuántica[15]. Al igual que Searle, su crítica se centra en la pretensión de lograr ese objetivo desde una comprensión computacional de la mente humana, no al objetivo en sí[16]. La pregunta principal para Penrose, cuya respuesta une a la posibilidad de una verdadera IA que llegue a ser consciente, es: "¿Cómo es posible que una conciencia pueda surgir realmente a partir de un objeto material?"[17].

Finalmente citaremos a Joseph Weizenbaum. Este profesor de informática en el MIT desarrolló en 1964 el programa Eliza simulando la conversación con un terapeuta, que es considerado el primer bot conversacional. Este hecho iba a suponer un antes y un después en su vida. Al observar cómo en el MIT eran muchos los que caían seducidos por el programa y lo conside-

---

Penguin Books, 2006), 465-469. Los argumentos de Kurzweil parecen estar apoyados en lo de Paul y Patricia Churchland, que afirman: "It is irrelevant that no unit in his system understands Chinese, since the same is true of nervous systems: no neuron in my brain understands English, although my whole brain does. For another, Searle neglects to mention that his simulation (using one person per neuron, plus a fleet-footed child for each synaptic connection) will require at least $10^{14}$ people, since the human brain has $10^{11}$ neurons, each of which averages over $10^{3}$ connections. His system will require the entire human populations of over 10,000 earths" (Paul M. Churchland and Patricia Smith Churchland, "Could a Machine Think? Classical AI is unlikely to yield conscious", *Scientific American* (February 1990): 32-37).

14 Julian Nida-Rümelin und Nathalie Weidenfeld, *Digitaler Humanismus. Eine Ethik für das Zeitalter der Künstlichen Intelligenz* (München: Piper, 2018), 115-117.

15 Roger Penrose, *The Emperor's New Mind* (Oxford: Oxford University Press, 1989). Este libro es considerado como "the most powerful attack yet written on strong AI" (Penrose, *The Emperor's New Mind*, xii).

16 Penrose, *The Emperor's New Mind*, 537-538.

17 Penrose, *The Emperor's New Mind*, 523. La cursiva pertenece al original. Original: "How is that a material object (a brain) can actually *evoke* consciousness?".


DOI: 10.59853/RAT62-2025-0201
ISSN: 1135-0547
Año XXXI - Nº 62 - 2025 / julio-diciembre

raban inteligente tratándolo como a una persona. Esto le llevó a advertir de los peligros que para el ser humano podían representar estas tecnologías, porque él era plenamente consciente de que Eliza era solo un programa ejecutándose en un ordenador, nada más, frente a las proclamas en el propio MIT de colegas como Marvin Minsky[18]. En la obra que podemos considerar como síntesis de su pensamiento, *Computer Power and Human Reason*[19], Joseph Weizenbaum defiende que hay que mantener siempre una distinción clara entre el ser humano y las máquinas, y que hay tareas que deben ser reservadas a las personas aunque técnicamente puedan ser realizadas por una máquina. El mayor peligro para él es que los seres humanos abdiquen de sus responsabilidades en favor de las máquinas. Por eso denuncia ya en 1984 la idolatría al ordenador que él ha ido observando y cómo se transmite a los niños, manifestación que hace desde el pesimismo, pues cree que ese es el camino que se va a seguir manteniendo: "El adoctrinamiento de las mentes de nuestros niños con una idolatría informática simplista y uniforme, y eso es casi con certeza lo que la mayor parte del aprendizaje de la informática es y será, es un fenómeno pandémico"[20].

## El autoengaño motivado por los lenguajes de programación

Cualquiera que haya programado utilizando los llamados lenguajes de bajo nivel, aquellos que son los que actúan directamente sobre la unidad central de proceso, tiene una experiencia muy distinta de programación de la que se tiene cuando uno se mueve con lenguajes llamados de alto nivel, que resultan más manejables a nivel de programación pero que han de ser traducidos a lo que realmente "entiende" el hardware del ordenador: el código máquina. Para comprenderlo vamos a hacer una ligera aproximación a lo que ocurre en el corazón del hardware. En realidad, es todo mucho más sencillo de lo que parece, todas las operaciones que se realizan involucran exclusivamente dos estados, un sistema binario, que expresamos como uno (1) y cero (0).

---

18 Marvin Minsky, *The Society of Mind* (New York: Simon & Schuster, 1986).

19 Joseph Weizenbaum, *Computer Power and Human Reason. From Judgement to Calculation* (Hardmonsworth: Penguin Books, 1984).

20 Weizenbaum, *Computer Power*, xix. Original: "The indoctrination of our children's minds with simplistic and uniformed computer idolatry, and that is almost certainly what most of computer instruction is and will be, is a pandemic phenomenon".

Para darnos cuenta de la sencillez, pensemos en lo contentos que estarían los escolares que tienen que aprender la tabla de multiplicar cuando les dijésemos que esta es toda la tabla para memorizar:

| Tabla de multiplicar del sistema binario | |
|---|---|
| 0 x 0 | 0 |
| 0 x 1 | 0 |
| 1 x 0 | 0 |
| 1 x 1 | 1 |

Tabla 1

El resto de operaciones resultan así mismo sencillas. Eso sí, pronto se encontrarían con la necesidad de escribir mucho más cuando les dijésemos que, por ejemplo, para trabajar con el número noventa y tres, que en decimal representamos como 93, tienen que escribir en forma de un código binario 1001 0011[21]. Porque en el hardware de nuestros ordenadores todo se reduce a trabajar con unos y ceros, eso sí, a través de circuitos electrónicos que manejan magnitudes eléctricas. No hay que perder de vista esa perspectiva, porque el hardware no "entiende" de unos y ceros, sino de magnitudes eléctricas.

Nada hay en todo esto que recuerde a un cerebro humano. Porque la verdadera maravilla tecnológica no está en las operaciones que se realizan, recordemos que son operaciones con ceros y unos, sino en la velocidad y el nivel de integración de los circuitos, esto es, el tamaño y capacidad de estos.

De modo que el hardware de cualquier ordenador actúa paso a paso realizando operaciones con ceros y unos. Y las instrucciones que mediante el programa le acabamos dando son así mismo códigos de ceros y unos ante los cuales el circuito lógico da una respuesta según ha sido diseñado. Por ejemplo, una operación básica, que incluye una instrucción y un dato con el que operar, podría ser algo así:

[21] Esta representación del número noventa y tres no es en sistema binario sino en código binario, que como podemos ver *codifica* cada uno de los dígitos decimales, correspondiéndole en este caso al 9 el 1001 y al 3 el 0011. Por tanto, no estamos en una representación del sistema binario, que sería 1011101. De modo que estamos hablando realmente de un número decimal codificado en binario (BCD).


DOI: 10.59853/RAT62-2025-0201
ISSN: 1135-0547
Año XXXI - Nº 62 - 2025 / julio-diciembre

| Código | Dato |
|---|---|
| 1010 0101 | 0010 0001 |

Tabla 2

Los circuitos lógicos internos de la unidad central de proceso, al tener ese código de entrada, harían con el dato lo que han sido diseñados para hacer al recibir ese código. Ni más, ni menos. Sin que el lenguaje que utilizamos nos haga olvidar que en realidad se trata de magnitudes eléctricas.

Hay quienes sostienen que de todo esto puede surgir una conciencia porque esta surgirá de todo sistema que maneja símbolos. Así lo hace el filósofo norteamericano Douglas Hofstadter[22], que lo expresa así: "La conciencia es aquella propiedad de un sistema que surge siempre que existen símbolos que obedecen a patrones desencadenantes"[23]. Este autor es un claro e influyente ejemplo de cómo se toma al ordenador, creación humana, como referencia para comprender al ser humano, y en concreto su cerebro, mente y conciencia. No solo no duda en usar la palabra hardware para hablar del cerebro[24], sino que en camino inverso proyecta sobre los ordenadores las capacidades mentales del ser humano. Para él es bastante plausible que un ordenador pueda llegar a hablar sobre sí mismo del mismo modo que los seres humanos y que, según él, defienda erróneamente también, como los seres humanos, su libre albedrío[25]. Toda la elaboración de este autor no es sino el marco filosófico de las ideas de los científicos e ingenieros defensores de una IA general, que simultáneamente, como Marvin Minsky, a quien cita Hofstadter, reducen al ser humano a su creación, el ordenador. Para Minsky los cerebros no son más que máquinas, los seres humanos máquinas de carne, y la libertad no existe, es solo una ilusión[26].

---

[22] Douglas R. Hofstadter, *Gödel, Escher, Bach: an Eternal Golden Braid* (New York: Basic Books, 1979), 384-385.

[23] Hofstadter, *Gödel, Escher, Bach*, 385. Original: "consciousness is that property of a system that arises whenever there exits symbols which obey triggering patterns".

[24] "It was only with the advents of computers that people actually tried to create 'thinking machines' [...]. As a result, we have acquired, in the last twenty years or so, a new kind of perspective on what thought is and what it is not. Meanwhile, brain researchers have found out much about the small-scale hardware of the brain" (Hofstadter, *Gödel, Escher, Bach*, 337).

[25] Hofstadter, *Gödel, Escher, Bach*, 388.

[26] Marvin Minsky, *The Society of Mind* (New York: Simon & Schuster, 1986), 288, 306.

Pero el debate sobre la manipulación de símbolos por parte de la IA deja de lado una realidad más profunda, por mucho que el leguaje quiera equiparar al ser humano con la máquina, al cerebro con el ordenador, la realidad es que, al profundizar en las entrañas de cada uno de ellos, la engañosa similitud se diluye para dar paso a la realidad. Como ya hemos visto, lo que verdaderamente manejan cualquiera de las tecnologías agrupadas bajo la etiqueta de IA son magnitudes eléctricas aplicando un sistema binario, de unos y ceros, de una simplicidad que desmonta cualquier pretensión de parecido con los procesos biológicos en el ser humano. Porque ahí radica una diferencia fundamental: la vida. Las llamadas redes neuronales, sobre las que profundizaremos más adelante, no son sino fórmulas matemáticas aplicadas en un programa y cualquier argumentación que se abra a la posibilidad de conciencia carece de fundamento alguno[27]. A una conclusión similar llega el sociólogo alemán Armin Nassehi tras estudiar los fundamentos de la sociedad digital, para él: " Quizás es demasiado contraintuitivo encontrar la diferencia crucial entre el ser humano y la máquina no tanto en las operaciones cognitivas como en el medio que forma su sustrato"[28]. Esta diferencia de sustrato no es sino la diferencia entre la vida y la no-vida, Nassehi lo afirma con un "quizás": " Quizás es simplemente tener la vida como fundamentación lo que hace a las dos máquinas diferentes la una de la otra, no la inteligencia"[29]. Las palabras finales son la clave para una afirmación fundamental: todas las discusiones sobre la inteligencia de las máquinas y su comparación con la inteligencia humana pierden pie porque se olvidan de que la verdadera cuestión no es de inteligencia sino de vida.

En definitiva, ¿cuál es el autoengaño motivado por los lenguajes de alto nivel? El de confundir lo que es un medio que facilita la tarea de programación con la realidad misma del sistema que es programado. Si esto ocurre con lenguajes de programación no es de extrañar que cuando el lenguaje directo

---

[27] Bender y Hanna en su obra ya citada se manifiestan claramente en contra de estas pretensiones, afirmando: "claims of AI sentience or consciousness are a kind of AI hype, and not at all grounded in reality" (Bender and Hanna, *The AI Con*, 22).

[28] Armin Nassehi, *Muster. Theorie der digitalin Gesellschaft* (Bonn: C. H. Beck, 2019), 262. Original: "Vielleicht ist es allzu kontraintuitiv, die entscheidende Differenz zwischen Mensch und Maschine weniger in den kognitiven Operationen selbst, sondern in ihrem medialen Substrat aufzufinden".

[29] Nassehi, *Muster*, 262. Traducción: "Vielleicht ist es schlicht das Leben las Grundlage, das die beiden Maschinen unterscheidet, nicht die Intelligenz".


DOI: 10.59853/RAT62-2025-0201
ISSN: 1135-0547
Año XXXI - Nº 62 - 2025 / julio-diciembre

de comunicación con el ordenador es el lenguaje natural, como ocurre en los LLMs, el autoengaño del usuario, fruto de la antropomorfización, lleve a creer que hay algún tipo de "mente" tras las respuestas que se generan ante los *prompts* que introducimos en ellos. Hay que hacer notar el efecto que tiene para un ser humano comunicarse en lenguaje natural con una máquina, en este sentido Mark Coeckelberg y David J. Gunkel consideran que tras el derrocamiento de la posición central que ha sufrido el ser humano con las revoluciones copernicana, darwiniana y freudiana[30], la revolución digital ha sustraído al ser humano lo que aún le quedaba como último dominio exclusivo: el lenguaje. Por este motivo nos vamos a centrar a continuación en la explicación de los LLMs.

## Las redes neuronales y los grandes modelos de lenguaje (LLMs)

### *Redes neuronales sin neuronas*

El origen de los LLMs debemos buscarlo en el procesamiento de lenguaje natural (NLP, *Natural Language Processing*). Y para hablar de NLP debemos retroceder hasta Alan Turing. Este, en el mismo artículo que comienza lanzando la pregunta "¿Pueden pensar las máquinas?"[31], termina preguntándose qué caminos hay que emprender para alcanzar la realización de una máquina inteligente. Y se responde al final del artículo marcando dos caminos:

> Podemos esperar que las máquinas eventualmente compitan con los hombres en todos los campos puramente intelectuales. Pero ¿cuáles son los mejores para empezar? Incluso esta es una decisión difícil. Mucha gente piensa que una actividad muy abstracta, como jugar al ajedrez, sería lo mejor. También se puede argumentar que lo mejor es dotar a la máquina de los mejores órganos sensoriales posibles y luego enseñarle a entender y hablar inglés. Este proceso podría seguir la enseñanza normal de un niño: se señalarían y nombrarían las cosas, etc. Repito

[30] Mark Coeckelberg and David J. Gunkel, *Comunicative AI. A Critical Introduction to Large Language Models* (Cambridge: Polity Press, 2025), 2.

[31] Alan Turing, "Computing Machinery and Intelligence," *Mind* 59, no. 236 (October 1950): 433. Original: "Can machines think?".

> de nuevo que no sé cuál es la respuesta correcta, pero creo que ambos enfoques deberían intentarse.[32]

Turing abre dos caminos, el de la inteligencia como una actividad abstracta que, según él, tiene su reflejo jugando al ajedrez, y el del lenguaje. El primer camino supuso una apuesta por la llamada aproximación simbólica, la de una inteligencia concebida como puro razonamiento matemático que en el ajedrez tenía el objetivo de vencer a un campeón humano. Esto ocurrió en 1996 cuando el ordenador Deep Blue de IBM venció al campeón mundial de ajedrez Gary Casparov. Sin embargo, el entusiasmo se disipó pronto, al mismo tiempo que Deep Blue era desmontado tras el juego, quedó claro que el resultado no era lo que muchos habían esperado. Al fin y al cabo, lo que Deep Blue hacía no era pensar, sino poner en práctica una gran velocidad de cálculo para comprobar todas las posibles jugadas y sus desarrollos. Nada había ahí que recordase al pensamiento humano. El segundo camino, el del NLP[33], es el de conseguir que los ordenadores se puedan comunicar con nosotros en lenguaje natural[34]. La primera dificultad que encontramos para conseguirlo es que los ordenadores no "entienden" el lenguaje, recordemos lo que comentábamos sobre cómo operan con magnitudes eléctricas implementando operaciones lógicas con unos y ceros. Esto supone que hay que buscar cómo unir la realidad de lo que es un ordenador con la realidad de lo que es el lenguaje humano. Y ese punto de unión van a ser algoritmos que "traduzcan" el lenguaje humano en información numérica que será manejada por el ordenador. La primera aproximación al problema fue la de la IA simbólica, también conocida como "Good-Old-Fashioned AI" (GOFAI),

---

32 Turing, "Computing Machinery", 460. Original: "We may hope that machines will eventually compete with men in all purely intellectual fields. But which are the best ones to start with? Even this is a difficult decision. Many people think that a very abstract activity, like the playing of chess, would be best. It can also be maintained that it is best to provide the machine with the best sense organs that money can buy, and then teach it to understand and speak English. This process could follow the normal teaching of a child. Things would be pointed out and named, etc. Again I do not know what the right answer is, but I think both approaches should be tried".

33 David Levy, *Robots Unlimited. Life in a Virtual Age* (Wellesley–Massachusetts: A K Peters, 2006), 245-266. David Levy define así el NLP: "Natural Language Processing (NLP) is the branch of Artificial Intelligence concerned with enabling computers to talk like you and me, to understand what is to said to them, to be able to conduct sensible conversations and even to translate into and out foreign languages" (Levy, *Robots Unlimited*, 245).

34 Coeckelberg and Gunkel, *Comunicative AI*, 13.


DOI: 10.59853/RAT62-2025-0201
ISSN: 1135-0547
Año XXXI - Nº 62 - 2025 / julio-diciembre

consistía en disponer de una base de datos, palabras, etiquetados adecuadamente y disponer en el programa de ordenador de las reglas necesarias para agruparlas de acuerdo con el lenguaje utilizado[35]. Esto permitió el desarrollo ya desde los años cuarenta del siglo XX de los modelos de lenguaje conocidos como "n-gram"[36]. En este modelo las palabras son clasificadas con un peso en función de su frecuencia de aparición en un conjunto de n palabras sobre las que se realiza la predicción en base a cálculos que ejecuta el programa. Los modelos más simples son los que se encontraban detrás de la escritura de texto letra a letra en el teclado numérico de los antiguos teléfonos móviles[37]. Estos sistemas predictivos sobre *n* palabras para aumentar su precisión debían incrementar *n*, pero hacerlo siguiendo el sistema clásico de programación generaba problemas que lo acababa haciendo inviable. Otra aproximación que hasta ese momento no había tenido grandes éxitos, aunque ya estaba presente desde los años cincuenta del siglo XX, va a emprender un camino renovado a partir de 2010; el de lo que se ha conocido como modelos neuronales del lenguaje[38], una aplicación de las denominadas redes neuronales[39]. Dos hechos facilitaron ese éxito, la reducción significativa del coste del hardware y la amplia disposición en la web de gran número de datos con los que entrenar esos modelos. Esto impulsaría el paso a los LLMs y su desarrollo actual. En la base de las redes neuronales se encuentra la pretendida simulación de un modelo de neurona, que no es sino un tipo de algoritmo en el que se ejecutan funciones matemáticas, conocido como perceptrón[40].

---

35 Coeckelberg and Gunkel, *Comunicative AI*, 14-16.

36 En la expresión *n-gram* expresa el número de palabras (n) sobre las que se hace la predicción.

37 Bender and Hanna, *The AI Con*, 23-28.

38 Bender and Hanna, *The AI Con*, 25.

39 Daniel Jurafsky y James H. Martin las definen así: "A neural network is a set of small computation units connected by weighted links. The network is given a vector of input values and computes a vector of output values. The computation proceeds by each computational unit computing some non-linear function of its input units and passing the resulting value on to its output units" (Daniel Jurafsky and James H. Martin, *Speech and Language Processing. An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition* (Upper saddle River – New Jersey: Prentice-Hall, 2000), 268-269.

40 Tzolkin Garduño Alvarado, Feliú Sagols Trancoso y Gunnar Wolf, "Las neuronas artificiales y su papel central en la inteligencia artificial", *Ciencia* 76, no. 1 (enero-marzo 2025): 70-71.

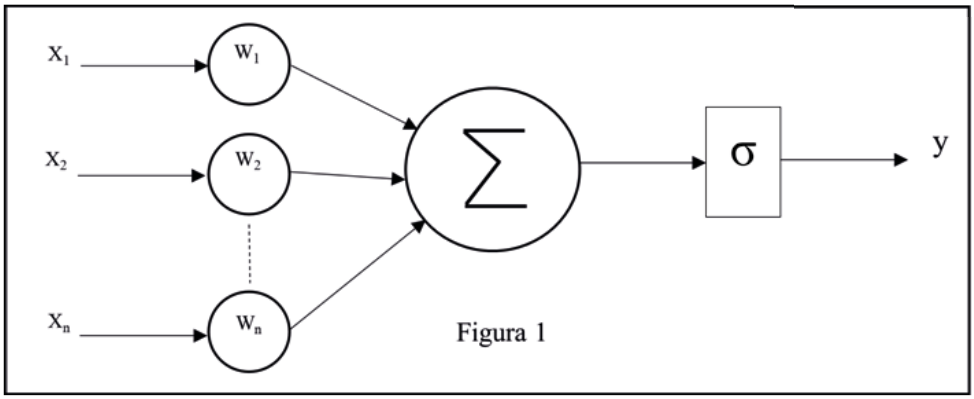


Figura 1

$$y = \sigma\left(\sum_{i=1}^{n} w_i\, x_i\right)$$

Como podemos ver en la figura 1, el nombre de neurona o red neuronal puede llevarnos al autoengaño de asimilarlas con las verdaderas neuronas biológicas. Aquí nos encontramos ante unas fórmulas matemáticas operadas en un ordenador. Para Bender y Hanna estamos ante la perpetuación de una falsa equivalencia entre el cerebro humano y lo que no deja de ser una máquina de cálculo de gran capacidad[41], que tiene su punto de origen en los perceptrones, que se encuentran en su base y a su vez se inspiraron vagamente, para su desarrollo en los años cuarenta del siglo XX, en el funcionamiento de las neuronas[42]. De modo que los modelos neuronales del lenguaje, y en general las redes neuronales, ni contienen neuronas ni su funcionamiento es equiparable al del cerebro. Julian Nida-Rümelin y Nathalie Weidenfeld lo afirman con rotundidad:

> El término 'redes neuronales' es engañoso en dos aspectos. En primer lugar, estas redes no constan de neuronas, sino de unidades transmisoras, y, en segundo lugar, estas redes solo se asemejan, en el mejor de los casos, a la inmensa complejidad y plasticidad del cerebro humano[43].

41 Bender and Hanna, *The AI Con*, 13.

42 Bender and Hanna, *The AI Con*, 25.

43 Nida-Rümelin und Weidenfeld, *Digitaler Humanismus*, 113. Original: "Der Ausdruck, neuronale Netze' ist dabai in dopelter Hinsicht irreführend. Erstens bestehen diese Netze nicht aus Neuronen, sondern aus Leiterstrukturen, und zweitens ähneln diese Netze allenfalls sehr


DOI: 10.59853/RAT62-2025-0201
ISSN: 1135-0547
Año XXXI - Nº 62 - 2025 / julio-diciembre

Desde ahí podemos fundamentar una afirmación central: la IA no piensa. Los mismos autores que acabamos de comentar dedican en su propuesta de humanismo digital un capítulo a afirmar que las llamadas inteligencias artificiales no se pueden equiparar en ningún caso al pensamiento humano[44]: "Los ordenadores y los humanos no piensan de la misma manera. O, para ser más precisos: un ordenador no piensa en absoluto como nosotros"[45]. Para ellos toda la realidad de la IA en su relación con el pensamiento humano se reduce a la categoría de simulación, porque los sistemas de software funcionan de acuerdo con un algoritmo mientras que los seres humanos no[46]. Una filosofía realista conduce a una interpretación mecanicista del ordenador que queda reducido a lo que es, una secuencia de estados eléctricos en una máquina.[47] Es lo que Joseph Weizenbaum expresaba al explicar su programa Eliza:

Pero una vez que un programa en particular es desenmascarado, una vez que su funcionamiento interno es explicado en un lenguaje suficientemente claro como para inducir la comprensión, su magia se desvanece; se revela como una mera colección de procedimientos, cada uno perfectamente comprensible. El observador se dice a sí mismo: "Yo podría haber escrito eso". Con ese pensamiento traslada el programa en cuestión del estante marcado como "inteligente", a aquel reservado para curiosidades, listo para ser discutido solo con personas menos ilustradas que él.[48]

---

entferntder immensen Komplexität un Plastizität des menschlichen Gehirns".

44 "Warum KIs nicht denken können" es como titulan un capítulo de su libro dedicado a la defensa de un humanismo digital: Julian Nida-Rümelin und Nathalie Weidenfeld, *Digitaler Humanismus. Eine Ethik für das Zeitalter der Künstlichen Intelligenz* (München: Piper, 2018), 108-119.

45 Nida-Rümelin und Weidenfeld, *Digitaler Humanismus*, 109. Original: "Computer und Menschen denken nicht auf die gleige Weise. Oder, um es zu präzisieren: Ein Computer denkt in unserem Sinne überhaupt nicht".

46 "Roboter und Softwaresysteme funktionieren nach einem Algorithmus, Menschen night. Darin liegt einer ihrer zentralen Unterschiede begründet" (Nida-Rümelin und Weidenfeld, *Digitaler Humanismus*, 111). Traducción: "Los robots y los sistemas de software funcionan de acuerdo con un algoritmo, los seres humanos no. Esta es una de las principales diferencias".

47 Nida-Rümelin und Weidenfeld, *Digitaler Humanismus*, 117-118.

48 Joseph Weizenbaum, "ELIZA–A Computer Program For the Study of Natural Language. Communication Between Man And Machine", *Communications of the ACM* 9, no. 1 (January 1966): 36-45, https://dl.acm.org/doi/10.1145/365153.365168. Original: "But once a particular

## *Los grandes modelos de lenguaje*

El modo de funcionamiento de los grandes modelos de lenguaje es el desarrollo de una representación probabilística de una lengua en el que se va prediciendo la palabra que viene a continuación[49], dadas unas determinadas palabras anteriores a través de una ponderación que da más peso a aquellas que tienen más probabilidad de ser la siguiente según los datos con los que ha sido entrenado el modelo. Al aplicar las redes neuronales se evita el problema de la programación clásica que requeriría un tamaño inmanejable. De este modo, convirtiendo las palabras en vectores los LLMs, partiendo del *prompt* podrán hacer todo el proceso necesario para a su vez generar una nueva secuencia de palabras en base a probabilidad y estadística. Por eso Coeckelberg y Gunkel pueden hacer una afirmación que es fundamental para nuestro propósito: "Dado que el LLM simplemente une palabras siguiendo probabilidades estadísticas, no conoce este hecho, al menos no en ningún sentido de "conocer" tal como es aplicado a los seres humanos"[50]. Ese unir simplemente palabras siguiendo probabilidades estadísticas nos lleva al centro del artículo: la no inteligencia de estos sistemas, aunque sean conocidos bajo la etiqueta de IA. Sin embargo, lo que hemos denominado el autoengaño de la antropomorfización nos va a llevar a la facilidad con la que podemos acabar atrapados en la ilusión de estar "hablando" con un "alguien" tras ese texto generado. Este riesgo, que hemos visto fue ya detectado en 1964 por Joseph Weizenbaum al desarrollar el programa Eliza, ha sido objeto de estudio dentro del campo de las interacciones entre seres humanos y ordenadores (*Human Computer Interaction,* HCI). Un campo que recibió el impulso en 1996 de Byron Reeves y Clifford Nass cuando en su estudio plantearon cómo el considerar persona a aquello que interactúa con nosotros, y tratarlo como tal, es un proceso automático fruto de nuestro desarrollo evolutivo, porque para ellos todo nuestro desarrollo tecnológico no puede obviar el cómo fun-

program is unmasked, once its inner workings are explained in language sufficiently plain to induce understanding, its magic crumbles away; it stands revealed as a mere collection of procedures, each quite comprehensible. The observer says to himself 'I could have written that'. With that thought he moves the program in question from the shelf marked 'intelligent', to that reserved for curios, fit to be discussed only with people less enlightened than he".

49 Coeckelberg and Gunkel, *Comunicative AI*, 15-16.

50 Coeckelberg and Gunkel, *Comunicative AI*, 22. Original: "Since the LLM is just stringing words together in a way that follows statistical probabilities, it does not know this fact – at least not in any sense of "knowing" as applied to humans".


DOI: 10.59853/RAT62-2025-0201
ISSN: 1135-0547
Año XXXI - Nº 62 - 2025 / julio-diciembre

cionan nuestros viejos cerebros[51]. En la misma línea ya en el año 2000 Daniel Jurafsky y James H. Martin plantearon cómo los seres humanos tienen la tendencia a actuar ante los ordenadores como si estos fueran personas:

> Independientemente de lo que las personas crean o sepan sobre el funcionamiento interno de los ordenadores, hablan de ellos e interactúan con ellos como entidades sociales. Las personas se comportan con ellos como si fueran personas, son educadas con ellos, los tratan como miembros de un equipo y esperan, entre otras cosas, que los ordenadores puedan comprender sus necesidades.[52]

Y cuando en ese ordenador con el que interactuamos nos encontramos con el lenguaje natural de un LLMs la ilusión es completa. Para darnos cuenta de la realidad de lo que ocurre con estos LLMs podemos volver la mirada al trabajo de Emily Bender y lo que ella, junto a otros colaboradores, denomina loros estocásticos[53]:

---

[51] Byron Reeves y Clifford Nass, *The Media Equation. How People Treat Computers, Television, and New Media Like Real People and Places* (Cambridge: CSLI, 1996), 12. En 1944 Fritz Heider y Mary-Ann Simmel realizaron un experimento para demostrar como un hecho natural en los seres humanos la antropomorfización de cualquier ser que sea percibido como animado. Su experimento consistía en mostrar figuras geométricas moviéndose por una pantalla. Los espectadores acaban interpretándolas en clave antropomórfica, y suponiendo en esos movimientos diversas motivaciones: miedo, necesidad, lucha, huida. Acceso el 22 de noviembre de 2025, *http://www.youtube.com/watch?v=n9TWwG4SFWQ*.

[52] Daniel Jurafsky and James H. Martin, *Speech and Language Processing. An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition* (Upper Saddle River – New Jersey: Prentice Hall, 2000), 8-9. Original: "Regardless of what people believe or know about the inner workings of computers, they talk about them and interact with them as social entities. People act toward computers as if they were people, they are polite to them, treat them as team members, and expect among other things that computers should be able to understand their needs".

[53] La definición de los LLMs como loros estocásticos, *Stochastic Parrots*, la encontramos en el artículo: Emily M. Bender, Timnit Gebru, Angelina McMillan-Major y Shmargaret Shmitchell, "On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?", *FAccT'21: Proceedings of the 2021 ACm Conference on Fairness, Accountability, and Transparency* (March 2021): 610-623. La propia Emily M. Bender en sus presentaciones sobre este tema denomina a los LLMs como "*Synthetic text extruding machines*" (Emily M. Bender, "Synthetic text extruding machines: A linguist-eye view on their narrow range of applicability", University of Washington, acceso el 23 de noviembre de 2025, *https://faculty.washington.edu/ebender/*

> El texto generado por un LM no se basa en la intención comunicativa, ningún modelo del mundo, ni ningún modelo del estado mental del lector. [...] pero tenemos que dar cuenta del hecho de que nuestra percepción del texto en lenguaje natural, independientemente de cómo se generó, está mediada por nuestra propia competencia lingüística y nuestra predisposición a interpretar los actos comunicativos como portadores de significado e intención coherentes, lo hagan o no [...]. Al contrario de lo que puede parecer cuando observamos su resultado, un LM es un sistema para unir aleatoriamente secuencias de formas lingüísticas que ha observado en sus amplios datos con los que ha sido entrenado, de acuerdo con información probabilística sobre cómo se combinan, pero sin ninguna referencia al significado: un loro estocástico.[54]

## Los costes humanos y ambientales de los grandes modelos de lenguaje

### *Trabajo fantasma*

Las tecnologías etiquetadas como IA comparten la pretensión de desplazar y sustituir el trabajo humano, sin embargo, la realidad es que se apoyan en el trabajo de una enorme cantidad de personas que permanecen ocultas. En 2019 esta realidad fue puesta sobre la mesa por Mary L. Gray y Siddharth Suri en su obra *Ghost Work. How to Stop Silicon Valley from Building a New Global Underclass*[55]. ¿A qué denominan estos autores trabajo fantasma? A los trabajadores que hacen posible el funcionamiento de muchas aplicacio-

*papers/Linguist-eye-view.pdf*. Las expresiones conjuntas de *máquinas extrusoras de texto sintético* y la de *loros estocásticos* expresan para Bender lo que verdaderamente son los LLMs.

54 Bender, Gebru, McMillan-Major y Shmitchell, "Stochastic Parrots", 616-617. Original: "Text generated by an LM is not grounded in communicative intent, any model of the world, or any model of the reader's state of mind. [...] but we have to account for the fact that our perception of natural language text, regardless of how it was generated, is mediated by our own linguistic competence and our predisposition to interpret communicative acts as conveying coherent meaning and intent, whether or not they do [...]. Contrary to how it may seem when we observe its output, an LM is a system for haphazardly stitching together sequences of linguistic forms it has observed in its vast training data, according to probabilistic information about how they combine, but without any reference to meaning: a stochastic parrot".

55 Mary L. Gray and Siddharth Suri, *Ghost Work. How to Stop Silicon Valley from Building a New Global Underclass* (Boston – New York: Houghton Mifflin Harcourt, 2019).


DOI: 10.59853/RAT62-2025-0201
ISSN: 1135-0547
Año XXXI - Nº 62 - 2025 / julio-diciembre

nes, sitios web y sistemas de inteligencia sin ser vistos, incluso conscientemente mantenidos en la sombra[56]. Basado en el trabajo online bajo demanda de millones de trabajadores, en diferentes partes del mundo, que realizan tareas sin las cuales buen número de funciones y aplicaciones que damos por sentadas no podrían ser llevadas a cabo. Estos autores citan, entre otros, el ejemplo del reconocimiento facial de Uber a sus conductores, cuando el algoritmo tiene dificultades con una identificación va a ser una persona en algún lugar del mundo la que en cuestión de segundos compare la imagen del conductor actual con la registrada en Uber, y dé una respuesta de la que depende el trabajo del conductor[57]. Pero este trabajo fantasma se realiza en unas condiciones que los autores denominan crueldad algorítmica[58], por las condiciones de inseguridad y explotación que ellos descubren tras ese tipo de trabajo online bajo demanda.

Los LLMs no son una excepción en cuanto al denominado trabajo fantasma o, como Bender y Hanna prefieren decir, trabajo oculto. Un ejemplo es OpenAI contratando trabajadores keniatas, a menos de dos dólares al día[59], para la revisión y clasificación de contenidos para entrenar a sus modelos de lenguaje. Para Karen Hao el éxito de la IA generativa ha ido de la mano con una explotación laboral a gran escala[60], ella afirma que esto es algo en común con los grandes imperios, cuya rápida acumulación de riqueza se basaría según Hao en pagar muy poco, e incluso nada, a una amplia mano de obra sobre la que se asienta. Pone de ejemplo a Venezuela de cómo las empresas de trabajo bajo demanda online buscan lugares en crisis económica donde resulte cada vez más barato el trabajo. Venezuela es un caso especialmente significativo para Hao, porque aunque las grandes compañías suelen buscar países donde el inglés sea al menos una de las lenguas, tal como ocurre en Kenia, Filipinas o India; la desesperación de muchos trabajadores venezolanos con buena formación les llevaba a trabajar por ínfimas cantidades de dinero, permitien-

---

56 "The human labor powering many mobile phone apps, websites, and artificial intelligence systems can be hard to see – in fact, it's often intentionally hidden" (Gray and Suri, *Ghost Work*, ix).

57 Gray and Suri, *Ghost Work*, xv-xvi.

58 Gray and Suri, *Ghost Work*, xxx.

59 Bender and Hanna, *The AI Con*, 58-64.

60 "Wide-scale labor exploitation" (Hao, *Empire of AI*, 194).

do a las compañías que los contrataban ampliar sus beneficios[61]. En opinión de Hao los criterios para la búsqueda de estos trabajadores serían: un elevado porcentaje de población con un buen nivel educativo y buena conexión a Internet, pero lo suficientemente pobres como para trabajar por muy poco dinero[62].

## *El coste ambiental*

Las compañías de IA generativa necesitan una constante inyección de capital, de miles de millones de dólares que les permitan mantenerse en marcha; a la vez que un incremento en el consumo de recursos, como muestra la imparable expansión de los centros de datos y de consumo de energía. Esto último, el consumo de energía ha llevado a recurrir a la energía nuclear por parte de Amazon, Google y Microsoft para alimentar centros de datos[63].

En junio de 2019 Emma Strubell, Ananya Ganesh y Andrew McCallum publicaron el artículo titulado "Energy and Policy Considerations for Deep Learning in NLP"[64], en el que presentaban sus estimaciones sobre el creciente consumo de energía de los LLMs. Este artículo y el de Emily Bender y Timnit Gebru ya citado, "Stochastic Parrots", fueron respondidos por las compañías implicadas con el cuestionamiento de los cálculos realizados. Karen Hao comenta cómo por parte de Google se respondió a través de Jeff Dean, quien propuso a Emma Strubell, coautora del artículo original, colaborar con un equipo de Google para mostrar cuál era su verdadera huella de carbón. Stru-

---

61 Hao, *Empire of AI*, 195-197.

62 Hao, *Empire of AI*, 203.

63 João da Silva, "Google turns to nuclear to power AI data centres", BBC, acceso el 22 de noviembre de 2025, https://www.bbc.com/news/articles/c748gn94k950. "Hungry for Energy Amazon, Google, and Microsoft Turn to Nuclear Power", The New York Times, acceso el 22 de noviembre de 2025, https://www.nytimes.com/2024/10/16/business/energy-environment/amazon-google-microsoft-nuclear-energy.html. "Amazon's $20B Nuclear-Powered Data Center Initiative in Pennsylvania", DATACENTERS.COM, acceso el 22 de noviembre de 2025, https://www.datacenters.com/news/amazon-s-20b-nuclear-powered-data-center-initiative-in-pennsylvania.

64 Emma Strubell, Ananya Ganesh, and Andrew McCallum, "Energy and Policy Considerations for Deep Learning in NLP", *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics* (June 2019): 3645-3650.


DOI: 10.59853/RAT62-2025-0201
ISSN: 1135-0547
Año XXXI - Nº 62 - 2025 / julio-diciembre

bell declinó su participación[65]. El resultado fue un artículo encabezado Dave Patterson titulado "The Carbon Footprint of Machine Learning Training Will Plateau, Then Shrink"[66].

La trampa del lenguaje que hemos comentado anteriormente tiene aquí también su reflejo cuando nos olvidamos de que tras expresiones como "la nube", lo que se encuentra son instalaciones de ordenadores, centros de datos, que Bender y Hanna describen como "ordenadores hambrientos de energía"[67]. Estos autores estiman el consumo global de estos centros de datos para 2034 en 1.580 Tw (1.580.000.000.000.000 w), el mismo consumo que un país como India.

## La inteligencia artificial en una sociedad digital

El sociólogo Armin Nassehi ha desarrollado lo que él denomina la primera teoría social de la sociedad digital[68], remarcando que no es equivalente a lo que se plantea como una teoría de la digitalización[69] cuyo eje central radica en que la cuestión de la digitalización es el de cómo la sociedad va adoptando la tecnología. Su perspectiva es diferente, para él "El problema de referencia de la digitalización se sitúa en la estructura misma de la sociedad".[70] Por eso comienza su estudio afirmando que la sociedad actual no es digital por haber adoptado la tecnología digital, sino que es ella misma digital antes de haber adoptado la tecnología, siendo esta una consecuencia de la propia digitalización de la sociedad: "La sociedad moderna siempre ha sido digital,

---

65 Hao, *Empire of AI*, 170-173.

66 David Patterson, Joseph Gonzalez, Urs Hölzle, Quoc Hung Le, Chen Liang, Lluis-Miquel Munguia, "The Carbon Footprint of Machine Learning Training Will Plateau, Then Shrink", TechRxiv, 2022, acceso el 22 de noviembre de 2025, https://doi.org/10.36227/techrxiv.19139645.v2.

67 Bender and Hanna, *The AI Con*, 157. Original: "power-hungry computers".

68 "Nichts weniger ist das Programm, alsdie erste Gesellschaftstheorie der digitalenGesellschaft vorzulegen" (Nassehi, *Muster*, 26). Traducción: "El plan es nada menos que presentar la primera teoría social de la sociedad digital".

69 Él mismo plantea como aquí se distancia de estudios como los de Sherry Turkle y Shoshana Zuboff (Nassehi, *Muster*, 12-15).

70 Nassehi, *Muster*, 318. Original: "Das Bezugsproblem der Digitalisierung ist in der Gesellschaftsstruktur selbts lokalisiert".

y por tanto, la técnica digital es, después de todo, solo la consecuencia lógica de una sociedad que es digital en su misma estructura fundamental".[71] Esta idea es compartida por otros autores, el mismo Weizebaum planteaba que la transformación del mundo a la medida del ordenador es algo que comenzó mucho antes de que los propios ordenadores existiesen, siendo estos fruto de esa mentalidad que ha ido redefiniendo al ser humano y a la sociedad, y que al dar a luz al ordenador ha acelerado ese mismo proceso en una retroalimentación continua[72]. También el filósofo Daniel Innerarity se manifiesta de un modo similar[73]. Y lo hace llevando la IA al lugar de consecuencia de todo ese proceso de digitalización anterior a las propias tecnologías digitales, en su opinión "la inteligencia artificial parece llamada a ser la lógica de legitimación de los gobiernos y de las sociedades digitales"[74].

## Conclusión

El término IA es una etiqueta tras la cual se agrupan diferentes tecnologías. Una etiqueta que les confiere un halo poderoso y mitológico. La esencia de ese mito no es sino el de considerarla realmente inteligente y estar en camino de igualar primero y superar después a la inteligencia humana, en el fondo al mismo ser humano[75]. Un mito que permite mantener siempre la tensión hacia un futuro que desvía nuestra mirada de los retos del presente. Ya en 1965 Herbert Simon predijo que en 1985 los ordenadores serían capaces de ejecutar cualquier tarea que realizase un ser humano. Poco más tarde, en 1967, Marvin Minsky predijo que en el plazo de una generación se alcanzarían las metas principales de la IA[76]. Este tipo de profecías incumplidas se

---

71 Nassehi, *Muster*, 11. La cursiva pertenece al original. Original: "Die gesellschaftliche Moderne immer schon digital war, dass die Digitaltechnik also letzlich nur die logische Konsequenz eimer in ihrer Grundstruktur *digital* gebauten Gesellschaft ist".

72 Weizenbaum, *Computer Power*, ix.

73 Innerarity, *Una teoría crítica*, 17-24. En este caso desde la perspectiva de la digitalización como paso siguiente al sistema burocrático de los estados.

74 Innerarity, *Una teoría crítica*, 20.

75 Erik J. Larson define el mito de la IA como la creencia en la llegada inevitable de una IA general que se sitúe al nivel del ser humano y posteriormente de una superinteligencia que la supere: Erik J. Larson, *The Myth of Artificial Intelligence. Why Computers Can't Think the Way We Do* (Cambridge – London: The Belknap Press of Harvard University Press, 2021), 1.

76 Philip Brey, "Hubert Dreyfus: Humans versus Computers", in *American Philosophy of*

han seguido sucediendo y están presentes en la actualidad, desde las de Elon Musk con la llegada inminente de los coches autónomos[77], hasta las del CEO de OpenAI Sam Altman en torno al futuro de la IA, tal como que un niño nacido hoy nunca será más inteligente que la IA[78], o que en el año 2030 o antes superará la IA a la humana[79]. Esto lo ha expresado Jenna Burrell del siguiente modo:

Dicen que "el poder reside en la máquina". Nos están distrayendo de la enorme riqueza y poder que están ganando con el auge de la IA, y del hecho de que no tiene por qué ser así. Les gustaría pasar por alto el verdadero trabajo de hoy, el afrontar los desafíos y lidiar con las posibilidades que moldearán el mundo que está por venir. En cambio, nos piden que miremos hacia el futuro, pero olvidan que el futuro nos pertenece a todos.[80]

Jenna Burrell lo plantea con claridad, esas profecías nos distraen de lo que realmente es la IA y lo que está en juego con el poder que se está acumulando en torno a los grandes desarrollos tecnológicos de nuestro mundo. Y mien-

---

*Technology*, ed. Hans Achterhuis (Bloomington-IN: Indiana University Press, 2001), 37-38.

77 "Here are Elon Musk's wildest predictions about Tesla's self-driving cars", The Verge, accessed November 22, 2025, *https://www.theverge.com/2019/4/22/18510828/tesla-elon-musk-autonomy-day-investor-comments-self-driving-cars-predictions.*

78 "Sam Altman makes chilling AI prediction: 'A child born today will never know any other world,' offers parenting advice for the new age", The Economic Times, accessed November 22, 2025, *https://economictimes.indiatimes.com/magazines/panache/sam-altman-makes-chilling-ai-prediction-a-child-born-today-will-never-know-any-other-world-offers-parenting-advice-for-the-new-age/articleshow/123239246.cms?utm_source=contentofinterest&utm_medium=text&utm_campaign=cppst.*

79 "Sam Altman predicts AI will surpass human intelligence by 2030", Business Insider, accessed November 22, 2025, *https://www.businessinsider.com/sam-altman-predicts-ai-agi-surpass-human-intelligence-2030-2025-9.*

80 Jenna Burrell, "Artificial Intelligence and the Ever-Receding Horizon of the Future," Tech Policy Press, (June 2023) accessed November 22, 2025, *https://www.techpolicy.press/artificial-intelligence-and-the-ever-receding-horizon-of-the-future/.* Original: "They are saying "the power resides in the machine." They are distracting us from the enormous amount of wealth and power they stand to gain from the rise of AI — and from the fact that it doesn't have to be this way. They would like to skip past today's real work, that of tackling the challenges and wrestling with the possibilities that will shape the world to come. Instead, they ask us to look to the future — but they forget that the future belongs to us all".

tras prometen resolver en el futuro todos los problemas, quedan oscurecidos los que ya hoy se crean con esas tecnologías, como los que hemos visto del coste humano y ambiental. A esta labor de distracción no son ajenos algunos de los planteamientos éticos que surgen en torno a la IA, son aquellos que caen en la utilización interesada de la ética, en un *ethics washing* que Daniel Innerarity explica así:

Ese lavado de cara ético forma parte del problema, porque apela a un uso que deja todo como estaba, que hace depender el sentido de la tecnología del modo en que es utilizada, como si no hubiera un condicionamiento estructural, y que parece desconocer la complejidad tecnológica.[81]

Esa clave de centrar la reflexión ética en el uso como forma de no abordar los verdaderos desafíos éticos de la IA, pertenece a la pretensión de una neutralidad de la técnica que no existe. Las tecnologías que desarrollamos son fruto de decisiones humanas, no realidades inevitables, que en su misma concepción y desarrollo ya han de verse sometidas al escrutinio ético de todo lo humano. Hemos visto el coste humano y ambiental del desarrollo de algunas tecnologías. No podemos solamente quedarnos con el resultado y pensar en cuál es su uso adecuado. La pregunta debe englobar también su propia concepción y desarrollo. Quedarnos en el uso de una determinada tecnología es aceptar como inevitable su existencia.

La IA no es inteligente, no desarrollará conciencia ni autonomía respecto al ser humano. La IA es una tecnología desarrollada por el ser humano, que siguiendo a Arvind Narayanan y Sayash Kapoor, podemos denominar como normal[82]. No porque no tenga y vaya a tener un gran impacto en la sociedad, sino porque hay que desmitificarla, liberarla de la tendencia a considerarla cualitativamente distinta, incluso autónoma o, como algunos han llegado a considerarla, un riesgo existencial para la humanidad. El riesgo existencial para la humanidad no estará en la IA sino en quienes la controlen, que siempre serán humanos.

---

[81] Daniel Innerarity, *Una teoría crítica de la* inteligencia *artificial* (Barcelona: Galaxia Gutenberg, 2025), 31.

[82] Arvind Narayanan and Sayash Kapoor, “AI as Normal Technology. An alternative to the vision of AI as a potential superintelligence” Knight First Amendment Institute at Columbia University, April 15, 2025, *https://knightcolumbia.org/content/ai-as-normal-technology.*


DOI: 10.59853/RAT62-2025-0201
ISSN: 1135-0547
Año XXXI - Nº 62 - 2025 / julio-diciembre

## Bibliografía

Bender, Emily M., Timnit Gebru, Angelina McMillan-Major and Shmargaret Shmitchell, "On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?", *FAccT'21: Proceedings of the 2021 ACm Conference on Fairness, Accountability, and Transparency* (March 2021): 610-623.

Bender, Emily M. and Alex Hanna. *The AI Con. How to Fight Big Tech's Hype and Create the Future We Want*. New York: Harper, 2025.

Brey, Philip. "Hubert Dreyfus: Humans versus Computers". In *American Philosophy of Technology*, edited by Hans Achterhuis, 37-63. Bloomington-IN: Indiana University Press, 200.

Burrell, Jenna. "Artificial Intelligence and the Ever-Receding Horizon of the Future." *Tech Policy Press*, (June 2023) accessed November 22, 2025, https://www.techpolicy.press/artificial-intelligence-and-the-ever-receding-horizon-of-the-future/.

Byron Reeves and Clifford Nass. *The Media Equation. How People Treat Computers, Television, and New Media Like Real People and Places*. Cambridge: CSLI, 1996.

Churchland, Paul M. and Patricia Smith Churchland. "Could a Machine Think? Classical AI is unlikely to yield conscious". *Scientific American* (February 1990): 32-37.

Coeckelberg, Mark and David J. Gunkel. *Comunicative AI. A Critical Introduction to Large Language Models*. Cambridge: Polity Press, 2025.

Crevier, Daniel. *AI. The Tumultuous History of the Search for Artificial Intelligence*. New York: Basic Books, 1993.

Dreyfus, Hubert L. *Alchemy and Artificial Intelligence*. Santa Mónica-CA: RAND Corporation, 1965. http://www.rand.org/content/dam/rand/pubs/papers/2006/P3244.pdf.

Dreyfus, Hubert L. and Stuart E. Dreyfus. *Mind Over Machine. The Power of Human Intuition and Expertise in the Era of the Computer*. New York: The Free Press, 1986.

Dreyfus, Hubert L. *What Computers Still Can't Do*. Cambridge: The MIT Press, 1994.

Garduño Alvarado, Tzolkin, Feliú Sagols Trancoso y Gunnar Wolf. "Las neuronas artificiales y su papel central en la inteligencia artificial". *Ciencia* 76, no. 1 (enero-marzo 2025): 68-75.

Gray, Mary L. and Siddharth Suri. *Ghost Work. How to Stop Silicon Valley from Building a New Global Underclass*. Boston – New York: Houghton Mifflin Harcourt, 2019.

Grazer, Brian, and Charles Fishman. *A Curious Mind: The Secret to a Bigger Life*. New York: Simon & Schuster, 2015.

Hao, Karen. *Empire of AI. Inside the reckless race for total domination*. London: Allen Lane, 2025.

Hofstadter, Douglas R. *Gödel, Escher, Bach: an Eternal Golden Braid*. New York: Basic Books, 1979.

Innerarity, Daniel. *Una teoría crítica de la inteligencia artificial*. Barcelona: Galaxia Gutenberg, 2025.

Julia, Luc. *IA génératives, pas creatives*. Paris: Le Cherche Midi, 2025.

Jurafsky, Daniel and James H. Martin. *Speech and Language Processing. An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*. Upper Saddle River – New Jersey: Prentice-Hall, 2000.

Kurzweil, Ray. "Locked in His Chinese Room". In *Are We Spiritual Machines? Ray Kurzweil vs. The Critics of Strong AI*, edited by Jay Richards, 128-163. Seattle: Discovery Institute, 2002.

Kurzweil, Ray. *The Singularity is Near*. New York: Penguin Books, 2006.

La Fontaine, Gustavo, " Sobre loros estocásticos. Una mirada a los modelos grandes de lenguaje", *LÓGOI Revista de Filosofía*, no. 45 (enero-junio 2024): 75-87.

Larson, Erik J. *The Myth of Artificial Intelligence. Why Computers Can't Think the Way We Do*. Cambridge – London: The Belknap Press of Harvard University Press, 2021.

Levy, David. *Robots Unlimited. Life in a Virtual Age*. Wellesley-Massachusetts: A K Peters, 2006.

McCorduck, Pamela. *Machines Who Think*. Natick-Massachusetts: A K Peters, 2004.

Minsky, Marvin. *The Society of Mind*. New York: Simon & Schuster, 1986.

Narayanan, Arvind and Sayash Kapoor. "AI as Normal Technology. An alternative to the vision of AI as a potential superintelligence". *Knight First Amendment Institute at Columbia University*, April 15, 2025, https://knightcolumbia.org/content/ai-as-normal-technology.

Nassehi, Armin. *Muster. Theorie der digitalin Gesellschaft*. Bonn: C. H. Beck, 2019.

Nida-Rümelin, Julian und Nathalie Weidenfeld. *Digitaler Humanismus. Eine Ethik für das Zeitalter der Künstlichen Intelligenz*. München: Piper, 2018.

Nilson, Nils J. *The Quest for Artificial Intelligence. A History of Ideas and Achievement*. Cambridge: Cambridge University Press, 2010.

Patterson, David, Joseph Gonzalez, Urs Hölzle, Quoc Hung Le, Chen Liang, Lluis-Miquel Munguia, "The Carbon Footprint of Machine Learning Training Will Plateau, Then Shrink", *TechRxiv*, 2022, acceso el 22 de noviembre de 2025, https://doi.org/10.36227/techrxiv.19139645.v2.

Penrose, Roger. *The Emperor's New Mind*. Oxford: Oxford University Press, 1989.

Searle, John R. "Minds, brains, and programs". *Behavioral and Brain Sciences* 3, no. 3 (1980): 417-457.

Searle, John. "I Married a Computer". In *Are We Spiritual Machines? Ray Kurzweil vs. The Critics of Strong AI*, edited by Jay Richards, 56-76. Seattle: Discovery Institute, 2002.

Strubell, Emma, Ananya Ganesh, and Andrew McCallum. "Energy and Policy Considerations for Deep Learning in NLP". *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics* (June 2019): 3645-3650.

Turing, Alan. "Computing Machinery and Intelligence." *Mind* 59, no. 236 (October 1950): 433-460.

Weizenbaum, Joseph. "ELIZA–A Computer Program For the Study of Natural Language. Communication Between Man And Machine". *Communications of the ACM* 9, no. 1 (January 1966): 36-45, https://dl.acm.org/doi/10.1145/365153.365168.

Weizenbaum, Joseph. *Computer Power and Human Reason. From Judgement to Calculation*. Hardmonsworth: Penguin Books, 1984.

Centro Regional de Estudios Teológicos de Aragón