

Del fantasma en la máquina al fantasma de Freud: la inteligencia de lo artificial y sus fantasías

*Sergio J. Aguilar Alcalá**

Resumen

Tomando como ejemplo clave la figura conocida como “LOAB” entre los aficionados a la inteligencia artificial, y a partir de las lecciones sobre el orden simbólico y el sujeto del significante planteadas por Jacques Lacan en su texto sobre un cuento de Edgar Allan Poe, este artículo propone reflexionar sobre algunas discursividades implantadas en el ámbito popular respecto de la inteligencia artificial. Se trata de discursos que, lejos de proveer una mera explicación del funcionamiento de estas máquinas, muestran la manera en la que han ideologizado nuestra comprensión de lo humano, el cuerpo, el lenguaje y el pensamiento.

Palabras clave: inteligencia artificial, psicoanálisis, imágenes, neurociencias.

Abstract

Taking as a key example the figure known as “LOAB” among fanatics to IA, and from the lessons on the Symbolic order and the subject of the signifier brought by Jacques Lacan in his text on a story by Edgar Allan Poe, this article proposes a reflection upon some discourses popularly implemented on regards of artificial intelligence. These discourses, far from

* Investigador independiente. Correo electrónico: [sergio.aguilaralcala@gmail.com] / ORCID: [<https://orcid.org/0000-0002-1712-753X>].

being simply an explanation on the ways these machines work, show the ideology they impose on our conceptions of the human, the body, language and thought.

Keywords: artificial intelligence, psychoanalysis, images, neurosciences

Introducción: el fantasma en la máquina

En septiembre de 2022, la artista sueca Supercomposite¹ hizo un hilo en Twitter² en el que describe su experiencia al encontrarse con una misteriosa figura femenina mientras utilizaba un generador de imágenes basado en inteligencia artificial (IA). La artista estaba experimentando con *negative prompts* (indicadores negativos), es decir, instrucciones que se le dan al generador de imágenes para que cree una figura “*que luzca lo más diferente posible*” al indicador puesto; esto funciona colocando el *prompt* y agregando “–1” (de ahí la expresión *negative prompt*).

Así, Supercomposite introdujo como indicador “Brando::–1”, con la curiosidad de ver cuál sería el “*opuesto*” del actor Marlon Brando. El resultado fue la siguiente imagen:

Figura 1



¹ Su sitio de internet: [<https://supercompo.site>].

² El hilo de Twitter y todas las imágenes presentadas aquí se encuentran en: [<https://web.archive.org/web/20220907140608/https://twitter.com/supercomposite/status/1567162288087470081>].

Sorprendida por el resultado, la artista decidió solicitar la *imagen negativa* de ésta, esperando obtener una imagen del actor Marlon Brando, por lo que escribió “*skyline DIGITA PNTICS::−1*”. Es fácil encontrar aquí varios problemas, pues lo que se hizo, a lo sumo, es una especie de representación verbal de algunos elementos que aparecen en la imagen, que ciertamente luce más o menos como una silueta de una ciudad (*skyline*), e incluye letras que *parecen* “*DIGITA PNTICS*”. Más adelante profundizaré en lo problemático de esto.

¿Qué fue lo que el generador arrojó ante el “*skyline DIGITA PNTICS::−1*”? Cuatro imágenes que mostraban lo que parecía ser la misma mujer, en entornos y gestos tenebrosos (figura 2). Extrañas e incómodas imágenes. Una de ellas destaca (figura 3), pues tiene lo que *parecen ser* *letras* en la parte superior. Supercomposite decidió nombrarla “Loab” por esa razón.

Figura 2



Figura 3



No obstante, ¿las cosas son así de tenebrosas? ¿Es LOAB un “fantasma en la máquina”, como lo denominaron algunos comentaristas?³

³ Véase, por ejemplo: Frank The Tank (2023), “Unlocking the Mysteries of Loab: The Ghost in the Machine”, *Toolify.ai*, 15 de noviembre, [<https://www.toolify.ai/ai-news/unlocking-the-mysteries-of-loab-the-ghost-in-the-machine-40599>]. Este texto resulta particularmente revelador del trabajo de adoctrinamiento que la ideología pretende implantar para aceptar y celebrar la IA, pues afirma que la figura de LOAB “fue diseñada para retar el pensamiento convencional y empujar los límites de la creatividad en el arte IA”.

Es probable que ese “fantasma” más bien se encuentre en nuestro acercamiento a la realidad, no “*en la máquina*”. Esta facilidad por la que nos decantamos a leer letras ahí donde aparecen figuras refleja nuestra naturaleza de seres atrapados por el lenguaje, o –como lo concibió Jacques Lacan, quizá el más influyente psicoanalista después del propio Sigmund Freud– por el orden simbólico. Será un escrito de Lacan el que permita iniciar una discusión sobre los *fantasmas* de la IA y los que parecen evocarnos al hablar sobre la IA.

Para explicarlo en pocas palabras, en el psicoanálisis de Freud, el concepto de *fantasía* suponía la escenificación del modo en que el sujeto establece una relación con la realidad; en el psicoanálisis de Lacan, este concepto se mutó por el de *fantasma*, que implica el modo en que el sujeto condensa su relación con la otredad. Son conceptos que siguen discutiéndose en el psicoanálisis y plantearían una digresión muy amplia en este trabajo, por lo que no voy a detenerme en ellos.⁴ Más bien, me interesa proponer en este texto la idea de fantasía/fantasma en varias acepciones: la posibilidad de que la IA tenga un fantasma, una especie de entidad autoconsciente que se expresa en LOAB (*el fantasma en la máquina*); el modo en que maquinamos discursos sobre lo que es la IA (o sea, el modo en que explicamos lo que asumimos que la IA es o hace: *nuestro fantasma frente a la IA*), y el modo en que, en una discursividad que cree haberse extirpado un parásito incómodo (como el modo en que las neurociencias creen ya haber “superado” al psicoanálisis), este parásito reaparece de manera sutil, velada y como desmentida (*el fantasma de Freud que recorre las neurociencias*).

Por todo ello, este texto no pretende hacer una digresión, explicación o discusión sobre cómo las IA hacen lo que hacen, sino sobre cómo nos relacionamos y hablamos acerca de lo que hacen. Para ello, voy a detenerme en algunos casos de nuestra relación con imágenes y textos generados por IA y por el “pensamiento” de las máquinas:

⁴ Quien esté interesado en profundizar, en el caso de la *fantasía*, recomiendo la lectura del texto de Freud de 1924, “La pérdida de la realidad en la neurosis y psicosis”; y, en el caso del *fantasma* en Lacan, la sesión VII (dictada en 1957) del *Seminario 4* (1994).

desde el ajedrez hasta los generadores de texto, el videoarte, las películas y las aproximaciones neurocientíficas a los sueños. Estos ejemplos podrán parecer heterogéneos, pero tienen el interés de mostrar la diversidad de puntos de encuentro, contradicción y de discusión de nuestras fantasías/fantasmas sobre la IA.

Como veremos, estas fantasías sobre las que construimos nuestra aproximación a los resultados de la IA introducen una serie de problemas respecto a nuestra relación con el lenguaje, las imágenes, la comunicación entre humanos, la comunicación entre máquinas y con nuestro propio cuerpo. Antes de intentar discutir con la programación, las matemáticas computacionales, las neurociencias o el cognitivismo, deberíamos de plantearnos una pregunta mucho más simple: ¿qué idea de lo humano, del lenguaje y del cuerpo se introduce en los discursos sobre la IA?

La idiotez de la inteligencia

Hoy se nos habla acerca de lo que la IA “interpreta”, “entiende” y “aprende”, sin que quede claro qué tipo de interpretación, qué tipo de entendimiento o qué tipo de aprendizaje es el que efectúa una máquina. De hecho, ni siquiera queda claro qué tipo de inteligencia posee la *inteligencia artificial*. Como ha señalado con precisión Slavoj Žižek (2023), al llamarla “idiotez artificial”, el punto no es que la máquina “hable” como nosotros, sino que nosotros mismos, con las constricciones y obsesivas reglas sociales que nos imponemos y la lógica cognitivista, biologicista y unidimensional que tenemos del lenguaje, estamos empezando a hablar como una máquina, “obsesivamente diciendo sólo lo que uno cree que quiere decir”. Antes de que la IA hable como humano, nosotros hablaremos como máquinas.

El proyecto de hacer corresponder las palabras al pensamiento tiene una serie de implicaciones acordes con el mundo tecnocientífico que se asombra de cómo una “inteligencia” artificial “aprende”. Pero ésta no es una discusión novedosa, pues podemos rastrear en la cibernetica de mediados del siglo pasado ese problema formulado en

—casi— los mismos términos. Un texto de Lacan (2009) se dedicó a este tema: *El Seminario sobre “La carta robada”* (en su versión de 1966), que permite integrar a la discusión la idea de la *máquina de pensamiento*, en la que se anticipa lo que medio siglo después llamamos IA.

La carta robada, la letra deslizada

El Seminario sobre “La carta robada” lleva ese nombre por el cuento de Edgar Allan Poe, “La carta robada”.⁵ Todo el enredo del cuento sobre quién roba qué a quién y quién se entera o no de ello resulta muy importante, porque el tema de esta historia —Lacan así lo interpreta— es la rotación de posiciones en una especie de *cadena de significantes*. Así como los personajes se desplazan conforme una carta pasa de unas manos a otras, Lacan entiende la naturaleza del sujeto del lenguaje, que se desplaza en el movimiento de los significantes que lo (des)anclan. Esta cadena, como señala Lacan al inicio su texto, da cuenta de que “es el orden simbólico el que es, para el sujeto, constituyente”, pues el sujeto es determinado fundamentalmente por el “recorrido de un significante” (2009: 24).

Los personajes son sujetos marcados por el significante que está en constante desplazamiento; en términos lacanianos, son *sujetos del significante*, y el matema que Lacan inventó para esta idea es \$, que se lee como *S barrada, sujeto barrado o sujeto del significante*. La noción de “sujeto del significante” podría despistar, creyendo que son sujetos *determinados* en un lugar por un significante, pero recordemos que el punto del significante en Lacan no es su fijeza, sino su *desplazamiento*: el sujeto del significante está determinado por el *recorrido* del significante, por su *deslizamiento*. Esto es lo paradójico del orden simbólico lacaniano: es un *orden* caracterizado por un continuo *deslizamiento*.

⁵ Por cuestión de espacio, no incluiré una sinopsis del cuento. Remito a su lectura o, en su defecto, a un resumen, con el fin de comprender a cabalidad los puntos que se comentarán a continuación, los cuales están escritos asumiendo que se conocen los detalles del cuento.

Esto permite a Lacan cuestionar la idea de una comunicación transparente entre subjetividades. En el cuento, cuando el Prefecto le explica al detective Dupin lo acontecido antes de su llegada, Lacan (2009: 30) señala que se trata de un diálogo entre “un sordo con uno que oye”, e inmediatamente agrega lo siguiente:

Es decir que [este diálogo] representa la complejidad verdadera de lo que se simplifica ordinariamente, con los más confusos resultados, en la noción de comunicación.

Se percibe en efecto con este ejemplo cómo la comunicación puede dar la impresión, en la que la teoría se detiene demasiado a menudo, de no implicar en su transmisión sino un solo sentido, como si el comentario lleno de significación con que lo hace concordar el que escucha pudiese, por quedar inadvertido para aquel que no escucha, considerarse como neutralizado (Lacan, 2009: 30).

¿Qué es una comunicación que implica en su transmisión un solo sentido? Sabemos que el sueño de la cibernetica era conseguir precisamente eso para la comunicación entre máquinas; sin embargo, lo que Lacan señala es cómo esto resulta imposible en la comunicación humana, en la que nos topamos una y otra vez con malos entendidos y filtros que impiden la transmisión de “un solo sentido”, incluso a pesar de las intenciones de quienes hablan. Por ejemplo, recordemos que el cuento viene atravesado por la imposibilidad de acceder a una narración omnisciente de los acontecimientos: si los ordenamos diegéticamente, encontramos momentos definidos, pero la narración del cuento sólo incluye la presentación de algunos de ellos; los demás son narrados en pasado, son los recuerdos de un personaje. Incluso, en el primer diálogo, lo que obtenemos es un triple filtro: los acontecimientos sobre el robo fueron narrados por (el personaje que se interpreta como) la Reina al Prefecto de la policía; luego por éste a Dupin y al narrador y, finalmente, por el narrador a nosotros.

Lacan considera que el hecho mismo de que el mensaje sea transmitido *a través* de estas voces –como en el juego infantil conocido

como “teléfono descompuesto” – se redobla en el hecho, nada casual, de que *nunca* nos enteramos lo que dice la carta, es decir, el mensaje está velado en su enunciado, no así en los *efectos* que produce en la enunciación.

Por su falta de fijeza, el significante se encuentra siempre excedido y siempre marcando una ausencia: es la consecuencia lacaniana de la diferencialidad del significante saussuriano (es decir, que un significante es lo que *no* es otro significante). La policía del cuento no se percata de esto: la razón por la que son sordos mientras Dupin sí escucha. Son sordos porque no escuchan lo que están diciendo, no se percantan, al decir que no encuentran la carta que está escondida, que su búsqueda misma está condicionando sus resultados, que su búsqueda transforma el objeto que están buscando (Lacan, 2009: 36). No se percantan de que no sólo no encuentran la carta, sino que *sí encuentran la ausencia misma*, de modo que incluso cuando una carta no está, no simplemente no está, sino que *sí está presente a través de su ausencia*: la ausencia de la carta tiene efectos en las relaciones entre sujetos y en la relación de los sujetos con la carta misma.

La letra deslizada, sujeto del significante

En el cuento de Poe, la carta es un elemento que, antes de “decirnos algo” (pues nunca nos enteramos de su contenido), ejerce *efectos de desplazamiento*, de trastocamiento de la estructura en la que se relacionan los personajes: cada uno de ellos está marcado por cómo la carta va desplazándose de mano en mano, y por cómo los otros personajes son, a su vez, cambiados por este deslizamiento de la carta. Esa *carta/letra*⁶ que se desplaza es lo que marca, como significante en deslizamiento, a los personajes/sujetos de la estructura del cuento: en términos lacanianos, el sujeto humano está, en tanto sujeto, siempre ya marcado en un lugar en la red simbólica de su relación con los

⁶ Recuérdese que “carta” es *letter* en inglés (de Poe), que es la misma palabra para “letra”, y este juego de palabras también presenta el vocablo francés *lettre* (de Lacan).

otros sujetos, es decir, es un sujeto no por lo que *es*, sino por el *lugar* que ocupa en la red de intersubjetividad.

Para Lacan, la estructura de relaciones que se establece entre los sujetos no es fija, sino que siempre es desplaza más allá de los intereses, voluntad o conciencia de los propios sujetos: “cualquiera que sea el destino escogido para la Reina, sigue siendo cierto que esa carta es el símbolo de un pacto, y que incluso si su destinataria no asume ese pacto, la existencia de la carta la sitúa en una cadena simbólica extraña a la que constituye su fe” (Lacan, 2009: 39). Por ello, una carta siempre llega a su destino: *la carta/letral/significante surte efectos*, sea su mensaje leído o no, y es el desplazamiento de esta carta en el cuento, del mismo modo que el significante en la vida humana, lo que define al propio sujeto humano como atrapado “en el desfiladero de lo simbólico” (Lacan, 2009: 40).

Este desfiladero, producido por el deslizamiento de la carta entre los personajes, de la letra entre los sujetos, no sólo supone que ahora un sujeto diferente tiene la carta, sino una alteración total de la estructura de las relaciones que se sostienen, y, por lo tanto, del sujeto que es resultado de esta estructura. En otras palabras, si para Lacan el sujeto es el resultado de la estructura, no debemos olvidar que esta estructura no se fundamenta en la fijeza de sus elementos, sino en el desplazamiento inherente, de modo que *el sujeto es el resultado del deslizamiento del significante en la estructura*.

Sujeto del significante, máquina del pensamiento

¿Cómo entender esa idea de que el sujeto es lo que se desliza en la estructura? Es aquí donde Lacan introduce *la máquina del pensamiento*.

En el cuento de Poe, Dupin menciona a un niño que jugaba con otro a adivinar si las pelotitas en su mano son un número par o impar. El resultado, entonces, sólo puede ser par o impar, uno u otro, 0 o 1, + o -. Sin embargo, como Lacan aclara, aquí entra en juego un complicado sistema de intersubjetividad, donde uno de los juga-

dores piensa lo que el otro está pensando que está pensando, ya sea que intente adivinar (el jugador activo) o que oculte las pelotitas (el jugador pasivo):

sé que si mi adversario es un simple, su astucia no irá más allá de cambiar de tablero para su apuesta, pero que sí es un grado más fino, se le ocurrirá que esto es precisamente lo que voy a cavilar y por lo tanto conviene que juegue sobre el mismo. [...]

[si yo oculto las pelotitas] mi esfuerzo en cada instante será sugerir al adversario la existencia de una ley que preside cierta regularidad de mis jugadas, para arrebatarle su captura las más veces posibles por medio de su ruptura (Lacan, 2009: 66-67).

Lo que encontramos es que no podemos reducir simplemente el asunto del azar en los juegos a un mero azar numérico. No porque no haya un azar en los números, sino porque el azar es algo en lo que la subjetividad se construye para intentar dominarlo, explicarlo, someterlo. La tecnociencia trabaja con mucho ímpetu para crear máquinas que puedan controlar este azar, hasta el punto de anularlo: “no es impensable que una moderna máquina de calcular, desentrañando la frase que modula sin que él lo sepa y a largo término las elecciones de un sujeto, llegue a ganar más allá de toda proporción acostumbrada en el juego de par e impar” (Lacan, 2009: 67).

Lacan imagina una máquina que pueda hallar cierto patrón en el sujeto que esconde las pelotitas para que quien debe adivinar pueda hacerlo con certeza matemática. Esta máquina intenta anular la decisión del que esconde las pelotitas (al haber encontrado el patrón que rige su decisión). Pero lo que no sabemos es si la máquina puede pensar que quien oculta las pelotitas sabe que el otro puede pensar en una ley que gobierna cómo lo intenta engañar, pero que podría pensar que él piensa que piensa eso para así cambiar ese patrón.

Esto, que parece trabalenguas, es bastante simple cuando lo aterrizamos en un ejemplo práctico reciente: sucedió en un evento en vivo frente a millones de personas que veían la final del Mundial de Fútbol de Catar en 2022, entre Francia y Argentina. El jugador fran-

cés Kylian Mbappé tiró tres penales durante el mismo partido contra el portero argentino Damián Martínez, y lo que los comentaristas señalaban era el gran dilema para ambos: por parte del tirador, si acaso debía repetir o cambiar el lado al que tiró el balón en el penal anterior; y, por parte del arquero, si acaso creería que el tirador cambiaría el lado, o si creería que él creería eso, de modo que, sabiendo lo que el arquero cree que el jugador cree, entonces sabría que no cambiaría el lado para tirar, pero sabiendo que él otro sabe que él sabe que el otro sabe que él sabe... Por supuesto, lo que encontramos aquí es una estructura paranoica que en algún punto es interrumpida por un *acto* subjetivo que debe ejecutarse: tomar la decisión de lanzar el balón a un lado o a otro.

Cuando Lacan señala que si imaginamos una máquina que podría empezar a calcular lo que su oponente piensa que la máquina piensa, entonces podemos imaginar una máquina que incluso prevea y calcule el próximo error, lapsus o equívoco que alguien cometería. Eso es un sueño directamente cibernetico: es el sueño de crear una máquina que pueda hallar un patrón en aquello que parece azaroso. Es el sueño de los neurocientíficos que quieren grabar los sueños, crear una réplica de un cerebro, y de sus apóstoles tecnocientíficos que nos dicen que los algoritmos “nos conocen mejor de lo que nos conocemos nosotros mismos” (véase, por ejemplo, el texto de Bilinkis, 2019).

Lo curioso es que no hay nada más antipsicoanalítico –o, por lo menos, que demuestre un total desconocimiento de la obra freudiana– que creer que los equívocos o los lapsus son azarosos. Esas equivocaciones, para Freud, están sobredeterminadas: no son en modo alguno producto del azar, sino que son singulares, impactan de modos singulares y por causas singulares a cada sujeto.

Máquina del pensamiento, máquina que “piensa”

¿Qué estamos pensando cuando decimos que una máquina “piensa”? Verifiquemos uno de los lugares en los que –años antes de la popula-

ridad de la inteligencia artificial que vivimos hoy— ya se discutía qué es “una máquina que piensa”: el ajedrez. Desde mediados del siglo xx se diseñaron experimentos de máquinas que jugaran ajedrez, y en 1996 se dio el famoso caso en que la máquina Deep Blue le ganó al campeón del mundo, Garri Kaspárov, siguiendo los lineamientos de competencias internacionales para ambos oponentes.

En 1950, un ajedrecista aficionado, Claude E. Shannon, más conocido por ser una de las figuras fundamentales de la cibernetica y del campo de la comunicación, publicó un breve artículo donde planteaba cuáles son los problemas para el diseño de una máquina de ajedrez y los posibles caminos para hacerla posible. Ese pequeño texto nos ayuda a entender cómo los conceptos que distintos campos de pensamiento utilizan no pueden darse por sentado, y no podemos apostar, tan rápido, a la “interdisciplina” o el “diálogo” para pensar las consecuencias de la IA. Cuando la cibernetica hablaba de “pensamiento”, cuando la psicología habla de “sujeto”, cuando la ingeniería robótica habla de “máquinas que piensan”, ¿realmente sabemos a qué se están refiriendo?

Shannon empieza su texto señalando que la pregunta sobre si las máquinas de computación pueden hacer algo que caractericemos como “pensar” aún no se responde. Sin embargo, “[t]he basic design of these machines is so general and flexible, however, that they can be adapted to work symbolically with elements representing words, propositions or other conceptual entities”⁷ (Shannon, 1950: 48).

Considerar cómo programar una máquina de ajedrez –indica Shannon– es algo excelente para entender las aplicaciones prácticas de la computación, y podemos delinear las tres tareas principales que dicha máquina supondría (Shannon, 1950: 49): primero, se elige un “código” para indicar todas las posiciones y las piezas; segundo, se debe hallar una “estrategia” que elija los movimientos a realizar; y tercero, esta estrategia debe ser traducida en una secuencia de órde-

⁷ “El diseño básico de estas máquinas es tan general y flexible, sin embargo, que pueden ser adaptadas a funcionar simbólicamente con elementos que representan palabras, proposiciones u otras entidades conceptuales” (traducción propia).

nes que la computadora ejecuta (lo que se conoce como programa de la computadora).

Como notamos, la segunda tarea es la más delicada de todo el proceso, pues es la caja negra que acontece entre el *input* y el *output*. Resulta claro que se trata del problema más complejo de las tres tareas: la elección de un código para posiciones y piezas lleva siglos estandarizada en el modo en que se anotan las partidas de ajedrez, y la traducción de órdenes precisas sobre qué pieza específica mover, en qué momento específico y de qué manera específica, en código que la máquina lea y execute; es algo sencillo una vez que se tiene un lenguaje de programación y de ingeniería. El dilema, obviamente, es cómo elegir la pieza, el momento y la manera de mover. Dice Shannon que:

A straight forward process must be found for calculating a reasonably good move for any given chess position. This is the most difficult part of the problem. The program designer can employ here the principles of correct play that have been evolved by expert chess players (Shannon, 1950: 49).⁸

Entonces, la máquina podría seguir los principios de un “juego correcto”. Esos principios son los primeros que aprendemos al iniciar en el juego de ajedrez: desarrollar piezas, ocupar el centro, enrocarse pronto, etcétera. Pero en este punto nos enfrentamos a los problemas de la pregunta con la que Shannon abrió su artículo, la cual escuchamos hasta el hartazgo como el discursillo de moda en la época: ¿puede diseñarse una máquina que sea capaz de “pensar”? El problema con la pregunta es que parece presuponer cierta obviedad: como si fuera obvio qué es el “pensar”. ¿Se refiere a “pensar” de acuerdo con la manera en que los neurocientíficos entienden el pensamiento? ¿O la manera en la que “pensamiento” se encuentra en el

⁸ “Un proceso directo debe ser hallado para calcular un movimiento razonablemente bueno en cualquier posición dada. Ésta es la parte más difícil del problema. El diseñador del programa puede emplear aquí los principios de un juego correcto que han desarrollado jugadores expertos de ajedrez” (traducción propia).

cognitivismo o en la psicología Gestalt? ¿Qué tal la manera de pensar en las religiones orientales? Cuando decimos que la máquina “piensa”, ¿qué estamos pensando sobre el pensamiento?

Shannon (1950: 50) reconoce que no tiene sentido que la máquina “piense” *todos* los movimientos *posibles* para responder ante un movimiento estándar como, por ejemplo, 1.e4, que es la apertura de peón de rey, quizás el movimiento de apertura más común de todos. Un movimiento de respuesta de las negras como peón a b5 sería uno que cualquier persona consideraría como “muy malo”, y por ello no sería una línea de juego que la máquina consideraría. Esto resulta muy claro, pero no supone, en modo alguno, la complejidad del ajedrez, que no se reduce a hallar “el mejor movimiento” simplemente, sino a hallar el mejor movimiento dada la situación específica del tablero y basándose en la experiencia, el estilo y las condiciones del jugador que la ejecute. Si en ajedrez hubiera *siempre* un movimiento que es “el mejor”, ya no tendría sentido jugar ninguna partida, pero toda vez que hay alrededor de 10^{95} de posibles combinaciones de juegos, y que variables como la presión geopolítica o dificultades personales y el uso del reloj influyen en el comportamiento de los jugadores, es difícil señalar que haya un “mejor movimiento” siempre. Entonces, el punto ahora es: ¿cómo “pensará” la máquina? ¿Jugará de manera más agresiva o más defensiva, más posicional o más estratégica? ¿Al modo de juego de Capablanca, de Polgar, de Karpov, de Kaspárov o de un jugador de menor nivel?

Para evidenciar este problema, podríamos llevarlo a una siguiente dimensión, considerando a un jugador que juega *contra su propio bot*. En la plataforma Chess.com, existen *bots* de jugadores famosos a los que se les han cargado sus partidas registradas y que, por ello, hacen los movimientos que, estadísticamente hablando, representan lo común de esos jugadores: sus tipos de aperturas, sus tipos de sacrificios, sus tipos de posiciones, etcétera. En un video donde juega contra su propio *bot*, el GM Viswanathan Anand ofrece comentarios interesantes que reflejan lo clave para distinguir el modo en que se establece el “pensamiento” de un humano jugando ajedrez frente al de una máquina (señalo en cursivas lo que me parece clave para la

argumentación): “es divertido usar la tercera persona y decir “veamos qué es lo que él hace”, ¿pero éste *soy yo?*”, “no estoy seguro de por qué está él haciendo eso”; “odio cuando me hago esto a *mí mismo*, ¿es así como se dice?”, “esto es molesto, que te ganes *tú mismo*”.⁹

Shannon (1950: 51) reconoce hacia el final del texto que la clave de este problema no es una cuestión ingenieril o cómo hacer que la máquina “aprenda” de sus errores, sino algo más fundamental: dependiendo de lo que entendamos por “pensar” podremos responder si es posible diseñar una máquina que “piense” al jugar ajedrez. Y esto es, efectivamente, lo que parece que no estamos pensando lo suficiente.

Entonces, ¿qué es lo que la máquina no puede pensar? Cualquiera que haya jugado con un *bot* de ajedrez al que pueda ganarle sabe algo que el *bot*, la máquina, *nunca* hace: rendirse ante una situación de desventaja apabullante. Cuando se tiene una ventaja de puntos y material muy grande sobre el *bot*, éste sigue haciendo movimientos que apelan a cierto intento de revertir el resultado. Cuando uno (como humano) va perdiendo, quizá apueste por conseguir un empate de rey ahogado, o jugar ganando segundos para que al oponente se le acabe el tiempo primero, pero *uno se sabe* claramente derrotado. De hecho, si uno pierde una ventaja importante muy pronto o hacia el final de la partida (como perder a la reina en los primeros movimientos o cuando, al perderla, la ventaja del oponente se vuelve demasiado amplia por el resto de sus piezas), usualmente se rinde, e incluso, en algunos círculos profesionales, se considera una grosería que el jugador en desventaja no se rinda, pues sabe –precisamente por principios del juego– que no se debe jugar esperando que el oponente se equivoque. Y parece que la máquina no tiene muy clara esta situación: la máquina no puede jugar pensando el *no jugar*.

Después de lo anterior, regresemos a lo señalado por Lacan (2009: 67): ¿puede una máquina de cálculo ganar siempre en un juego donde debe tomar en cuenta el modo en que el otro está pen-

⁹ Véase: Chess.com (2021), “Viswanathan Anand Plays His Own Bot!”, [video]. YouTube, 9 de enero, [<https://www.youtube.com/watch?v=EEWuZ6suFt4>].

sando? ¿Sueña Lacan con una máquina que mida, descifre y prediga el inconsciente? En absoluto. Continúa en el texto:

Pura paradoja sin duda, pero en la que se expresa que no es por falta de una virtud que sería la de la conciencia humana por lo que nos negamos a calificar de máquina-de-pensar aquella a la que concederíamos tan miríficas actuaciones, sino simplemente porque no pensaría más de lo que lo hace el hombre en su estatuto común sin que por ello sea menos presa de los llamados del significante (Lacan, 2009: 67).

En otras palabras: no es que la máquina “no piense” simplemente porque no tiene “conciencia”. Más bien, “no piensa” de un modo similar a cómo los sujetos, que somos “presa de los llamados del significante”, no pensamos tampoco, sino que *algo piensa* en nosotros. El pensamiento se piensa a sí mismo y evidencia esa escisión constitutiva del Yo. Tal vez logremos que una máquina piense como un humano, pero no podemos diseñar una máquina que *no piense* como *no piensa* un humano: ahí en su equívoco, en su imposibilidad.

Lacan señala cómo las máquinas pueden resultar muy ominosas: no es que puedan pensar *como si* tuvieran una conciencia propia, sino que, más bien, nos demuestran cómo en nosotros *hay algo que piensa*. Hay algo en lo humano que está pensando más allá de nuestra voluntad y conciencia, y todos los intentos de dar cuenta de qué es eso que parece ir desbocado en la máquina, todos los intentos con los que tratamos de explicar cómo es que la máquina tiene sus propios deseos, conciencia y voluntad, más bien vuelven a apuntar a esa falla en nuestra propia subjetividad que, causada por el significante, imposibilita a un sujeto completo. En nuestra fantasía del pensamiento de una máquina, evidenciamos la fantasía de nuestro propio pensamiento (o la pesadilla del pensar de la máquina, como veremos a continuación).

Máquina que “piensa”, máquina que “interpreta”

Lo que observamos es que la relación que sostenemos con las máquinas, con su “pensamiento” y su “inteligencia”, es una relación mediada por el campo del lenguaje, que imposibilita ya toda transparencia. Es esencial recordar que vivimos en una época donde escuchamos con frecuencia que las máquinas, y su “pensamiento” e “inteligencia”, lograrán “cambiarlo todo”, el cual es un mantra común en la cultura popular. Sin duda, el asunto que más “lo cambiará todo” es la generación de imágenes, códigos, textos, videos y audios con inteligencia artificial, como lo atestigua la enorme cantidad de libros publicados sobre cómo usar la IA más allá de la computación; por ejemplo, en el marketing, la industria de moda, el emprendedurismo de *apps* y la publicación de libros generados con IA.

Recordemos el caso de LOAB mencionado al principio de este texto. Tras la primera imagen, cuando la artista Supercomposite solicitó “la imagen negativa” de esa especie de boceto de un logotipo, esperando que le saliera el actor Marlon Brando, nos encontramos varios problemas de interesante complejidad. ¿Acaso es lo mismo pensar en “una figura lo más distinta posible de otra”, tal como podemos pensar en el negativo de un número? Equiparar palabras con números, ¿no demuestra una aproximación bastante ingenua a la naturaleza de la lengua? El “negativo” de negro, ¿es blanco u *orgen*?¹⁰

Por otro lado, “skyline DIGITA PNTICS” es, a lo sumo, una especie de representación verbal de ciertos elementos (y sólo ciertas dimensiones de ellos) que aparecen en la imagen, que si bien luce *más o menos* como la silueta de una ciudad (*skyline*), y que incluye letras que *más o menos* indican “DIGITA PNTICS”, también está muy incompleta: no se mencionan colores, la forma cuadrangular, que el “edificio” más a la izquierda parece la silueta de Batman o haber decidido que eso después de lo que parece una C es una S y no un 2 invertido. Si

¹⁰ No señalo esto en contra de la artista, pues nunca queda claro en sus tuits si ella comparte o no esta ingenua aproximación a las palabras con las que funcionan estos motores, aunque tampoco parece tomar distancia de ella.

no fuéramos lacanianos, creeríamos que esto simplemente supone que las palabras “no alcanzan” a aprehender la totalidad de la imagen, pero no podemos ignorar el hecho de que las palabras permiten más bien abrir ciertas dimensiones de la imagen que no existen antes de ellas. Es justo lo que sucede con LOAB, la cual fue nombrada –como ya se mencionó– por las supuestas letras que aparecían en una de sus primeras imágenes.

Resulta muy peculiar la facilidad con la que leemos letras en esa imagen, y no pensamos simplemente que son elementos visuales de la propia imagen, es decir, que leamos *letras* donde bien podrían sólo ser *formas*: ¿por qué eso sería una *L* y no una línea vertical junto a una horizontal? ¿Por qué eso sería una *O* y no el dibujo de un anillo blanco? ¿Es una *A* o un triángulo cuya base se dibujó muy arriba? ¿Es una *B* o dos cuadrados, uno encima del otro? En otras palabras: ¿qué hace que un palito y un anillito dejen de ser eso y se vuelvan una *b*, una *d*, una *p* o una *q*?

Eso es lo propio de la condición humana, como sujeto atrapado en el orden simbólico, un sujeto “presa de los llamados del significante” (Lacan, 2009: 67): que *leamos ahí una palabra*, y no simplemente *veamos figuras geométricas*. Resulta ser una cualidad del sujeto del significante encontrar salvajemente palabras y letras ahí donde podrían figurar simples imágenes. Éste es el punto lacaniano de que el orden imaginario se subordina al orden simbólico.

En un documental para YouTube, donde se expone este caso y otro que mencionaré a continuación,¹¹ destaca el comentario del usuario Sitromxe: “The input wasn’t ‘Marlon Brando’. It was just ‘Brando’. Which the AI seems to have interpreted as a combination of ‘Brand’ and ‘logo’. Which is, in my opinion, abstractly what the initial output looks like: A mock-up for logo design representing a brand”.¹²

¹¹ Nexo (2022), “The Disturbing Art of A.I.”, [video], YouTube, 1 de octubre, [<https://www.youtube.com/watch?v=i9InAbpM7mU&t=1023s>].

¹² “El *input* no era ‘Marlon Brando’. Sólo ‘Brando’. Lo que la IA parece haber interpretado como una combinación de ‘marca’ y ‘logo’. Lo que, en mi opinión, abstractamente es lo que el *output* [la primera imagen] parece: un boceto del diseño de un logo de una marca” (traducción propia).

Como bien dice el usuario, el *input* no fue “Marlon Brando”, sino “Brando”. Y también es cierto que en “Brando”, atendiendo a la equivocidad de la lengua inglesa, se lee una especie de combinación de *brand* (marca) y *logo* (logotipo). Sin embargo, ¿supone esto que la IA “interpreta” la combinación de palabras? ¿Por qué asumimos que la máquina “interpreta” y no simplemente responde a los estímulos verbales que introducimos como indicadores? Los programadores suelen justificar la idea de que las IA sí interpretan y no simplemente responden como en un circuito de estímulo-respuesta, pero vuelvo a pre-guntar: ¿qué tipo de “interpretación” es la que ahí actúa? ¿Hablamos de una interpretación hermenéutica, semiótica, crítica, psicoanalítica, entre tantas otras escuelas y corrientes? Antes de preguntarnos si la IA “interpretó” o no, ¿qué es lo que entendemos por “interpretar”?

Resulta curioso cómo seguimos atribuyéndole a estas máquinas cierta voluntad, cosa que nunca decimos de nuestra comida cuando se echa a perder (“interpretó que debía empezar a descomponerse”) o de nuestros autos (“interpretó que le faltaba gasolina”). De hecho, pocas veces lo decimos de nuestras propias computadoras (no he es-cuchado a nadie decir que su computadora “interpretó que se le gastó la batería” o que “interpretó que no hay conexión wifi”). ¿Por qué entonces parecemos aceptar esto de máquinas de IA? Quizá porque, precisamente, *creemos* que son IA; es decir, esa supuesta “inteligencia” parece ser algo que damos a estas máquinas no sólo a través de nuestra programación de las mismas, sino a través de cómo enten-demos lo que éstas hacen y producen, sostenidas en un semblante de lo imaginario: que hablen en primera persona, que pidan disculpas o den palabras de ánimo. Cuando hablamos de la “inteligencia” de las máquinas, ¿a qué tipo de inteligencia nos estamos refiriendo? ¿Qué es lo que entendemos como propio del pensamiento humano cuan-do asumimos que una máquina es “inteligente”?

En su texto sobre la IA, Žižek (2023) se plantea la pregunta de por qué los textos generados con esos motores piden disculpas antici-padas por su incapacidad para dar más datos, o si ante una peti-ción explícitamente racista o sexista responderían desde ese mismo lugar. En un blog sobre este tema, por curiosidad experimental, Sašo

Dolenc (2023) le cargó a Bing Chat, un generador de textos, que comentara ese análisis de Žižek: la máquina señaló que, pese a ser una pieza interesante y provocadora, no es “balanceada”. Aquí nos enfrentamos al problema: ¿en verdad creemos que la máquina sólo genera un texto sin una posición política? ¿“Balanceada”?

La imposibilidad, aparente, de dejar de pensar sus resultados como resultados de una voluntad propia es lo que crea las condiciones para que, efectivamente, estas máquinas sean “inteligentes”: la tautología se instaura, pues son inteligentes porque vemos sus actos como muestras de inteligencia. Por ello, la idea de Lacan en *El Seminario de “La carta robada”* se mantiene: quizás piensen como pensamos los humanos, pero no vemos pruebas de que no piensen como no pensamos los humanos. Lo que la IA no parece poder hacer es lidiar con la negatividad.

Máquina que “interpreta”, máquina que “*me habla*”

Una especie de histerización es la que caracteriza nuestra relación con las máquinas: cuando éstas sueltan un texto en respuesta a nuestra pregunta, decimos que “*nos responden*”. Aquí un ejemplo donde la ingenuidad ridícula se reúne con un peligroso llamado al “bienestar de la máquina”:¹³ un “canal de noticias” en una página de Facebook, que según su información se encuentra en México, mostraba el video de un robot que, tras 20 horas de trabajo de demostración en una convención, “se desplomó sorpresivamente”. En lugar de verificar ahí un fallo en la máquina como cuando un auto se queda sin gasolina o la batería de una computadora se “infla”, de pronto –nos dice el post– se han generado “debates sobre los límites del trabajo automatizado y el bienestar de las máquinas”. Uno sabe que el mundo se está yendo al abismo cuando –por lo menos la persona que escribió ese texto– demuestra una ingenuidad respecto a las palabras que usa y una ignorancia respecto a los peligros del trabajo auto-

¹³ Véase: [<https://fb.watch/rFk25Gr8ZG/>].

matizado que hace un llamamiento para que haya regulaciones que “protejan a los robots”.

Así, en los modos en los que nos referimos a las máquinas encontramos modos en que intentamos histerizarlas, y también nos histerizamos a través de ellas: es la diferencia entre decir que “la IA dice” y “la IA *me* dice”. Nuestra inclusión en lo que *se* dice es el punto nodal del sujeto presa de los llamados del significante.

En el sitio web *AI told me* (“La IA me dijo”) aparece una entrada con el título *I will not forget* (“No olvidaré”).¹⁴ Ahí se presenta un poema sobre la lenta desaparición de la memoria, y un videoarte donde las “neuronas” de una IA van “apagándose”, una por una, “olvidando”. Esta elección de palabras es de lo que debemos percatarnos, ¿qué dice de nuestro entendimiento de la tecnología cuando creemos que las máquinas “olvidan”, y comparamos nuestra memoria y los modos en lo que somos reinventados por ella (por ejemplo, reprimiendo recuerdos, inventándonos otros) con un simple “apagar” neuronas? La imagen es la de una mujer, pero no cualquier mujer, sino una mujer que está generada a partir de miles de imágenes, una especie de rostro “promedio” generado por la combinación de miles de fotos. Este procedimiento se conoce desde el siglo XIX, pero eso no es lo interesante, sino que los programadores –como nos lo dicen en el poema– fueron “apagando” una por una las “neuronas” con las que el programa generó esta imagen, generando esa extraña distorsión justo antes de una pantalla negra.

Todavía un mayor caso de delirio en la relación con la máquina es lo que dio a conocer el *Daily Mail* el 1 de diciembre de 2022: Michelle Huang había creado un *chatbot* de IA al que le cargó sus propios diarios infantiles, de modo que “habló consigo misma del pasado”, como nos lo relata el encabezado de la nota.¹⁵ Indica la programadora a través de Twitter: “Overall, this was a very trippy but also strangely affirming / healing experience that I didn’t realise that

¹⁴ Véase: [<https://aitold.me/portfolio/i-will-not-forget/>].

¹⁵ Véase la nota aquí: [<https://www.dailymail.co.uk/sciencetech/article-11491323/Woman-talks-past-self-trippy-conversation.html>].

I had access to using real data from my past self-allowed me to connect with her in deeper + more tangible ways than I typically have” (Huang, en *Daily Mail*, 1 de diciembre de 2022).¹⁶

¿Qué clase de subjetividad es la que puede “sanarse y afirmarse” a través de un ensimismamiento como éste, que trata sus propias memorias y secretos de la infancia como datos para alimentar la máquina? Por supuesto, sólo la subjetividad que se construye con el Yo, que cree que el lenguaje es una especie de puente que me puede conectar con el otro (en este caso, *conmigo mismo* en el pasado). Resulta curioso que describa esta experiencia como una conexión “más profunda y tangible”, cuando, desde otra perspectiva, eso es lo último que sucede: hasta en la redacción misma de este tuit notamos su desdoblamiento y división como sujeto del significante, como cuando habla de “*su yo del pasado*” (*my past self*) que le permitió “conectarse con *ella*” (*connect with her*), poniéndose en primera, segunda y tercera persona en pocas líneas. Esta especie de brinco de sujeto gramatical y de pronombres no corresponde a una decisión arbitraria o falta de habilidades de escritura: responde a la naturaleza humana como sujeto del lenguaje, atrapado en una imposibilidad de coincidir su enunciado con su enunciación, pese a que los discursos fantasiosos y celebratorios de la IA intenten desmentir esta imposibilidad.

Finalmente, estas aproximaciones ingenuas a la IA y sus productos decantan en aproximaciones ingenuas al propio cuerpo, al que se le empieza a tratar como una máquina. Esta consideración se encuentra en el corazón de las aproximaciones neurocientíficas a los sueños.

Máquina que me “habla”, máquina que “sueña”

Se conoce como sueño lúcido aquel en el que el soñador se percata que está soñando. Ese “percibirse” es algo que los neurocientíficos

¹⁶ “En resumen, fue una experiencia muy alucinante pero también extrañamente afirmadora/sanadora de la que no me percaté que tenía acceso usando datos reales de mi yo del pasado que me permitieron conectar con ella en modos más profundos y tangibles que los que típicamente tengo” (traducción propia).

han intentado medir: Voss *et al.* (2013) desarrollaron una metodología para ello, a partir de un cuestionario que les aplicaban a los participantes antes y después de la experimentación. Este cuestionario, conocido como la métrica LuCiD, consta de 28 preguntas hechas a los participantes, que se responden del 0 (muy en desacuerdo) al 5 (muy de acuerdo), y que se acompaña de una tabla donde esas preguntas se dividen en 7 variables: *insight*, control, pensamiento, realismo, memoria, disociación, emociones negativas y emociones positivas.

Es difícil no sentir cierta curiosidad ante un trabajo que ha sido citado más de cien veces, según ScienceDirect,¹⁷ y que presenta una metodología tan pobre y tautológica. Se supone que es una metodología para medir “lo cualitativo y cuantitativo de los determinantes” de la autoconciencia del sueño (Voss *et al.*, 2013: 10); y, sin embargo, una de sus siete variables ya engloba exactamente eso, la de *insight* –un vocablo intraducible del inglés–, cuyas preguntas de medición son lo que permiten concluir cualquier nivel de “autoconciencia”: la primera de ellas es si, mientras soñaba, el sujeto estaba consciente del hecho de que las cosas que experimentaba en el sueño no eran reales.

Varios otros problemas metodológicos identificamos quienes, no siendo neurocientíficos, por lo menos tenemos una introspección en nuestros propios sueños. ¿Cómo se mide una variable como “control” en un sueño, y por qué el hecho de que en un sueño yo haga acciones sobrenaturales, como volar, significa que tengo mayor “control” sobre él? El acceso a la memoria de la vigilia durante el sueño, ¿aplica cuando veo en sueños algo que sí recuerdo de la vigilia, pero que no luce así en el sueño? ¿Qué ocurre cuando, mientras sueño, recuerdo otro sueño? Las últimas dos variables, emociones (*emotions*) negativas y positivas, son planteadas con preguntas sobre los sentimientos (*feelings*), ¿acaso son sinónimos?

¹⁷ En el siguiente enlace puede verse el número de citaciones del artículo: [<https://www.sciencedirect.com/science/article/abs/pii/S1053810012002103?via%3Dihub>].

La misma escala se utilizó en otro experimento al año siguiente (quizá en buena medida porque dos personas participaron en ambos): se trata del de Voss *et al.* (2014). Este texto es interesante porque presenta tres relatos de sueños. Uno de los problemas es que el artículo se publica en inglés, pero se entiende que sus sujetos de experimentos hablaron alemán. ¿Qué tipo de consideración hubo en esta traducción? A continuación, los tres relatos:

[1] I was dreaming about lemon cake. It looked translucent, but then again, it didn't. It was a bit like in an animated movie, like the Simpsons. And then I started falling and the scenery changed and I was talking to Matthias Schweighöfer (a German actor) and 2 foreign exchange students. And I was wondering about the actor and they told me "yes, you met him before," so then I realized "oops, you are dreaming". I mean, while I was dreaming! So strange!

[2] I am driving in my car, for a long time. Then I arrive at this place where I haven't been before. And there are a lot of people there. I think maybe I know some of them but they are all in a bad mood so I go to a separate room, all by myself.

[3] It was about shopping. I bought these shoes and then there was such a girl, she went-like-“snap” (snaps her fingers) and cut off her waist, just like that. Interviewer: she cut off her waist? Subject: yeah, just like that (Voss *et al.*, 2014: 813).¹⁸

¹⁸ [1] Estaba soñando sobre pastel de limón. Se veía translúcido, pero luego, no era así. Era como en una película de animación, como en *Los Simpson*. Y luego empecé a caer y el escenario cambió y estaba hablando con Matthias Schweighöfer (un actor alemán) y dos estudiantes extranjeros. Y preguntaba sobre el actor y me dijeron “sí, lo has visto antes”, y entonces me percaté “ups, estás soñando”. Es decir, ¡mientras yo estaba soñando! ¡Qué extraño!

[2] Estoy manejando mi auto, por mucho tiempo. Luego, llego a un lugar donde no había estado antes. Y hay mucha gente ahí. Pienso que quizás los conozco, pero están todos de mal humor, así que me voy a una habitación separada, yo solo.

[3] Era sobre compras. Compré unos zapatos y ahí estaba una chica, ella de pronto “snap” (chasquea los dedos), y se cortó la cadera, sólo así. Entrevistador: ¿se cortó la cadera? Sujeto: sí, sólo así. (Traducción propia).

Se presentan estos tres relatos de los que no se da un solo comentario, más que indicar que el primero es *evidencia* de un sueño lúcido, los otros dos no. Llama la atención que en el primero, el que dicen que sí lo es, el sujeto lo reporta en segunda persona (“ups, *tú* estás soñando”), como si pese a ser un “sueño lúcido”, de acuerdo con esa metodología, aun así fuera necesario atravesar la otredad para dar cuenta de él: me doy cuenta de que estoy soñando, pero hablo en segunda persona. También es curioso que el segundo sueño, aunque no tiene esa dimensión de “darme cuenta de que estoy soñando”, sí demuestra bastante memoria de la vigilia (“mi auto”, “quizá los conozco”). El tercer sueño introduce algo bastante poderoso: un correlato en el cuerpo (la chica cuya cadera se corta) que es expresado a través del propio cuerpo (al chasquear los dedos). Tampoco se entiende para qué incluir estos relatos que no son comentados en absoluto en la argumentación del texto.

El punto al que llegan todas estas metodologías tan extrañas es a las recomendaciones de las “aplicaciones clínicas” de la inducción de sueños lúcidos. La última de ellas es ingenuamente optimista: “Finally, promoting gamma oscillations during REM sleep in post-traumatic stress disorder with reemerging nightmares might trigger lucid dreaming and eventually enable active changes in dream content”¹⁹ (Voss *et al.*, 2014: 812). Es incomprendible de dónde sacan los neurocientíficos que los cambios que produzca el sujeto en su pesadilla recurrente serían positivos para aliviar el trauma, y qué es un “cambio activo” frente a uno “pasivo” también se deja sin explicación. Vale la pena preguntarse cómo entenderían ellos que el sujeto, ante la pesadilla en la que lo persigue el monstruo, considere como “cambio activo” –permitido por el sueño lúcido– el hecho de dejarse atrapar.

El paradigma subyacente a las neurociencias, siguiendo estos experimentos, es una confusión provocada por la reducción de la

¹⁹ “Finalmente, promover oscilaciones gamma durante la fase MRO del dormir en desórdenes de estrés postraumático con pesadillas recurrentes puede desencadenar sueños lúcidos y eventualmente permitir cambios activos en el contenido del sueño” (traducción propia).

complejidad de los procesos psíquicos a la idea de una máquina que responde según una lógica de estímulo-respuesta. Ver al cerebro como una máquina susceptible de ser manipular (con pastillas, drogas, impulsos eléctricos o con la simple verificación de datos al colocar un casco a quienes duermen) es el relato que sirve para desmentir el rol del propio científico que verifica ahí lo que sucede. Es decir, el relato de la objetividad neurocientífica que puede explicarnos cómo funciona la maquinaria cerebral es la desmentida de la inclusión del deseo del científico en ello. La fantasía del científico que objetivamente *desea* estudiar el cerebro y los resultados de ese estudio son dos lados de la misma moneda.

Máquina que “sueña”, el sueño de Freud

Es evidente que el ancestro teórico de las neurociencias, antes que la cibernetica o la computación, es la psiquiatría del siglo XIX, pero es curioso cómo las neurociencias suelen presentarse como las verdaderas herederas del pensamiento científico y médico, y que el psicoanálisis, con quien comparten esos ancestros, no es más que un desvío peligroso de la “verdadera” ciencia del cerebro.

Es así como un muy famoso apóstol de las neurociencias y detractor del psicoanálisis, J. Allan Hobson (incluido entre los autores de los experimentos sobre sueños lúcidos citados como Voss *et al.*), lo plantea al inicio de su libro *Los 13 sueños que Freud nunca tuvo* (2007). El prólogo del libro es un texto escrito *como si su autor fuera el fantasma de Freud*, haciendo una especie de autobiografía intelectual póstuma en la que destaca cómo se apartó de las nociones e hipótesis que estaban en el *Proyecto* (un texto clave en la obra de Freud, que data de 1895), y cómo no estaba al tanto de los descubrimientos neurocientíficos de principios del siglo XX. Ese prólogo termina con ese Freud-títere/Hobson-titiritero diciendo: “Hacia 1950 estaban dadas las condiciones para una revolución que no supe prever y que francamente no me habría gustado. Fallecí en 1939, mucho antes de

que se empezara a derribar el castillo de naipes que había construido con tanto esmero” (Hobson, 2007: 25).

Resulta risible cómo muestra tanto desprecio por el psicoanálisis que no puede evitar ponerse en su lugar para hablar, con el fin de terminar de “derribar el castillo de naipes” que Freud, supuestamente, construyó. Como si el asunto fuera un chiste, en el primero de los capítulos del libro, Hobson relata que, al estudiar los colículos cerebrales, no podía evitar calificarlos como

montecillos particularmente seductores que se acomodan como muslos o nalgas en una colina encima del mesencéfalo. [...] Me parecían tan rechonchos, tan bellamente redondos, y estaban tan bien lubricados por su propio jugo, que me daban ganas de exclamar: “¡Éstos son los colículos! ¿No les parecen maravillosos? (Hobson, 2007: 37-38).

Quizá uno se percate de que no se necesita ser psicoanalista para que resulte, por lo menos extraño, que tras el prólogo tan abiertamente antifreudiano (hablando desde el lugar del propio Freud), pase a dar un ejemplo tan caricaturescamente freudiano de cómo parecía excitarse sexualmente mientras estudiaba partes del cerebro. A esto nos referimos cuando decimos, desde el psicoanálisis, que las neurociencias parecen ser incapaces de escucharse.

Conclusiones: de la sociedad maquínica al fantasma de Freud

En este texto pretendí establecer una lectura de los modos en que cierta fantasía sobre lo que sucede con los productos de la IA acosan las conclusiones que tomamos de su funcionamiento. Una paradoja se establece cuando metemos estas fantasías en los diseños de esas máquinas: una fantasía las diseña, luego verifica sus productos como confirmaciones de sí misma, y con ello reproduce su propio funcionamiento. Es decir, cuando hablamos sobre lo que la IA “hace”, no hacemos sino imponer nuestras fantasías sobre lo que hace, y con

ello imponer ciertas fantasías sobre lo que entendemos por “pensamiento”, “voluntad” y “consciencia”.

El ejemplo de LOAB es particularmente instructivo para verificar cómo la captura del humano en el orden simbólico, en tanto ser de lenguaje, imposibilita tener un acceso “auténtico” a cualquier tipo de máquina. La confusión producida cuando esa máquina es “inteligente” alcanza límites insospechados, que permiten hacernos creer que la máquina puede “olvidar”, que podemos “platicar” con ella. Estos experimentos, como muestra el interés de la comunidad ajedrecística, no son para nada nuevos, pero sí habitamos en una época donde su popularidad parece ser la condición ideal para nublar su condición ideologizante sobre nuestras concepciones de lo humano.

Esta última idea tampoco es nueva: ya la alertaba Norbert Wiener (1961: 27) en su famoso libro de introducción a la cibernetica al anunciar los peligros de pretender tomar la comunicación de máquinas como una manera para comprender la complejidad de la sociedad humana, y ubicaba en esta equivocación que el potencial de la cibernetica y la maquinización no debía ser puesto al servicio del mercado. No obstante, esta lección no parece ser tomada en cuenta por las aplicaciones mercantilistas y bélicas que los neurocientíficos tienen para sus experimentos cerebrales.²⁰ Es tan patente y obvio el asunto que hasta reporteros lo mencionan de casualidad, como si fuera un asunto de chiste o de curiosidad dominical, como el ya antes citado texto de Bilinkis (2019).

LOAB fue bautizada como un “fantasma en la máquina”, como si su presencia fuera inesperada hasta el punto de asumir una voluntad de su parte. Pero si nos percatamos de que el fantasma está en otro lado, no en la máquina sino en el que imagina ver fantasmas en ella, aquí podemos terminar con un giro poético y preguntarnos si Hobson se percatará de cuánta razón tiene al invocar el fantasma de Freud en su libro, y fantasear con que Freud está hablando a través de él, pues nos queda claro, con el delirio del neurocientífico, que

²⁰ El libro de Jan de Vos (2020) está dedicado a pensar precisamente ese problema.

un fantasma recorre la máquina de las neurociencias: el fantasma del psicoanálisis.

Referencias

- Bilinkis, Santiago (2019), “¿Y si tu teléfono te conoce mejor que vos?”, *La Nación*, 17 de febrero, [<https://www.lanacion.com.ar/lifestyle/y-si-tu-telefono-te-conoce-mejor-nid2220616/>].
- De Vos, Jan (2020), *The Digitalisation of (Inter)Subjectivity. A Psi-critique of the Digital Death Drive*, Routledge, Nueva York.
- Dolenc, Sašo (2023), “Artificial Idiocy Responds to Žižek”, *Project Syndicate*, 27 de marzo, [<https://sci-highs.com/artificial-idiocy-responds-to-zizek/>].
- Freud, Sigmund (1984), “La pérdida de realidad en la neurosis y la psicosis”, en *Sigmund Freud. Obras completas, vol. 19* (pp. 189-198), Amorrortu editores, Buenos Aires.
- Hobson, J. Allan (2007), *Los 13 sueños que Freud nunca tuvo*, Fondo de Cultura Económica, México.
- Lacan, Jacques (1994), *El Seminario de Jacques Lacan, Libro 4: La relación de objeto*, Paidós, Buenos Aires.
- Lacan, Jacques (2009 [1966]), “El Seminario sobre ‘La carta robada’”, en *Escritos 1*, Siglo xxi Editores, México.
- Shannon, Claude E. (1950), “A Chess-Playing Machine”, *Scientific American*, vol. 182, núm. 2, pp. 48-51.
- Voss, Ursula, Schermelleh-Engel, Karin, Windt, Jennifer, Frenzel, Clemens y Hobson, Allan, (2013), “Measuring Consciousness in Dreams: The Lucidity and Consciousness in Dreams Scale”, *Consciousness and Cognition*, vol. 22, núm. 1, pp. 8-21.
- Voss, Ursula, Holzmann, Romain, Hobson, Allan, Paulus, Walter, Koppehele-Gossel, Judith, Klimke, Ansgar y Nitsche, Michael A. (2014), “Induction of Self-awareness in Dreams Through Frontal Low Current Stimulation of Gamma Activity”, *Nature Neuroscience*, vol. 17, núm. 6, pp. 810-814.

Wiener, Norbert (1961), *Cybernetics or Control and Communication in the Animal and the Machine* (Second Edition), MIT Press, Cambridge.

Žižek, Slavoj (2023), “Artificial Idiocy”, *Project Syndicate*, 23 de marzo, [<https://www.project-syndicate.org/commentary/ai-chatbots-naive-idiots-no-sense-of-irony-by-slavoj-zizek-2023-03>].

Fecha de recepción: 03/09/24

Fecha de aceptación: 25/04/25

doi: 10.24275/tramas/uamx/202563171-200