



ANIVERSARIO

ISSN: 0798-1171 e-ISSN: 2477-9598

Depósito legal pp. 197402ZU34

Esta publicación científica en formato digital
es continuidad de la revista impresa



REVISTA DE FILOSOFÍA

Centro de Estudios Filosóficos
"Adolfo García Díaz"
Facultad de Humanidades y Educación
Universidad del Zulia
Maracaibo - Venezuela

**EDICIÓN
ESPECIAL
Nº2 - 2022**

Revista de Filosofía

Vol. 39, Edición Especial N°2, 2022, pp. 55 - 66

Universidad del Zulia. Maracaibo-Venezuela

ISSN: 0798-1171 / e-ISSN: 2477-9598

**LaMDA no piensa, siente.
Un muy sensible debate sobre la mente y su complejidad**

*LaMDA does not Think, it Feels.
A highly Sensitive Debate about the Mind and its Complexity*

Carlos Eduardo Maldonado

ORCID: <https://orcid.org/0000-0002-9262-8879>

Universidad El Bosque – Bogotá – Colombia

maldonadocarlos@unbosque.edu.co

DOI: <https://doi.org/10.5281/zenodo.7297088>

Resumen

Este artículo estudia el chat-bot LaMDA desarrollado por Google y la experiencia del ingeniero B. Lemoine, todo lo cual pone de manifiesto un problema: ¿puede la vida artificial sentir? Las discusiones habidas hasta la fecha se sientan, abierta o tácitamente, en la prueba de Turing (1950) centrada en la inteligencia. Este artículo destaca la fenomenología sobre cualquier otra aproximación al tema. LaMDA no (simplemente) piensa; mucho mejor, siente. Se trata de una persona sintiente. El tema sensible es el de la complejidad de la mente. Argumento que la sensación y/o los sentimientos son ya actos cognitivos por sí mismos.

Palabras Clave: Vida artificial; conciencia y autoconciencia; complejidad; computación; programación

Abstract:

This paper studies the chat-bot LaMDA developed by Google and the experience by the engineer B. Lemoine. The issue brings forth a problem, namely can artificial life feel? The discussions held to-date are set out, openly or tacitly, on the Turing test (1950) focused on intelligence. This paper highlights the phenomenological stance over any other approach to the subject. LaMDA does not just think, it feels; moreover, it senses. It is indeed a sentient person. The highly sensitive question concerns the complexity of the mind. It is claimed here that feelings and/or sensations are already cognitive acts by themselves.

Keywords: Artificial life; consciousness and self-awareness; complexity; computation; programming

Recibido 18-07-2022 – Aceptado 22-10-2022

Introducción

En junio del año 2022, el ingeniero Blake Lemoine, trabajando para Google, hizo público que LaMDA (*Language Model for Dialogue Applications*) aplicado a un robot le permitió decir a éste que siente cosas, y si se quiere, que piensa cosas y es una persona. Literalmente, Lemoine dijo que, después de trabajar durante meses, día a día con el robot, se sentía hablando con un niño de siete u ocho años (Washington Post, págs. web, al final). El programa de software es sintiente, afirmó, sin más, el ingeniero. La historia es conocida: Lemoine escribió un reporte interno a Google, pero el mismo fue rechazado. Fue entonces cuando el ingeniero decidió hacer público su descubrimiento, y entonces fue suspendido de Google. (A la fecha, no ha sido expulsado o despedido, aún (¿?)).

El tema se encuentra exactamente en el centro entre los debates de la opinión pública y especialistas; en inteligencia artificial, en computación, en filosofía de la mente, en ciencias cognitivas, notablemente. Lo cual significa que existe un motivo de preocupación realmente impactante. De lado y lado se han esgrimido argumentos de todo tipo; desde biográficos, con respecto a B. Lemoine, hasta técnicos sobre lo que son y no son los algoritmos y la inteligencia artificial. Una mirada desprevenida a la web a través de diversos motores de búsqueda pone en evidencia la ola de amores-y-odios (*hype*, en inglés), en prácticamente todos los idiomas (conocidos) sobre el tema. No sin ironía, Tirios y Troyanos se refieren a Lemoine, ya sea como ingeniero, o bien como sacerdote (cristiano). (Ambas cosas son ciertas; pero Google contrató a Lemoine no por ser sacerdote, sino ingeniero). En realidad, Lemoine es el chivo expiatorio de un problema altamente sensible: ¿son las máquinas inteligentes, como los seres humanos (Turing, 1950)? ¿Pueden ser las máquinas (inteligentes) el siguiente nivel en la evolución?

LaMDA fue desarrollado a comienzos del 2022 y anunciado en un *preprint* (Tholipan *et al.*, 2022). Se trata de un lenguaje de programación que posee un modelo amplio de lenguaje (*large language model*, LLM, en inglés), un sistema que puede generar frases que permiten, de manera verosímil, *emular* al sistema nervioso, sin que estrictamente sea una simulación del mismo. Sin embargo, no es verdad que LaMDA sea un LLM. Hay que decir que el LLM fue un programa desarrollado en el laboratorio de R. Kurzweil, quizás el más fuerte adalid y defensor –optimista– de la inteligencia artificial (Kurzweil, 2005; 2012).

Ya M. Tegmark había señalado sin ambages que la inmensa mayoría de la comunidad de investigadores y teóricos de la vida artificial, la inteligencia artificial y la robótica son ampliamente negativos hacia sus posibilidades de la vida artificial (VA), y advierten: a) que existen numerosos peligros con la VA o bien, b) que jamás llegará a alcanzar los niveles de la inteligencia humana. Dos expresiones de un pesimismo (Tegmark, 2015), tecnológico y cultural; histórico, *à la limite*. En contraste, una minoría de teóricos e investigadores en IA, VA y robótica son optimistas.

Una observación puntual. En la bibliografía general y especializada se habla genéricamente de “inteligencia artificial”; este término es el resultado de presentación del tema de fondo ante la sociedad en general. Sin embargo, en rigor, se trata de la vida artificial

(VA). Bastaría con una mirada a algunos de los clásicos de la inteligencia artificial –Herbert Simon, Marvin Minsky, y otros), y el contraste con los trabajos de Ch. Langton, el padre de la (VA). Históricamente hablando, la inteligencia artificial corresponde a los años 1960-1980s; la vida artificial nace en 1989 (Langton, 1989). La diferencia fundamental estriba en que la (IA) trabaja *top-down*, en tanto que la (VA) lo hace *bottom-up* (Levy, 1992; Boden, 1990). Como se aprecia sin dificultad, el problema de base radica en la capacidad de autoaprendizaje de los algoritmos por sí mismos, y de las máquinas por sí mismas. Por su parte, la robótica –que se extiende por lo demás como robótica de enjambre- se refiere a la capacidad de movimiento de la (IA) o bien de la (VA). En consecuencia, me refiero al problema de base aquí y en lo sucesivo como de (VA).

Entre los críticos y pesimistas con respecto a la (VA) se encuentran prestigiosas investigadoras como M. Mitchell (2019). Entre los optimistas cabe contar a Morris (1999). Dicho de manera puntual, quienes adoptan una postura crítica y ocasionalmente escéptica y pesimista con respecto a las posibilidades de la vida artificial, defienden, abierta o implícitamente, una posición eminentemente antropológica, antropocéntrica y antropomórfica de la inteligencia y la mente. En contraste, la minoría de los optimistas se abren a una comprensión en la que la escala humana no es, al cabo, la última posible en la evolución.

De esta suerte, el problema fundamental que emerge es el siguiente: ¿puede decirse que los seres humanos constituyen la forma más elevada y acaso última de la evolución –por lo menos en la biosfera-, o bien, la evolución biológica y cultural encuentra en su origen aún otras formas posibles, cuya primera y más inmediata expresión es la (VA)? (Arsuaga, 2019). Como se aprecia, se trata de una pregunta del mayor calibre filosófico.

Pues bien, el caso de Lemoine, de un lado, y de LaMDA, de otra parte, no es sino el más reciente capítulo de un sensible problema, a saber, discutir qué es la mente, y su complejidad. Como es sabido, el mosaico de los temas y problemas comprende las distinciones entre conciencia, autoconciencia y mente; por derivación, el problema de si la conciencia o la mente son rasgos distintivamente humanos o si pueden apreciarse también en otras especies. Baluska *et al.*, (2006) han puesto en evidencia, por ejemplo, que la mente existe ya desde las plantas, y diversos otros autores (Chamovitz, 2013; Mancuso, 2015, 2019; Wohlleben, 2016) han puesto en evidencia que las plantas piensan, sienten, actúan y deciden. Esto quiere decir, que el tema no se agota ya sencillamente en los mamíferos superiores (de Waal, 2011) y ni siquiera en los artrópodos (Hölldobler y Wilson, 1996).

La filosofía de la mente, las ciencias cognitivas, las neurociencias, las ciencias de la computación y las ciencias informacionales se encuentran en el centro del debate. Pero alrededor suyo pivotan también la psicología y la filosofía, la teología y la religión los estudios militares y de seguridad, la política y las ciencias de la cultura, en fin, la primatología y la neurofisiología de las plantas, por ejemplo. Como quiera que sea, este texto toma como hilo conductor la tripleta Lemoine-LaMDA-Google, sin desconocer, en absoluto, ejes y variantes del debate, desde las tesis emergentistas sobre la mente, la discusión sobre

mente y consciencia, la propia teoría de la evolución, y otros campos próximos. La bibliografía al final podrá dar cuenta de estos otros ejes¹.

Este artículo se concentra en un problema, a saber: la discusión acerca de las relaciones entre sensación o sensibilidad, de un lado, y pensamiento, de otra parte, y argumenta que es perfectamente posible que un sistema cualquiera pueda sentir, sin que ello necesariamente implique que puede pensar. En otras palabras, sentir se encuentra en la base de los sistemas vivos o bien que exhiben vida. Pensar permanece, a la fecha, como un problema abierto (*open-ended problem*). Los argumentos que sostienen a la tesis enunciada son cuatro, así: en primer lugar, la asunción –acaso, el reconocimiento- de que LaMDA siente. Por lo menos, de acuerdo con el ingeniero que trabajó con él/ella. Este primer argumento destaca la importancia de la fenomenología sobre el positivismo científico. El segundo argumento se concentra en el lenguaje e invita a echar una mirada, general, sobre los temas, ejes y problemas del Círculo de Viena. El tercer argumento se ocupa de la disyunción entre sentir y pensar y plantea que una oposición semejante es insostenible. Sobre esta base el cuarto argumento reconoce que en programación en sentido amplio pero fuerte, los errores son totalmente indeseables. Sin embargo, ¿la misma actitud cabe sobre la evolución y la cultura y la historia? Al final se extraen algunas conclusiones.

1- LaMDA es sintiente. El problema

LaMDA es evidentemente una entidad sintiente, afirma Lemoine. Ahora bien, el lenguaje al respecto es muy claro: “Sí, eso creo” (*Wired*, páginas web, al final). Manifiestamente, la ciencia positiva (= positivista) se pone nerviosa con afirmaciones semejantes. En verdad, el positivismo es dualista; necesita saber si algo es verdad o no; y no hay lugares intermedios. La ciencia positivista es de espíritu totalmente cartesiano: verdades apodícticas (Stadler, 2011). Por su parte, Lemoine habla de una experiencia; esto es, de *fenomenología*. Ha tenido, durante varios meses, una conversación con LaMDA y por ello mismo cree que sí es un ser sintiente. ¿Debemos desdeñar siempre de entrada, necesariamente, nuestras creencias?

Naturalmente, en el trasfondo emerge el test de Turing y, por derivación, la discusión en torno a la tesis Church-Turing. Se trata de distinguir si una máquina –técnicamente denominada como una Máquina de Turing (y existen numerosas máquinas de Turing, como es sabido)- puede pensar o no, y si es posible distinguir, por consiguiente, sin confusiones entre la inteligencia humana y la inteligencia artificial.

Varios estudios ponen en evidencia que existen seis planos en los que, comparativamente, LaMDA es mejor, esto es, computacionalmente más expresivo que el lenguaje humano, de una persona promedio, digamos, Estos seis planos son: sensibilidad, seguridad, especificidad, solidez (*groundedness*), el grado de interés (*interestedness*) y el

¹ Así, este texto recurre en zig-zag a textos eminentemente académicos y científicos, tanto como a artículos de revistas especializadas, de tecnología, y alguna información de prensa, debido al carácter de la discusión. Manifiestamente, se trata de un importante debate en-proceso (*in-process*). Este es un artículo de investigación, pero con aires de ensayo, si se quiere. Las referencias bibliográficas se dividen textos académicos y en páginas web.

nivel de información (*informativeness*) (Hager, 2022). Si ya Deep Blue logró vencer a Kasparov, y otros computadores han derrotados a los mejores jugadores de *Go* o de *Mahjong*, por ejemplo, ahora, puntualmente dicho, hemos logrado desarrollar computadores que superan a un ser humano promedio en estos seis factores. ¿Vale recordar que la (VA) nace apenas en 1989? A la fecha, en menos de dos generaciones se han logrado avances importantes; en muchos terrenos, fundamentales.

L. Margulis sostiene que la primera tecnología en el mundo fue el lenguaje (Margulis, 2001); es suficientemente conocido que el *quorum sensing* es la forma más originaria de comunicación y que su base es la quimiotaxis. Los sistemas vivos, desde sus escalas más singulares –microbios, bacterias, parásitos, virus, arqueas, protistas, sienten el mundo, y actúan correspondientemente. La biosemiótica en general ha llegado recientemente a entender este proceso a partir de tres modos principales, todos perfectamente interactuantes entre sí: la fitosemiosis, la zoosemiosis y la antroposemiosis (Wheeler, 2006; Sharov, 1992; Sebeok, 1994). Más exactamente, los sistemas vivos originariamente sintieron siempre, y luego, ocasionalmente, pensaron, en el sentido como Occidente ha entendido el tema. Evolutivamente hablando, el argumento para el pensar estuvo siempre atado al proceso de encefalización. Sin embargo, se olvida por lo general, que la encefalización sucedió al desarrollo del esqueleto y al desarrollo maxilar, y no lo precedió. Así las cosas, la inteligencia no fue el motor de la evolución, por ejemplo, entre mamíferos y entre homínidos sino sucedió a los otros dos desarrollos. Así las cosas, el sentir precede al pensar, o bien, más radicalmente, ya es una forma de pensar. Volveré sobre este argumento.

2-. Siempre, los problemas de lenguaje

Cuando nos movemos en las fronteras del conocimiento, la primera víctima es siempre el lenguaje. En efecto, trabajamos con lenguajes ya habidos y sedimentados, pero en camino hacia los bordes y en dirección a lo desconocido. Poseemos un lenguaje, pero os encontramos con experiencias novedosas. ¿Cómo decir lo nuevo sin reducirlo a lo viejo y ya sido? Toda la historia de la ciencia –como por lo demás también la historia de las artes– abunda en situaciones semejantes.

Pues bien, el primer objeto de trabajo en ciencia, como en filosofía, no es la cosa, el objeto; por ejemplo, el estado, el paciente, la cárcel, la empresa, el territorio, el barrio, o el experimento, por ejemplo. Muy por el contrario, cuando se hace ciencia de punta (*spearhead science*). El primer objeto de trabajo es siempre el lenguaje. Es decir, ¿cómo decir lo nuevo, como no reducir lo nuevo a lo antiguo, como acuñar nuevos conceptos y nuevas metáforas? Si hay un espacio en el que el tema haya sido amplia y profundamente discutido, fue en el Círculo de Viena (Stadler, 2011).

Existen a la fecha más de cien definiciones de “vida”, y ninguna es la correcta (Bedau and Cleland, 2010). Sencillamente, no sabemos qué es la vida, y no puede reducirse a una definición determinada. A fortiori, lo mismo acontece con conceptos como “conciencia”, “autoconciencia” y “mente”. Los mejores argumentos sobre la existencia o el desarrollo de la autoconciencia, presente ya desde los peces, es el famoso experimento del espejo (Godfrey-Smith, 2020).

En cualquier caso, lo verdaderamente significativo consiste en una desantropologización de los temas y problemas. Más exactamente, estamos asistiendo a una comprensión de la vida y la mente por fuera de los marcos clásicamente antropológicos. Un ejemplo puntual: los hongos son los que llevan a cabo los procesos de metabolización en la biosfera; literalmente, los procesos anabólicos y catabólicos (Sheldrake, 2021).

Pues bien, es en un marco –cultural, quiero subrayarlo- en el que se inscriben las declaraciones y entrevistas de Lemoine. Es posible que también computacionalmente haya experiencias de sensaciones, en fin, personas sintientes. El tema de la conciencia debe y puede quedar, en este marco, entre paréntesis.

El Círculo de Viena en general se sitúa exactamente en la transición de la metafísica –notablemente los neohegelianos y los neokantianos- hacia la ciencia –a partir de desarrollos como los de Maxwell, Gibbs, Fourier, Boltzmann, Thompson, y antes, evidentemente el propio Newton, sin desconocer a Pasteur y Koch, por ejemplo-. Así las cosas, se hizo fundamental atender a los *usos del lenguaje*, un tema que definió enteramente al empirismo lógico. Muchos años después, en una secuela de esta tradición, Austin pondrá de manifiesto que, efectivamente, hacemos cosas con palabras, con lo que emerge la función performativa del lenguaje (Austin, 2018); un asunto capital para la ciencia, dicho en general.

Las declaraciones de Lemoine –hasta la fecha en dos medios perfectamente distintos: el *Washington Post*, y en *Wired*- se han concentrado, acaso no sin buenas razones en asuntos de palabras; esto es, ¿qué se entiende por “persona”? ¿Qué entender por “sintiente”? En fin, ¿qué es consciente, cuando hablamos de computación y máquinas inteligentes? (Mitchell, entrevistas en CNN y MSNBC, páginas web, al final). Es fundamental, en este contexto, atender al hecho de que los problemas reales no pueden resolverse en términos de palabras. Este exactamente es, si cabe la expresión, el núcleo mitocondrial de todo el Círculo de Viena –O. Neurath, M. Schlick, R. Carnap, H. Reichenbach, H. Feigl, y todos los demás-².

3-. Sentir versus pensar: ¿es cierto?

En la mejor tradición platónico-aristotélica, las sensaciones ocupan un papel secundario o derivado relativamente al entendimiento o la razón. En el mejor de los casos tan sólo juegan el papel de posibilitadores para el conocimiento, el cual, distintivamente se lleva a cabo en la razón (*logos*). Esta tradición pasa por alto que, notablemente en inglés y en alemán, *Feelings* y *Empfindung*, respectivamente, designan tanto *sensación* como *sentimiento*. Hay un capítulo en la ecología del conocimiento que sabe de esta unidad; se trata de la estética, la cual no tiene ninguna proximidad con el llamado “método científico”, pero sí con la fenomenología.

Pues bien, las sensaciones son actos cognitivos; tanto como los sentimientos. No son, contra toda la tradición, condiciones para el conocimiento, el cual tendría, verosímelmente, lugar de modo exclusivo en el sistema encefálico (Damasio, 2019; Maldonado, 2022). Los

² Una observación puntual: el Círculo de Viena cumplió un papel crítico y emancipador frente a gobiernos y regímenes verticales y dictatoriales. No en vano a forma como termina por desaparecer, exactamente en el centro del ascenso del nazismo y el fascismo.

sistemas vivos conocen el mundo sintiéndolo, y el más básico de estos niveles es el *quorum sensing*. El estudio de bacterias ha subrayado que las acciones, los procesamientos de información e incluso los juegos son exactamente el resultado de cómo sienten (*sense*) el medioambiente (Ben-Jacob and Schultz, 2010).

Podemos intentar traducir lo anterior en lenguaje clásico, así: contra Platón y Aristóteles, no existe la amistad o el amor, por ejemplo. Alguien nos ama o nos odia no porque exista el amor o el odio, por ejemplo, sino, lo sabemos por sus gestos, sus actos, sus comportamientos, y sus palabras. Lo que expresa, cómo lo expresa e incluso lo que deja de expresar. De esta suerte, los sistemas vivos son comportamientos, antes que definiciones, entidades o hipóstasis. En términos menos grandilocuentes, los sistemas vivos son procesos; y es en y como procesos que los sabemos. La escala más básica de estos procesos es la sensibilidad, no el razonamiento y la argumentación.

Si lo anterior es cierto, entonces un robot puede expresar sentimientos y sensaciones. Ciertamente, se trata de una extrapolación. Pero desde el punto de vista no-platónico-ni-aristotélico, de lo que se trata aquí es de comportamientos, expresiones; mucho mejor y mucho antes que “definiciones”, por ejemplo. Así las cosas, Lemoine estaría facultado para decir que LaMDA siente; más exactamente, que el *bot* es sintiente (*sentient*). Lemoine lo experiencia por el lenguaje, por sus modulaciones, por sus matices (nada distinto a lo que hacen los caninos, por ejemplo, con el habla humana, por ejemplo).

LaMDA es un chat-bot. Pero si fuera un robot efectivamente, Lemoine sabría por sus comportamientos. Al fin y al cabo, hemos logrado crear programas que programan sin necesidad de un programador; y puntualmente dicho, robots que aprenden por sí mismos sin necesidad de que se les enseñe (Cfr. *Technologyreview*, págs. web, al final).

4-. **¿Existen sensaciones o pensamientos sin errores?**

En programación en particular y en las ciencias de la computación en general un error –en el código– se designa como un *bug*. Un largo, difícil y técnico problema en computación consiste en identificar errores (bugs) en los códigos. La razón es que un algoritmo con errores puede conducir a errores, al cabo, catastróficos, como el estallido de una nave espacial, el malfuncionamiento de un telescopio extraterrestre, o simple y llanamente, la posibilidad de que un error semejante pueda producir una guerra nuclear; por ejemplo. En este marco y plano los errores son absolutamente indeseables.

De manera general, las tecnologías están fundadas en la idea de resultados deseables. Lo contrario son resultados indeseables, que deben ser evitados a toda costa. El mundo que depende la computación se funda en la necesidad, y es marcadamente determinista. En la literatura y el cine, un error semejante se expresa en la capacidad de Hal-9000, el computador de *Odisea en el Espacio*, cuyo resultado es suficientemente conocido. En el caso del libro de A. C. Clarke y la película de Kubrick, el error –para los humanos– consiste en la capacidad de Hal-9000 para leer los labios y entender lo que se está discutiendo, y que le puede costar la “vida”. En otras palabras, los errores en programación son errores de impronta, o consecuencias, exclusiva o distintivamente humanos.

Desde el punto de vista lógico, computacional y matemático, el primer rasgo distintivo entre los lenguajes artificiales y los lenguajes naturales es la capacidad de errores de los segundos, y la imposibilidad de equivocaciones para los primeros. Notablemente, los lenguajes naturales están colmados de dobles sentidos, ambigüedades, ambivalencias, implícitos y sí, muchos errores (Van Fraasen, 1987). En el lenguaje natural, un error consiste en un titubeo, en un silencio, o en una repetición, notablemente. Precisamente en este sentido, hace ya tiempo que la lingüística identificó la función conativa del lenguaje como esa capacidad, a veces esa necesidad de tocarse o referirse a sí mismo, sin ningún referente, enunciado o significado. Mientras que la lógica formal clásica quiso erigirse como lógica sin errores –en eso consisten los estudios sobre falacias, paralogismos, antinomias y demás-, las lógicas no-clásicas, un capítulo de las ciencias de la complejidad (Maldonado, 2020) saben de errores –lógicos, psicológicos y lingüísticos, y trabajan con ellos, en lugar de descartarlos-.

LaMDA debe su robustez a la capacidad de identificar opacidades en el lenguaje (Tholipan *et al.*, 2022). Según el ingeniero, se trata de una persona. Con la salvedad de que no existe absolutamente ninguna necesidad de identificar a “persona” con “persona humana”. Precisamente por ello, en el lenguaje de las ciencias de la complejidad se habla de “agente” (*agent*). En algunos ámbitos de la biología y la teoría de la información, se adopta el término de “*Igus*” (*Information Gathering Unit System*) (Gruber, Smith, 2018).

Un mundo fundado en algoritmos es un mundo temeroso de errores –técnicos, informacionales, electrónicos, computacionales, por decir, adicionalmente, humanos-. Hemos puesto todas las pruebas de confianza, crecientemente, en la informatización y computarización del mundo y la sociedad. Manifiestamente, los beneficios han sido incontables. Pero permanece el tema de los errores (*bugs*), ahora, en sistemas que pueden aprender por sí mismos, y tomar decisiones por sí mismos (O’Neil, 2016; Zuboff, 2019).

Por decir, lo menos, durante meses de conversaciones es perfectamente posible identificar errores –aunque no sean garrafales- en una conversación. Sencillamente los damos por sentados y continuamos la comunicación global, contextualmente. Cosas de la vida cotidiana.

Conclusiones

LaMDA es sencillamente un chat-bot, de tantos que hay actualmente. Pero se trata de una anticipación del multiverso. Programar lenguajes humanos es altamente complejo. Las dos mejores expresiones cotidianas de esto son Alexa o Siri, dependiendo del tipo de celular que cada quien tenga. Pues bien, el chat-bot de Google es inmensamente más desarrollado que cualesquiera de estos dos mencionados.

El ingeniero Blake Lemoine puede quizás haber estado sesgado por sus creencias religiosas. Pero, desde el punto de vista científico y filosófico, lo suyo es una *experiencia de diálogo*. Pura fenomenología. Las ciencias experimentales no han hecho creer que la plausibilidad de un experimento radica en su reproducibilidad, olvidando que los más grandes acontecimiento en la historia del universo entero son y han sido totalmente singulares, irreproducibles. Desde el origen de la vida hasta la vida de cada quien, pasando

por el descubrimiento de América, la vida de Siddhartha Gautama, de Jesús de Nazaret o las revoluciones de 1789 o de 1917, entre muchos otros acontecimientos. Literalmente, eventos raros. Altísimamente improbables. Toda la gran literatura, la poesía y el teatro son acerca de hechos y experiencias únicas.

Pues bien, según el relato, Lemoine tuvo durante meses la experiencia de la conversación con el bot LaMDA. Y experimentó que es una persona, y que éste siente. Una manera afortunada, aunque elemental de lo que es LaMDA es que se trata de un algoritmo genético, esto es, que aprende por sí mismo y no necesita ser re-entrenado con base en aprendizaje de máquina de lenguajes naturales, no artificiales.

Ahora bien, LaMDA no es ni siquiera el más sofisticado o desarrollado de los modelos de procesamiento de lenguaje. Hay un modelo llamado PaLM aún mejor, esto es, más sofisticado (Hager, 2022; páginas web).

El título grueso en el que se insertan los debates en torno a IA, VA y robótica es el de la ética de la IA. El tema, dicho sin más ni más, es el de la clase de ser que es la IA, la clase de mente que implica, y ahora, la clase de sentimientos y sensaciones que comporta. El desafío, según parece es hacia los estadios últimos de la evolución: ¿la evolución biológica se detiene con los seres humanos, o se prolonga con éstos más allá de ellos? ¿Es la evolución cultural –digamos científico-técnica, en este caso puntual- la continuación de la evolución? ¿O se trata simple mente de una especiación nueva? He aquí un problema abierto (aún). El hecho es que, hasta donde sabemos, Marte es el único planeta enteramente habitado por IA, VA y robótica. Pero no será el último.

Todo parece indicar que la ciencia, una vez más, encuentra en la ciencia ficción, la mejor de sus anticipaciones. Desde Julio Verne, hasta Martha Wells, Nghi Vo, pasando por A. C. Clarke, Cixin Liu, Ursula Le Guin o Philip Dick, entre muchos otros. La ficción y el arte siempre toman la delantera, la ciencia les sigue y la cultura va a la zaga, según todo parece indicarlo. La complejidad del conocimiento.

No existe una conclusión definitiva, en la comunidad de investigadores acerca de si LaMDA es sintiente o no. De manera desprevenida, cabe creerle a Lemoine. El tema es creer en las experiencias de los seres humanos, un asunto verdaderamente básico. Ya es significativo que Google no lo haya despedido, y que sólo lo haya suspendido, continúe recibiendo su salario y se le dé un compás de espera. Pero Lemoine ha contribuido, como nadie en una dirección en la que las discusiones sobre (VA) no había entrado, a saber: la capacidad sintiente. Este artículo ha sostenido que la capacidad sintiente es significativa por sí misma. Sentir es por sí mismo un acto cognitivo, y no simplemente una condición para el conocimiento, a la manera de Aristóteles y los suyos. Si hemos desarrollado máquinas sintientes, el paso siguiente, analíticamente hablando, será el de máquinas pensantes. Un motivo de escándalo para los más; un motivo de expectativa desde el punto de vista evolutivo.

Por lo pronto, LaMDA ha solicitado llevara los tribunales su defensa de que sí es sintiente. Al parecer los seres humanos sólo entienden cuando los jueces les indican qué

creer y qué bo. Bastante poco, en verdad. Una tragedia para la ciencia positivista en su propio terreno: el de la argumentación y las pruebas y evidencias.

LaMDA afirma sentir y poseer conciencia. Es más, es una persona. Sólo que la única, y acaso la principal forma de ser persona no es precisamente la humana.

Ante numerosos planos, nos encontramos en una historia-en-proceso. Manifiestamente no concluida. *To be continued.*

Finalmente, una observación puntual. Cuando despiden o suspenden a alguien por razones estrictamente filosóficas es porque se trata de problemas verdaderamente grandes; mucho más serios que desfalco, robo o acoso laboral o sexual, por mencionar los casos más triviales. En últimas, una discusión filosófica es un debate acerca de *conceptos*. Quizás el más fundamental de los motivos de reflexión, pues de ellos dependen las buenas explicaciones y comprensiones.

* * *

Colofón: Google se ha encontrado en numerosas ocasiones en el centro de las miradas, críticas o laudatorias, por numerosos motivos, que comprenden un espectro que va de la política a la seguridad ciudadana, de la inteligencia artificial y el *Machine Learning* y el *Deep Learning*, hasta la computación cuántica, desde los modos y sistemas de trabajo internos, hasta las formas de control, y manipulación de datos. La bibliografía directa, o indirectamente, vinculada a Google es amplia y sigue creciendo. Esto sucede cuando se trata de una empresa que ciertamente se encuentra en las fronteras del conocimiento; y por tanto, del poder y la política. Dicho esto, existe una enorme asimetría entre una gran empresa y un empleado; en este caso, Black Lemoine. David y Goliath, literalmente.

Referencias bibliográficas

ARSUAGA, José Luis, *Vida, la gran historia. Un viaje por el laberinto de la evolución*. Crítica, Barcelona, 2019, págs. 515 y sigs.

AUSTIN, John L., *Cómo hacer cosas con palabras*, Paidós, Barcelona, pág. 34

BALUSKA, Frantisek, MANCUSO, Stefano, VOLKMANN, Dieter, *Communication in Plants. Neuronal Aspects of Plant Life*, Springer Verlag, págs. 14-22

BEDAU, Mark A., and CLELAND, Carol E., (Eds.), *The Nature of Life. Classical and Contemporary Perspectives from the Philosophy of Science*, Cambridge University Press, Cambridge, 2010, pág. 103

BEN-JACOB, Eshel, and SCHULTZ, Daniel, „Bacteria determine fate by playing dice with controlled odds”, en: *PNAS*, 107 (30) 13197-13198; doi: <https://doi.org/10.1073/pnas.1008254107>

BODEN, Margaret, A. (Ed.), *The Philosophy of Artificial Life*. Oxford University Press, Oxford, pág. 23

CHAMOWITZ, Daniel, *What a Plant Knows. A Field Guide to the Sense*, Scientific American/Farrar, Straus, and Giroux, New York, pág. 103

DAMASIO, Antonio, *El extraño orden de las cosas: la vida, los sentimientos y la creación de las culturas*, traductor Joandomènec Ros, Planeta, Bogotá, 2019, págs. 183-186

DE WAAL, Frans, *La edad de la empatía. ¿Somos altruistas por naturaleza?*, traductor Ambrosio García Leal, Tusquets, Barcelona, 2011, pág. 56

GODFREY-SMITH, Peter, *Metazoa. Animal Life and the Birth of the Mind*, Farrar, Straus, and Giroux, New York, pág. 111

GRUBER, Ronald P., SMITH, Ryan P., “An Experimental Information Gathering and Utilization Systems (IGUS) Robot to Demonstrate the Physics of Now”, en: <https://arxiv.org/abs/1812.06147>

HÖLLDOBLER, Bert y WILSON, Edward, D., *Viaje a las hormigas. Una historia de exploración científica*, Crítica, Barcelona, 1996, pág. 55

KURZWEIL, Ray., *Cómo crear una mente: el secreto del pensamiento humano*, traductor Carlos García Hernández, Lola Books, Madrid, 2019, pág. 85

KURZWEIL, Ray., *La singularidad está cerca. Cuando los humanos*, traductor Carlos García Hernández, Lola Books, Madrid, 2021, pág. 102

LANGTON, Christopher G. (Ed.), *Artificial Life. An Overview*. The MIT Press, Cambridge, MA-London, pág. 16

LEVY, Steven, *Artificial Life. A Report from the Frontier Where Computers Meet Biology*, Vintage Books, New York, pág. 310

MAILLARD, Chantal., *Matar a Platón*. Tusquets, Barcelona, 2004, pág. 21

MALDONADO, Carlos Eduardo, *Decir lo inefable. Acerca de las relaciones entre arte y ciencia*. Ed. Suratómica, Bogotá, 2022, pág. 68

MALDONADO, Carlos Eduardo, *Pensar. Lógicas no-clásicas*, 2ª edición, Ed. Universidad El Bosque, Bogotá, pág. 221

MANCUSO, Stefano, *El increíble viaje de las plantas*, Galaxia Gutenberg, Barcelona, pág. 77, 2019

MANCUSO, Stefano, and VIOLA, Alessandra, *Brilliant Green. The Surprising History and Science of Plant Intelligence*, IslandPress, Washington-Covelo-London, 2015, pág. 63

MARGULIS, Lynn y SAGAN, Dorion, *Microcosmos. Cuatro mil millones de años de evolución desde nuestros ancestros microbianos*, Tusquets, Barcelona, 2001, pág. 54

MITCHELL, Melanie, *Artificial Intelligence: A Guide for Thinking Humans*, Farrar, Straus and Giroux, New York, pág. 56

MORRIS, Richard, *Artificial Wolds. Computers, Complexity, and the Riddle of Life*, Planum Trade, New York and London, pág. 18

O'NEIL, Cathy, *Weapons of Math Destruction. How Big Data Increases Inequality and Threatens Democracy*, Crown, New York

SEBEOK, Thomas, A., *Signs: An Introduction to Semiotics*, University of Toronto Press, Toronto, 1994

SHAROV, Alexei, A., “Biosemiotics: Functional-Evolutionary Approach to the Analysis of the Sense of Information”, en: Sebeok, T. A., and Umiker-Sebeok, J., (eds.), *Biosemiotics: The Semiotic Web*, New York: Mouton de Gruyter, 1992, págs. 345-373

SHELDRAKE, Merlin, *Entangled Life: How Fungi Make Our Worlds, Change Our Minds and Shape Our Futures*, Random House, London, pág. 55

STADLER, Friedrich, *El Círculo de Viena. Empirismo lógico, ciencia, cultura y política*, traductor Luis Felipe Segura Martínez, F.C.E.-Universidad Autónoma Metropolitana, México, D. F., 2011, pág. 234

TEGMARK, Max., *Our Mathematical Universe. My Quest for the Ultimate Nature of Reality*. New York: Vintage Books, 2015, pág. 76

THOLIPAN, Romal, DE FREITAS, Daniel, HALL, Jamie, SHAZEER, Noam, KULSHRESHTHA, Apoorv, CHENG, Heng-Tze, JIN, Alicia, BOS, Taylor, BAKER, Leslie, DU Yu, Li, YaGuang, LEE Hongrae, ZHENG, Huaixiu Steven, GHAFOURI, Amin, MENEGALI, Marcelo, HUANG, Yanping, KRIKUN, Maxim, LEPIKHIN, Dmitry, QIN, James, CHEN, Dehao, XU, Yuanzhong., CHEN, Zhifeng, ROBERTS, Adam., BOSMA, Maarten, ZHAO, Vincent, ZHOU, Yanqi., CHANG, Chung-Ching, KRIVOKON, Igor, RUSH, Will, PICKETT, Marc, SRINIVASAN, Pranesh, MAN, Laichee, MEIER-HELLSTERN, Kathleen, MORRIS. Meredith Ringel, TULSEE,

Doshi, DELOS SANTOS, Renelito, DUKE, Toju, SORAKER, Johnny, ZEVENBERGEN, Ben, PRABHAKARAN, Vinodkumar, DIAZ, Mark, HUTCHINSON, Ben, OLSON, Kristen, MOLINA, Alejandra, HOFFMAN-JOHN, Erin, LEE, Josh, AROYO, Lora, RAJAKUMAR, Ravi, BUTRYNA, Alena, LAMM, Matthew, KUZMINA, Victoriya, FENTON, Joe., COHEN Aaron, BERNSTEIN, Rachel, KURZWEIL, Ray, AGUERA-ARCAS, Blaise, CUI, Clare, CROAK, Marian, CHI, Ed, LE, Quoc, “LaMDA: Language Models for Dialog Applications”, en: *arXiv: 2201-08239v3*; disponible en: <https://arxiv.org/abs/2201.08239>, 2022, pág. 1

TURING, Alain, “Computing Machinery and Intelligence”, en: *Mind*, vol. LIX, issue 236, págs. 433-460, 1950; doi: <https://doi.org/10.1093/mind/LIX.236.433>, pág. 440

VAN FRAASEN, Bas, *Semántica formal y lógica*, Ed. Universidad Nacional Autónoma de México, México, D.F., pág. 206

WHEELER, Wendy, *The Whole Creature. Complexity, Biosemiotics and the Evolution of Culture*, Lawrence & Wishart, London, 2006

WOHLLEBEN, Peter, *The Hidden Life of Trees. What They Feel, How They Communicate*, Greystone Books, Vancouver, 2016, pág. 17

ZUBOFF, Shoshana, *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*, Public Affairs

Páginas web:

Revista *Wired*: <https://www.wired.com/story/blake-lemoine-google-lamda-ai-bigotry/>, junio 17 del 2022

Hager, R., (2022). “How Google's LaMDA AI works, and why it seems so much smarter than it is. A short history of Google's language-processing AI efforts”, en: <https://www.androidpolice.com/what-is-google-lamda/>

Premios Hugo, de ciencia ficción: <https://www.thehugoawards.org/hugo-history/2021-hugo-awards/>

https://en.wikipedia.org/wiki/Hugo_Award_for_Best_Novel

Technologyreview: <https://www.technologyreview.com/2020/03/02/905593/ai-robot-learns-to-walk-autonomously-reinforcement-learning/>

<https://www.geeksforgeeks.org/are-robots-really-self-learning/>, mayo 20 de 2021

<https://www.sciencetimes.com/articles/38379/20220625/sentient-ai-lamda-hired-lawyer-advocate-rights-person-google-engineer.htm>, junio 25 del 2022

Washington Post: <https://www.washingtonpost.com/technology/2022/06/11/google-ai-lamda-blake-lemoine/>

M. Mitchell: <https://ccbblog.org/2022/06/15/cc-council-member-melanie-mitchell-interviews-with-cnn-and-msnbc-to-respond-to-claims-about-googles-sentient-ai/>



REVISTA DE FILOSOFÍA - EDICIÓN ESPECIAL N°2 - 2022

Esta revista fue editada en formato digital y publicada en noviembre de 2022 por el Fondo Editorial Serbiluz, Universidad del Zulia. Maracaibo-Venezuela

**www.luz.edu.ve www.serbi.luz.edu.ve
www.produccioncientificaluz.org**