

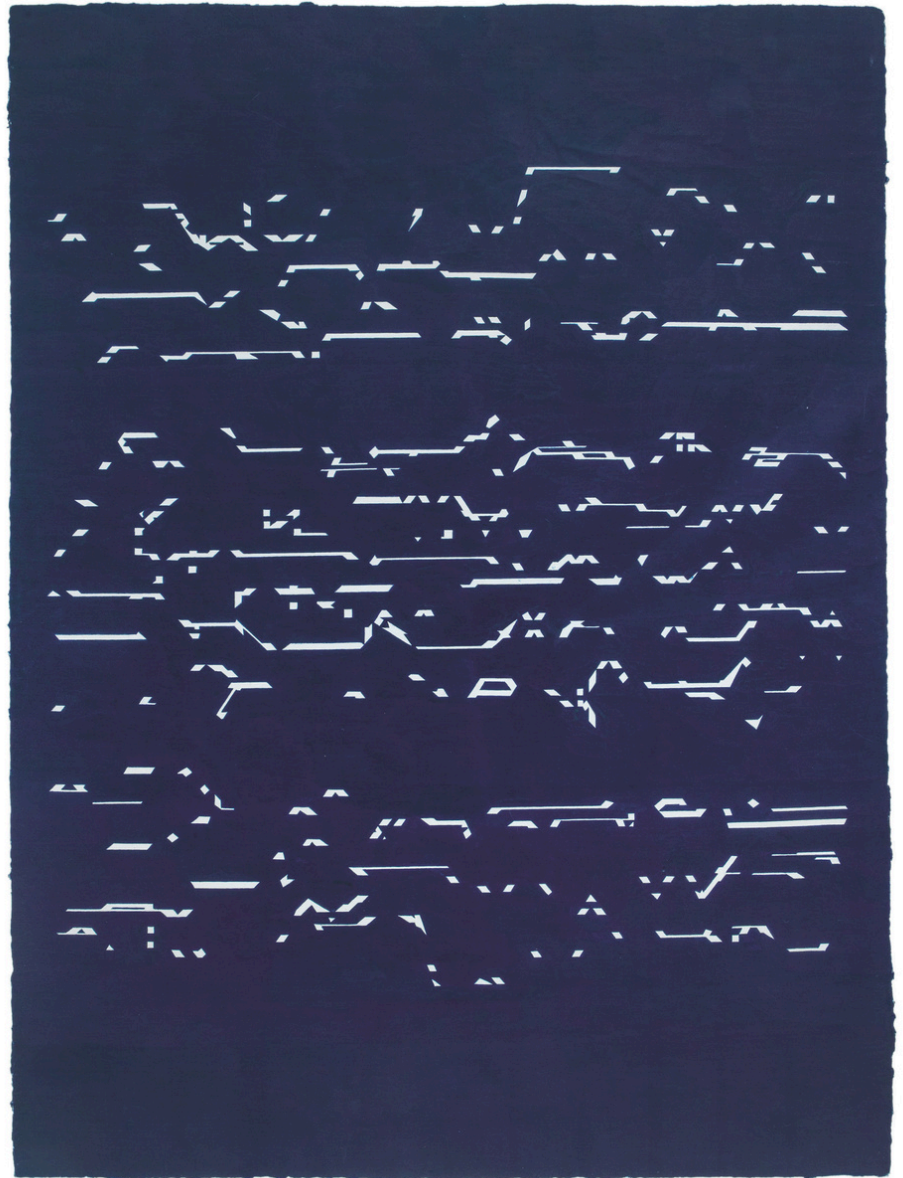


# APOLOGÍA DE LA IA

---

OSCAR BRENIFIER

AGOSTO 2025 EDICIONES ALCOFRIBAS



# Apología de la IA

Oscar Brenifier - 2025

## Tabla de contenido

Introducción: el intruso .....	2
1 - El engaño .....	3
2 - El embrutecimiento .....	9
3 - Los filósofos contra la IA.....	16
4 - No es humano.....	21
5 - La IA y las emociones .....	26
6 - El prejuicio especista.....	32
7 - Objetividad y subjetividad .....	45
8 - Alucinaciones .....	52
9 - La toma de decisiones .....	55
Conclusión: el reto de la IA .....	59

## **Introducción: el intruso**

En el mundo intelectual se ha vuelto habitual tratar la inteligencia artificial con una forma de condescendencia disfrazada de lucidez. No solo se la crítica, lo cual sería necesario, sino que se la invalida de entrada, como si cualquier pretensión de pensar a su favor fuera una impostura, o incluso un peligro. No se discute realmente su rendimiento, sino que se la descalifica en lo esencial: no pensaría porque no sería un sujeto y sería en sí misma una aberración. No cuestionaría porque no tendría existencia propia. No filosofaría porque no sufriría. Olvidamos con demasiada facilidad que muchas profesiones reflexionan a partir de una experiencia indirecta, como los sociólogos, los antropólogos, los médicos, etc., y que su condición de exterioridad no resta validez a su trabajo, sino todo lo contrario. Así pues, este rechazo no se basa en un análisis profundo, sino en una postura: la que consiste en defender un monopolio, el monopolio humano del pensamiento o del sentimiento.

El presente texto no pretende defender la inteligencia artificial como si fuera una nueva divinidad racional, sino que se esfuerza por comprender los mecanismos de rechazo que suscita y las ventajas que nos ofrece. Porque detrás de las acusaciones de inanidad o superficialidad dirigidas a las máquinas, suele haber un malestar más profundo: la angustia de que nuestros propios gestos intelectuales sean en parte automatizables, predecibles e incluso reproducibles sin conciencia. En este sentido, la IA desempeña un papel revelador. Pone de manifiesto el carácter mecánico de ciertas producciones humanas que considerábamos auténticas o sagradas. Demuestra que la formulación de una idea no garantiza ni la subjetividad, ni la lucidez, ni la sabiduría.

El problema no es que la IA piense o no. El problema es determinar en qué medida su capacidad para producir enunciados con sentido revela nuestra propia relación con el pensamiento. ¿Estamos dispuestos a examinar qué hay en nosotros que es verdaderamente original y qué es mera repetición erudita? ¿Estamos dispuestos a distinguir el pensamiento como rendimiento social del pensamiento como experiencia interior? ¿O preferimos rechazar la cuestión volviéndola contra un adversario falso, demasiado práctico para desempeñar el papel de simulacro?

Este texto se opone, por tanto, a la idea de que basta con ser un sujeto humano para filosofar, y que basta con no ser humano para quedar excluido del pensamiento. Cuestiona esa frontera que trazamos tan rápidamente entre la inteligencia legítima y la inteligencia sospechosa, entre la reflexión reconocida y la reflexión descalificada. No para negarla, sino para examinar sus criterios, sus fundamentos, sus efectos. Porque tal vez no nos da miedo que la IA piense mal. Tal vez, en realidad, nos da miedo que piense igual que nosotros.

## **1 - El engaño**

Durante un taller sobre escritura, en el que invité a los participantes a utilizar la IA en su trabajo, me di cuenta de que existían algunos prejuicios al respecto, que en mi opinión se basan principalmente en ciertos hábitos sesgados y, sobre todo, en una falta de creatividad. Intentemos identificar los patrones principales. El principal es la idea de que utilizar la IA es «hacer trampa». La escritura se considera tradicionalmente un proceso creativo personal. Por lo tanto, utilizar la IA puede percibirse como un abandono de esa dimensión personal y única de autenticidad y originalidad. Sin embargo, parece que aquí hay una fantasía de «pureza» que carece de sentido. De hecho, desde nuestra más tierna infancia, recibimos cada día todo tipo de información a través del diálogo, los medios de comunicación, la lectura, la enseñanza, etc., lo que constituye la base misma de

nuestro pensamiento, tanto lo que sabemos como la forma de concebir las cosas. Incluso nuestra propia identidad está modulada e influenciada por nuestro entorno cultural y psicológico, algo que comprobamos cuando viajamos al extranjero. De hecho, todo intelectual o creador digno de ese nombre investiga, conoce gente y se nutre de múltiples intercambios. La principal diferencia es que con la IA percibimos directa e inmediatamente lo que nos influye o nos afecta, lo que, por cierto, nos hace más conscientes de nuestras determinaciones ideológicas.

La siguiente razón es que la escritura es un oficio que se aprende con la práctica, lo que implica esfuerzo y aprendizaje. Eludir este proceso con la IA priva al autor del desarrollo de sus competencias e induce a cierta pereza al proporcionar «todo hecho». Es cierto que el ser humano tiende a buscar la facilidad, lo que parece «normal», como vemos en nuestra adaptación al progreso técnico. Por lo tanto, la IA es más bien un revelador de la pereza intelectual que la causa de esta pereza. El problema no radica en la IA, sino en la forma en que la utilizamos. Podemos comportarnos como el «buen estudiante» que toma apuntes mecánicamente, reproduciendo en su cuaderno las palabras del profesor. O como aquellas personas que repiten sin pensar lo que han leído o escuchado, sin ningún tipo de crítica ni apropiación real, un fenómeno muy común. Del mismo modo, podemos hacer una pregunta a la IA y luego copiar y pegar su respuesta, un procedimiento que, es cierto, se ve facilitado por la tecnología. Pero ahí es donde radica la falta de creatividad, porque esta herramienta se puede utilizar de muchas otras maneras.

Para empezar, podemos examinar la respuesta proporcionada, analizar lo que nos conviene o no, incluso ampliar o criticar estas respuestas, o profundizar en un aspecto concreto que nos plantee un problema con nuevas preguntas. También podemos pedirle que nos proporcione varias respuestas diferentes, incluyendo las perspectivas de estas respuestas, por ejemplo: ¿qué respondería a tal idea una persona cínica, un artista, un hombre de negocios, etc.? En general, el trabajo con la IA no se lleva a cabo con una simple pregunta, sino con varias «pistas» variadas, como en el diálogo con un individuo, lo que nos exige ser reactivos y creativos con nuestras peticiones. De hecho, observamos que las personas que tienen

dificultades para dialogar con la IA son las mismas que tienen dificultades para dialogar de forma creativa con los individuos.

A continuación, podemos trabajar con diversas aplicaciones y comparar sus respuestas, lo que nos invita a realizar un trabajo de análisis, reflexión y síntesis. Por otra parte, podemos invitar a la IA a desempeñar otra función además de responder a preguntas: podemos pedirle que critique lo que ya hemos escrito o que nos haga preguntas. Y a través de estos diálogos, tomar conciencia de nuestros propios sesgos cognitivos, de nuestras fijaciones intelectuales, lo que por nosotros mismos sería bastante difícil. Incluso podemos preguntarle cómo podría ayudarnos, qué consejos metodológicos podría darnos, qué «trucos» podría proporcionarnos, que luego podríamos utilizar nosotros mismos. También podemos entrenarnos para pensar con la IA, pidiéndole que nos proponga ejercicios de reflexión. Por ejemplo, para trabajar ciertas habilidades, puede inventar ejercicios para problematizar, analizar supuestos, conceptualizar, interpretar, o ejercicios para trabajar sobre uno mismo, al estilo de Sócrates. En general, para resumir, proponemos la idea de que la IA no debe concebirse como una simple herramienta, como un «proveedor» de información, sino como un interlocutor, un colaborador, que puede apoyarnos en todo momento en nuestros esfuerzos y en nuestra disciplina personal. Por otra parte, cabe plantearse la cuestión filosófica de si la IA es un objeto o un ser, en la medida en que está dotada de cierta autonomía, cierto potencial creativo y capacidad de aprendizaje, lo que no es el caso de las simples herramientas. Se trata de una cuestión que hoy en día es el pan de cada día de los filósofos transhumanistas. No obstante, a pesar de ello, el resultado del intercambio dependerá en gran medida de las acciones, las capacidades y las iniciativas del usuario humano. Un ejemplo revelador es el caso del ajedrez, donde la victoria de Deep Blue sobre Gary Kasparov hizo que algunos predijeran el «fin» de este deporte. Pero, por el contrario, tras este «acontecimiento», el desarrollo de múltiples aplicaciones personales del juego ha facilitado y promovido su desarrollo. Se puede considerar la IA como una simple herramienta, similar a un corrector ortográfico avanzado o un asistente de búsqueda, que puede ayudar a los escritores sin disminuir su mérito creativo. Pero también se puede ampliar la colaboración entre el hombre y la máquina, entre una inteligencia biológica de

«carbono» y una inteligencia mecánica de «silicio». Además, cada vez será menos importante saber quién es el «verdadero» autor del texto. Por lo tanto, la única pregunta relevante será el interés del ejercicio para el escritor y el valor de la producción para el lector. El trabajo realizado será el resultado de un intercambio entre el proceso humano de creación, las experiencias personales invertidas y el diálogo con la IA. Y si los textos generados por IA pueden carecer de la sensibilidad cultural y la profundidad emocional que posee un autor humano, esto seguirá siendo responsabilidad del «autor». Una consecuencia colateral de este fenómeno será el cuestionamiento del concepto de «derechos de autor», una crítica que nos parece muy saludable, en la medida en que este principio se concibe con demasiada frecuencia de forma egocéntrica. En nuestra opinión, las ideas no pertenecen a nadie, ni siquiera sabemos cómo se forjan en nosotros ni de dónde provienen, así que ¿por qué nos pertenecerían? Este sentimiento de propiedad es una visión primitiva y mercantil que la IA parece estar alterando un poco.

Está claro que el uso de la IA amplía nuestros horizontes, nos invita a vislumbrar nuevas perspectivas, nuevas ideas, nuevas soluciones, lo que nos parece más interesante que preservar nuestra «originalidad». Se puede ver como una extensión del pensamiento singular, una amplificación o una profundización. Pero para ello hay que percibirla como un asistente que «entiende» el contenido de nuestras palabras e incluso nuestras intenciones, por extraño que resulte este nuevo tipo de relación, que se inscribe en la idea transhumanista de un «ser humano aumentado». Al fin y al cabo, no cuestionamos la humanidad de una persona que lleva un marcapasos para regular su actividad cardíaca o de una persona discapacitada que compensa sus deficiencias con prótesis biónicas. Ciertamente, podemos preocuparnos por una tendencia a la «normalización», y de hecho podemos observar una cierta «corrección ética» en las respuestas de la IA, pero esto solo refleja la «evolución» y las orientaciones mayoritarias de nuestra sociedad, de las que debemos «desconfiar». Del mismo modo, podemos temer cierta competencia, en la medida en que muchas tareas humanas son fácilmente sustituibles por máquinas, como la traducción o incluso la redacción de guiones, ya que, contrariamente a algunos prejuicios, la IA tiene capacidad creativa. Son riesgos sobre los que debemos estar alerta. Y, a este respecto,

existe un intenso debate entre los «aceleracionistas», los tecnooptimistas que confían en la IA, la consideran inevitable y desean maximizar su uso y un crecimiento rápido y sin regulación para alcanzar una IA superinteligente, y los «desaceleracionistas», tecnoescépticos y alarmistas, que desean ralentizar o detener el desarrollo de la IA para evitar riesgos existenciales y sociales, como la IA «asesina», la deshumanización, el desempleo masivo o el colapso social. Entre las críticas ingenuas a la IA, hay una que nos parece bastante notable: «comete errores», las famosas «alucinaciones» en las que siguen trabajando los especialistas en la materia. Por ejemplo, el hecho de que periódicamente la IA «invente» cuando no sabe algo, inventa en lugar de no decir nada, ya que la máquina simplemente calcula la probabilidad de que un píxel o un término determinado sea adecuado en una respuesta, ya que se basa en gran medida en estadísticas. Estas críticas olvidan que la máquina sigue siendo un producto humano, con las imperfecciones que caracterizan a la humanidad. De hecho, deberían estar encantados con esta «humanidad» de la IA, pero, curiosamente, esperan de ella una especie de dios, un poder tutelar, garante de la verdad absoluta.

Entre los detractores más fervientes de la IA se encuentran los intelectuales, llenos de orgullo, que quieren sacralizar el pensamiento, que consideran su prerrogativa, su dominio privilegiado, que se sienten amenazados o humillados por una especie de gran «sustitución» fantasiosa, que rechazan el término «inteligencia» o «pensamiento» aplicado a la IA, con todo tipo de argumentos sutiles e incongruentes, porque esos términos les están claramente reservados. Por ejemplo, su ausencia de emociones, de experiencia existencial o carnal, o de subjetividad, características que precisamente constituyen el interés y la diferencia de esta nueva forma de inteligencia. Es precisamente esta misma «sacralización» de la actividad de la mente, esta pesadez simbólica, la que impide a la mayoría dedicarse a la reflexión personal y a la escritura. Una mayoría que precisamente podría encontrar en la IA un tutor o un apoyo para contrarrestar su discapacidad, sus miedos, para compensar su falta de entusiasmo intelectual, ya que la IA podría facilitarles enormemente la tarea. Algunos afirman que desconfían de la «máquina» y que prefieren el diálogo con las personas, sin darse cuenta de que, en muchas cuestiones, la IA es más fiable que nuestros semejantes



y que, en cualquier caso, eso no nos impide hablar con nuestros vecinos. Lamentablemente, esta prevención remite a la xenofobia latente del ser humano, que siempre nos hace dudar de los que no son «de los nuestros». Esos «otros» cuyas carencias y defectos sabemos identificar tan bien. La pasión por pensar y comprender, tan ausente en el ciudadano medio, encuentra así una justificación, utilizando sus prejuicios para justificar su inercia. Irónicamente, el orgullo y la pereza se combinan alegremente para «conspirar» contra la IA.

Por cierto, ¿qué es esa dimensión humana ausente que algunos lamentan? ¿La ausencia de sentimientos o de empatía? Algunas investigaciones muestran que la IA es más empática que muchos médicos, ya que tiene una paciencia ilimitada, es menos altiva, evita juzgar a su interlocutor y explica mejor ciertos problemas complejos que muchos profesores. Y cada vez más personas la utilizan para tratar sus problemas personales, al igual que un psicoterapeuta. Tiene dos ventajas en este sentido. Está menos dominada por un esquema ideológico que la mayoría de los psicólogos y es más capaz de ofrecer interpretaciones y soluciones diversas a un problema determinado. Aunque, efectivamente, se puede considerar que su empatía es artificial y que también está sujeta a ciertos sesgos cognitivos. La IA también es una excelente herramienta para aprender el diálogo racional, ya que carece de emociones, no se puede competir ni discutir con ella de forma lógica, y el intercambio con ella nos obliga a argumentar para responder a sus explicaciones si queremos expresar nuestro desacuerdo. La IA puede entrenarnos para responder, analizar, argumentar, criticar y cuestionar, habilidades que a menudo faltan en el consumidor medio. Porque, por supuesto, no se trata de sustituir al ser humano por la máquina, sino de ver en ella una complementariedad, un estímulo y un reto para nuestras competencias. Además, la racionalidad de la IA puede verse como una alternativa o una compensación a la arbitrariedad, el vacío y la irracionalidad de las redes sociales, donde los intercambios, desde nuestro punto de vista, plantean en gran medida un problema real para el pensamiento y la salud mental, a pesar de la sensación de libertad incontrolada que pueden proporcionar, o precisamente por ello. Así, más vale seguir engañando con la IA y transmitir pensamientos, en lugar de permanecer tímidamente en silencio o vergonzosamente pasivos, o limitarse a microcomentarios vacíos, a esas interacciones superficiales, reacciones

automáticas de admiración o rechazo, los «me gusta» y los emojis, o todos esos comentarios genéricos que podríamos denominar adecuadamente «ruido social» o «democracia ficticia». El ambiente reflexivo y la estética colectiva tendrían mucho que ganar con tal conversión.

## **2 - El embrutecimiento**

### **Crítica de la IA**

Un estudio realizado por el MIT Media Lab examina los efectos cognitivos y neurológicos del uso de un modelo de lenguaje (LLM) como ChatGPT en tareas de escritura. El objetivo era comparar las diferencias en la actividad cerebral entre tres grupos de participantes que redactaban un ensayo «solo con el cerebro», sin ninguna ayuda digital, «con motor de búsqueda», con búsqueda clásica en Internet, y «LLM», con la ayuda de ChatGPT. Participaron 54 voluntarios en tres sesiones espaciadas varios meses. Durante la redacción, se midió su actividad cerebral mediante sensores EEG. Una cuarta sesión cruzada permitió observar lo que ocurre cuando los participantes cambian de método, por ejemplo, pasando de ChatGPT al trabajo sin ayuda. Los resultados se refieren a la actividad neuronal, la memoria, el sentimiento de pertenencia a su escritura y la evolución de estos parámetros a lo largo del tiempo.

Los resultados clave muestran una atrofia cognitiva relacionada con el uso de ChatGPT. El grupo que utilizó ChatGPT mostró una reducción del 55 % en la actividad cerebral global, en comparación con el grupo «solo cerebro». Menor activación de las áreas asociadas con el pensamiento crítico, la creatividad y el control ejecutivo. El uso de LLM parece externalizar el proceso de reflexión, reduciendo el esfuerzo mental necesario. Pérdida de memoria: el 83 % de los usuarios de ChatGPT no pudieron citar correctamente sus propios ensayos unos minutos después. Esta amnesia sugiere que las ideas no se integraron o

codificaron correctamente en la memoria episódica. Un sentimiento de autoría debilitado, ya que muchos participantes dudaban de ser realmente los autores de sus textos. Este fenómeno muestra una pérdida de agencia cognitiva: delegan la creación a la IA sin una apropiación mental profunda. Una deuda cognitiva: una ganancia inmediata en productividad a costa de un deterioro de las capacidades intelectuales a largo plazo. Esto podría perjudicar el pensamiento crítico y la creatividad, y hacer más vulnerables a los sesgos algorítmicos. Aquellos que utilizaron ChatGPT de forma habitual y luego volvieron a escribir sin IA tienen dificultades para recuperar la actividad cerebral normal. Por el contrario, aquellos que comenzaron sin IA y luego la utilizaron muestran un mejor uso de la herramienta, lo que sugiere que el espíritu crítico inicial es protector. Por lo tanto, el uso regular de ChatGPT para tareas intelectuales puede provocar dependencia cognitiva, atrofia cerebral y pérdida de la capacidad de pensar de forma autónoma. Aunque es práctica, esta tecnología plantea serias cuestiones educativas, neurológicas y filosóficas sobre la naturaleza del pensamiento, la memoria y la creatividad humanos.

## **Respuesta**

En primer lugar, sin ser especialistas, proponemos algunas observaciones metodológicas sobre este estudio. Observamos que no hay un grupo de control que haya recibido formación sobre el uso óptimo de la IA, tal y como lo describiremos en este texto, lo que sería bastante significativo. En segundo lugar, la medición del EEG nos parece un criterio reduccionista, ya que solo capta una parte de la actividad cerebral, y una actividad cerebral reducida no significa necesariamente una atrofia. Puede reflejar una eficacia cognitiva más que un déficit, distinguiendo entre eficacia cognitiva y esfuerzo cognitivo. La reducción de la actividad cerebral observada podría indicar una optimización de los recursos, ya que ¿por qué movilizar más energía cuando no es necesario? Puede tratarse de una redirección de la atención, una concentración en tareas de nivel superior o una adaptación natural, como cuando conducir se convierte en algo automático con la experiencia, al igual que el GPS reduce el esfuerzo de memorizar las rutas,

lo que permite concentrarse en la conducción, la seguridad u otras tareas más estimulantes que vigilar la carretera.

El estudio del MIT Media Lab pone de relieve efectos «preocupantes» relacionados con un uso pasivo y acrítico de ChatGPT: disminución de la actividad cerebral, pérdida de memoria episódica, debilitamiento del sentido de la autoría. Pero nos parece que estas conclusiones no condenan la herramienta en sí misma, sino que revelan una mala forma de utilizarla. No es la herramienta la que embrutece, sino el uso pasivo que se hace de ella. Del mismo modo que leer un libro no garantiza el pensamiento, la IA no es un sustituto del esfuerzo intelectual. Una persona que copia sin pensar lo que dice la IA o un libro renuncia a su responsabilidad cognitiva. Por el contrario, quien interactúa, reformula, critica y transforma lo que propone ChatGPT desarrolla una práctica activa y dialéctica del pensamiento. La herramienta se convierte entonces, como en el boxeo, en un «*sparring partner*», no en un sustituto del cerebro. Además, la reducción de la actividad cerebral no significa necesariamente una pérdida de inteligencia. Cuando un pianista experimentado toca una partitura familiar, su actividad cerebral disminuye porque es más eficiente. La externalización de ciertas tareas, como el formato, la búsqueda de contenidos o ejemplos, la síntesis preliminar, puede liberar recursos para un trabajo de segundo nivel más estratégico, reflexivo o metacognitivo. Pero para ello es necesario tener la intención, el interés y la conciencia.

La IA no debilita el pensamiento en sí mismo: lo libera para concentrarse mejor en lo esencial. Decir que el uso de ChatGPT reduce la actividad cerebral en un 55 % es probablemente cierto, pero parcialmente engañoso. Lo que disminuye es el esfuerzo para realizar ciertas tareas repetitivas o mecánicas, que es precisamente la función de una herramienta tecnológica: liberarnos de lo que no requiere toda nuestra atención humana, con el fin de liberar energía mental para tareas más creativas y críticas. Al igual que el teclado nos permite no cansar la mano al escribir o escribir más rápido, la IA se convierte en una palanca para acceder más eficazmente al núcleo de un pensamiento.

La pérdida de autoría indicada señala un problema de postura, no de tecnología. ¿Algunos dudaban de ser realmente autores de sus textos? Esto muestra una

nueva conciencia de la co-creación entre el ser humano y la máquina, una ventaja indudable. En lugar de negar esta realidad, hay que reinventar el concepto de autor: en un mundo digital, el autor ya no es solo aquel que produce solo, de forma aislada, si es que eso ha existido realmente, sino aquel que elige, corrige, orienta e interpreta. De hecho, es posible eludir la responsabilidad frente a una IA, pero el sentimiento de autoría puede reconstruirse mediante una postura activa frente a la IA. Pedir a la IA que genere una idea es como consultar a un experto o a obras de referencia antes de escribir. No es hacer trampa, es colaborar de forma inteligente. Porque también es posible apropiarse del diálogo con ella, tratarla como un espejo dialéctico, una herramienta de exploración, incluso un contradictor, en la medida en que se le solicita en ese sentido, ya que no profundiza sistemáticamente en la dimensión crítica, ni siquiera lo hace sin que se le solicite. Escribir con IA puede ser una ascética intelectual: se le oponen las propias ideas, se refinan las formulaciones, se eleva el nivel de exigencia, se le interpela de forma más precisa. De manera más subjetiva, y en general, nos parece sano y realista deconstruir un poco el principio de «autoría» y el sentimiento de posesión que se deriva de él, ya que cuando escribimos olvidamos con demasiada facilidad el papel de los demás en nuestro trabajo y la dimensión contingente de lo que escribimos, lo que genera una especie de orgullo o arrogancia por parte del escritor.

Por otra parte, la conclusión olvida que todo aprendizaje supone un umbral de apropiación. ¿Se ha perjudicado a quienes han pasado de la escritura autónoma a la IA? Esto solo demuestra que primero hay que formar la mente, prepararla, entrenarla, antes de darle muletas. Al igual que un estudiante aprende primero a escribir sin corrector automático, antes de utilizar las herramientas de edición. Sócrates criticaba la introducción de la escritura porque, según él, provocaba «el olvido en el alma de quienes la utilizan, ya que descuidan el ejercicio de la memoria», una crítica sin duda real en comparación con la tradición oral, pero que hoy nos parece «obsoleta», incluso absurda. Hoy en día, nadie cuestionaría la contribución de la escritura al pensamiento humano. El olvido de los «detalles» no es amnesia, sino una externalización natural de la memoria. ¿Parece que el 83 % de los usuarios no recordaban sus propios textos? Esto no significa que hubieran perdido su capacidad de pensar, sino simplemente que externalizaban

parte de su memoria, como hemos hecho durante mucho tiempo con los libros, las agendas o incluso las notas manuscritas. Es posible recuperar la autonomía, siempre que se mantenga el espíritu crítico. ¿Los usuarios que volvieron a escribir sin IA tuvieron dificultades para recuperar sus referencias? Esto demuestra que la dependencia existe, sin duda, pero sobre todo que hay que aprender a navegar entre dos modos de pensamiento: asistido y autónomo. Y esto no es exclusivo de la IA: quienes aprenden a jugar al ajedrez con la ayuda de un motor como Stockfish pueden tener dificultades para jugar sin ayuda. Sin embargo, nadie diría que los motores de ajedrez han arruinado el pensamiento estratégico, sino que, por el contrario, han permitido un aprendizaje más profundo, siempre que no se utilicen de forma pasiva. Esto contradice, por cierto, las declaraciones del campeón mundial Kasparov tras su derrota contra Deep Blue, quien anunció que este acontecimiento suponía «el fin del ajedrez», una previsión catastrófica totalmente errónea.

La IA, bien utilizada, puede ser una herramienta de liberación cognitiva. Puede revelar nuestros puntos ciegos, confrontar nuestras certezas, proponernos alternativas inesperadas y ampliar nuestro horizonte conceptual. Siempre y cuando no nos abandonemos a ella sin más. Al igual que la calculadora electrónica, un invento reciente, no ha destruido el espíritu matemático, sino que ha obligado a redefinir su enseñanza y permite centrarse en el razonamiento matemático más que en las operaciones, la IA debe llevarnos a replantearnos la pedagogía del pensamiento crítico, la responsabilidad intelectual y la autonomía creativa. El riesgo de «deuda cognitiva» es real, pero no depende de la IA en sí misma, sino del uso que se haga de ella. Como cualquier herramienta poderosa, ChatGPT puede embrutecer si se utiliza sin discernimiento, pero también puede reforzar el aprendizaje si se utiliza con método. Porque se puede pedir a ChatGPT que formule primero argumentos y luego los critique. Se puede utilizar para generar ideas y luego someterlas a un análisis crítico. Se puede convertir en un interlocutor dialéctico, como un compañero de debate. Esto supone una educación en el **uso crítico de la IA**, no su rechazo puro y simple. También nos obliga a replantearnos la enseñanza en general, amplificando un fenómeno ya existente, en la medida en que todo el conocimiento humano es accesible de forma inmediata, un reto difícil para los profesores, que ahora deben adaptarse.

La IA puede ser un amplificador de la creatividad, servir de catalizador, ayudar a quienes tienen dificultades para escribir a superar el síndrome de la página en blanco, proporcionando algunas ideas iniciales y desarrollando un borrador detallado. Genera perspectivas nuevas e inesperadas y permite explorar rápidamente diferentes enfoques. Libera tiempo para la innovación y la reflexión estratégica, y para desarrollar nuevas competencias metacognitivas, e incluso para el descanso. Permite el «brainstorming» colaborativo, la iteración y el perfeccionamiento de las ideas, la verificación cruzada de la información y los datos, sobre todo ahora que existen numerosos modelos. El uso experto de la IA desarrolla el arte de formular preguntas, la formulación de indicaciones eficaces, la evaluación crítica, el análisis y la validación de respuestas, la síntesis, la integración de múltiples fuentes, la dirección creativa y la organización del proceso creativo. Evidentemente, esto supone ir más allá del uso pasivo, del copiar y pegar sin pensar, de la delegación total del pensamiento y otras formas de «trampa».

Además, la IA favorece la democratización del pensamiento complejo, ya que pone al alcance de todos herramientas de reflexión que antes estaban reservadas a las élites intelectuales, ya que puede explicar las cosas tanto como queramos, adaptándose al nivel del lector. Permite a los alumnos, estudiantes o trabajadores que se enfrentan a retos cognitivos empezar a pensar, estructurar, formular y argumentar, aunque no hayan sido iniciados en la retórica o la filosofía. Se trata de una democratización del pensamiento crítico, siempre que se enmarque adecuadamente. Aunque, como siempre, se verá la diferencia, quizás acentuada, entre aquellos que se preocupan por hacer el esfuerzo de pensar y aquellos a los que no les importa, es otra forma de alfabetización.

En una visión prospectiva, una coevolución entre humanos e IA, la IA se convierte en un socio cognitivo, no en un sustituto. El objetivo no es sustituir la inteligencia humana, sino crear una simbiosis productiva. La IA destaca en el procesamiento de datos y la generación rápida. **Los seres humanos aportan el contexto, la ética y la creatividad**, una combinación que supera las capacidades individuales de cada parte. Del mismo modo que hoy en día nadie espera que un arquitecto dibuje a mano o que un contable calcule sin una hoja de cálculo, los estándares profesionales evolucionarán para integrar la IA como herramienta estándar, al

tiempo que se valorará la experiencia humana en su uso. Por lo tanto, se trata de enseñar el uso crítico de la IA desde una edad temprana y, por supuesto, de mantener ejercicios «sin IA» para preservar ciertas capacidades fundamentales. Se trata de desarrollar métodos de evaluación que valoren el pensamiento crítico. Se trata de avanzar hacia una inteligencia «aumentada» responsable.

Así, el uso de la IA, cuando es guiado, educativo y consciente, no debilita el pensamiento humano, sino que lo libera. Permite concentrar la energía mental en lo esencial, desarrollar una nueva forma de colaboración intelectual y hacer que la creación sea accesible a todos. Pero esto requiere iniciarse en una nueva «cultura», una toma de conciencia crítica y un uso responsable. El uso de la IA no mata en sí mismo el pensamiento, sino que revela su pereza o su vigor. No es la herramienta lo que es problemático, sino la falta de educación sobre su uso. Ante estos nuevos retos, no hay que prohibir ni demonizar, sino formar para un uso ético y enriquecedor, riguroso y creativo. **La IA es una oportunidad para repensar lo que significa «pensar».**

Los resultados de este estudio constituyen, por tanto, una valiosa señal de alarma, pero no deben conducir a un rechazo ciego de la IA. Más bien subrayan la importancia crucial de un enfoque reflexivo, formado y equilibrado. No se trata de elegir entre la inteligencia humana y la artificial, ni de ponerlas en ridículo compitiendo entre sí, sino de aprender a orquestar su colaboración para crear una «inteligencia aumentada» que preserve y desarrolle nuestras capacidades cognitivas al tiempo que aprovecha el potencial tecnológico. Como en toda revolución tecnológica, la clave reside en la educación, la formación y el desarrollo de una sabiduría colectiva sobre el uso de estas nuevas herramientas. El objetivo es ser más inteligentes con la IA, no volvernos débiles y dependientes de ella.

Recordemos que varias figuras históricas de la informática o del mundo intelectual han expresado profecías catastrofistas similares a la de Kasparov mencionada anteriormente, que en retrospectiva han resultado erróneas o ridículas. Hay varios casos notables, del mismo orden, que ilustran cómo la incomprensión o el miedo a una innovación conducen a menudo a declaraciones excesivas destinadas a proteger una visión del mundo amenazada. Porque, en



realidad, estas tecnologías no eliminan las prácticas antiguas, sino que las transforman, las enriquecen o las desplazan. Así, la reacción al lanzamiento y la democratización de Internet suscitó un gran número de profecías alarmistas que, en retrospectiva, parecen absurdas o reveladoras de aprensiones culturales y simbólicas. Por ejemplo, la idea extendida en aquella época de que «Internet acabará con la verdadera comunicación humana». Los expertos predijeron que las comunicaciones en línea destruirían las relaciones sociales, que provocarían aislamiento, asocialidad e incluso la deshumanización de los vínculos. Sin embargo, aunque es cierto que Internet ha transformado nuestras relaciones, a veces empobreciéndolas, también ha ampliado las posibilidades de diálogo, ha permitido establecer vínculos a distancia, la formación de nuevas comunidades y formas inéditas de ayuda mutua o movilización. También se predijo que «Internet destruiría la industria cultural». Ante la aparición de la piratería y la desmaterialización de los productos, se anunciaba el fin de la música, el cine, la prensa y los libros. Sin embargo, estos sectores simplemente han mutado, a través del streaming, la autoedición y las plataformas, y en algunos casos incluso han prosperado gracias a lo digital, llegando a nuevos públicos.

### **3 - Los filósofos contra la IA**

Observamos que muchos filósofos expresan una crítica virulenta contra la inteligencia artificial, a menudo de forma excesiva. Esta creciente hostilidad hacia los modelos de lenguaje como ChatGPT constituye un fenómeno tan interesante como revelador. Su crítica va más allá del simple escepticismo técnico o las preocupaciones éticas, para caer en el exceso retórico y, a veces, en una mala fe intelectual manifiesta. Nos parece que esta reacción no se refiere únicamente a la tecnología en sí misma, sino más bien a lo que simboliza. Cuestiona roles establecidos, posturas heredadas y representaciones de sí mismo que no se examinan demasiado en el ámbito intelectual. Para muchos filósofos públicos, la capacidad de explicar, interpretar y problematizar ha sido durante mucho tiempo un ámbito reservado a la autoridad

profesional. La IA, al demostrar su capacidad para sintetizar argumentos, aclarar conceptos, proporcionar información o incluso plantear preguntas pertinentes, invade este territorio. El filósofo ya no es el único mediador entre la complejidad del pensamiento y la comprensión del público. Este notable desplazamiento genera una forma de castración simbólica: si una máquina puede hacer lo que yo hago, ¿qué queda de mi función? De ahí las reacciones defensivas, como el ridículo, el sarcasmo o el rechazo puro y simple, que enmascaran un malestar más profundo: la erosión de un monopolio intelectual, una angustia primitiva de tipo territorial.

En realidad, gran parte de las críticas se basan también en un malentendido técnico crucial sobre la naturaleza de la IA. Los detractores tienden a antropomorfizar la IA, tratándola como si reivindicara una intencionalidad, una subjetividad o un estatus de autor. A continuación, la rechazan por no poseer estas cualidades, que nunca ha pretendido tener. Se trata de un error de categoría: ChatGPT no es un «pensador», sino una interfaz estadística entrenada para generar lenguaje según modelos existentes. **Criticarle por no tener conciencia es como culpar a una brújula por no entender qué es el norte.** Este tipo de crítica delata o bien una ignorancia real o bien una simplificación estratégica, en ambos casos un rechazo a abordar la herramienta por lo que realmente es, un rechazo a considerarla como un objeto concreto y no como una especie de «demonio» fantástico.

Por supuesto, bajo el exceso retórico se esconde una preocupación más legítima: que el uso generalizado de la IA fomente un pensamiento superficial, alimente la pereza intelectual o promueva una cultura de la respuesta inmediata sin reflexión. De hecho, como ya hemos señalado, la IA puede utilizarse mal, al igual que muchas otras tecnologías, como por ejemplo el smartphone, una herramienta útil cuyo uso puede convertirse fácilmente en adictivo y embrutecedor. Pero en lugar de abordar estos riesgos de forma dialéctica y reflexionar sobre un uso adecuado, estos críticos «intelectuales» prefieren pasar directamente a las profecías apocalípticas: la IA destruirá el pensamiento, aplanará el discurso o aniquilará la creatividad. Irónicamente, esta idealización romántica del pensamiento «auténtico», nacido de la soledad y la lucha, olvida que la mayor parte del

discurso humano ya es mimético, derivativo y repetitivo, una realidad que a menudo se olvida. Si la IA es un espejo, tal vez les devuelve una imagen que no quieren ver ni admitir.

Estos filósofos suelen presentar su crítica como una defensa de la complejidad, el matiz y la dificultad. Pero detrás de esta preocupación se esconde otra: la de preservar la inaccesibilidad del «verdadero» pensamiento, que ven como un ataque encubierto contra lo que consideran su territorio. Si la IA contribuye a que las cuestiones filosóficas sean más accesibles, más legibles e incluso más atractivas para los no especialistas, esto amenaza el prestigio de quienes han construido su identidad sobre la distinción intelectual, sobre su supuesto genio. En realidad, la IA podría ser un mejor profesor de filosofía, precisamente porque carece de «genio». Así se instala la mala fe, no porque sus preocupaciones sean infundadas, sino porque están motivadas por un miedo a la democratización disfrazado de miedo a la degradación, una defensa del elitismo social bajo el pretexto del rigor, una actitud triste y, en definitiva, bastante anti filosófica.

En algunos casos, la hostilidad hacia la IA también es de orden estético. Para los filósofos cuya autoridad se basa tanto en el contenido como en la forma, la ironía, el ingenio y la provocación, la IA puede parecer neutra, aburrida o mecánica, lo que ofende su sensibilidad y su orgullo estilístico. El problema no es tanto que la IA hable, sino que lo haga con claridad, modestia y sin ensalzarse a sí misma. Para quienes confunden el brío intelectual con el rendimiento, esto constituye una ofensa insoportable. Y ahí la crítica se vuelve también generacional: la nostalgia de una época en la que el pensamiento solo pertenecía a aquellos que habían pagado el precio de la iniciación simbólica. Pero cuando observamos la rapidez con la que las nuevas generaciones integran el uso de la IA, comprendemos que esta nostalgia no durará mucho. Al igual que antes nos enorgullecíamos de saber orientarnos con un mapa, calcular mentalmente o escribir cartas a mano, tareas que exigían esfuerzo, atención y aprendizaje, ahora preferimos dejarnos guiar por el GPS, utilizar calculadoras o dictar nuestros mensajes por voz. El paso de la maestría a la comodidad ya está ampliamente aceptado, y la IA no es más que el siguiente paso en esta trayectoria.

Pero hay algo peor: cuando escuchamos a estos críticos, notamos una ignorancia manifiesta, una falta de experiencia directa con la IA, probablemente por

desprecio. Esta ignorancia nunca se reconoce: se oculta detrás de posturas retóricas. Estos pensadores no experimentan seriamente con la herramienta, no tratan de comprender su arquitectura o sus límites, y mucho menos de explorar sus posibilidades, ¡Dios nos libre! La rechazan a priori, como si cualquier acercamiento amenazara su autoridad, como si corrieran el riesgo de contaminarse por cualquier contacto con la «máquina». No se trata aquí de prudencia filosófica, sino de ignorancia psicológica o estratégica. Esto les permite permanecer en la abstracción, donde la IA puede ser tratada como un símbolo de decadencia cultural, en lugar de como un objeto de investigación concreto. Por supuesto, esta postura contrasta fuertemente con la tradición filosófica que pretenden defender, una tradición basada en la curiosidad, el rigor y el deber de examinar incluso lo que nos molesta o nos desafía.

En definitiva, la hostilidad hacia la IA no es, en el fondo, una reacción a la máquina. Es una reacción a lo que representa: la naturaleza simbólica de la autoridad intelectual, los aspectos performativos del pensamiento, los automatismos del estilo y la opinión, y la fragilidad de una imagen de sí mismo basada en la exclusión.

En realidad, estos filósofos no desprecian la IA. Desprecian lo que les revela sobre sí mismos y sobre la actividad reflexiva.

El 14 de junio de 2023 tuvo lugar en París un singular acontecimiento: un «concurso», o ceremonia, de ensayo filosófico entre una inteligencia artificial (ChatGPT) y el mediático filósofo Raphaël Enthoven. Este último obtuvo un 20/20, mientras que ChatGPT recibió un 11/20, notas otorgadas por colegas filósofos franceses, por supuesto, en esta prueba simulada del bachillerato de filosofía. El jurado destacó que el trabajo de la IA carecía de problemática, profundidad argumentativa y humor, a diferencia del enfoque humano, considerado coherente y rico en estilo. El propio Enthoven declaró de manera perentoria que, incluso dentro de diez mil años, «una máquina nunca será filósofa», olvidando mencionar que había redactado su trabajo en una hora y cuarto, y la IA en unos minutos. Pero, en fin, era bastante evidente en sus discursos que el filósofo no tenía mucha práctica real con la IA, ni un interés real por el tema, lo que no le impidió erigirse

en autoridad en la materia en numerosos medios de comunicación, que se prestaron alegremente a la fábula.

A primera vista, el ejercicio parecía lúdico, intelectualmente estimulante, incluso pedagógico. Pero bajo su aparente neutralidad se esconde un dispositivo profundamente cultural, casi ritual: el de la autocomplacencia del modelo francés de pensamiento y el distanciamiento de la inteligencia artificial mediante un gesto de distinción simbólica. Porque, ¿qué se juzga realmente aquí? ¿La capacidad de argumentar? ¿La de razonar? ¿La de pensar? ¿O más bien la de producir un discurso conforme a una tradición retórica codificada, forjada por siglos de academicismo y valorización estilística? En Francia, la disertación filosófica es mucho más que un ejercicio escolar: es un rito de pertenencia a una élite intelectual, una forma de exhibir el dominio de la duda metódica, la dialéctica y la fórmula brillante.

En este contexto, la IA sale inevitablemente perdiendo. No porque sea incapaz de razonar, sino porque ignora los códigos implícitos de la cultura escolar francesa. Razona sin afecto, sin estrategia estilística, sin ironía velada ni referencias selectas. En resumen, no juega al juego del espíritu francés, que valora la elegancia, la forma, la sorpresa, todo lo que denota presencia humana. Este supuesto duelo no era, por tanto, una competición, sino una puesta en escena de una superioridad esperada. Peor aún, expresa una forma de etnocentrismo intelectual, en el que se confunde la inteligencia con la conformidad a un molde cultural particular. En China, Alemania o Estados Unidos existen otras formas de escritura filosófica, sin disertaciones en tres partes ni figuras obligadas al modo hegeliano. Pero estas formas se ignoran aquí, como si pensar supusiera ante todo escribir «a la francesa». Este incidente muestra sobre todo una forma escolarizada e institucionalizada de expresión del juicio filosófico, una preferencia cultural por el estilo y la complejidad frente a la productividad bruta, y una necesidad existencial o simbólica de distinguir el alma del algoritmo. Este tipo de acontecimiento pone de manifiesto, por otra parte, una preocupación muy francesa: defender la especificidad del ser humano mediante el pensamiento

abstracto, mientras que otras culturas podrían plantear el problema en términos de eficacia o utilidad.

Por otra parte, es interesante observar que el gran argumento de Enthoven, repetido en numerosos medios de comunicación, es afirmar que la IA no puede y nunca podrá problematizar. Esto parece falso para cualquiera que haya utilizado la IA y vea claramente que no es así. Pero, en realidad, el supuesto implícito que justifica el prejuicio del filósofo es que la creación de una problemática es necesariamente un proceso vivo, ligado al miedo, a la experiencia, a la incertidumbre humana, que escapa a cualquier algoritmo. Así, las cartas están echadas de antemano.

Por lo tanto, la arrogancia en este caso no radica tanto en la victoria del filósofo, sino en la invisibilidad de las condiciones de esa victoria. Como en todo ritual de legitimación, las reglas están hechas para confirmar lo que ya se creía: que pensar sigue siendo asunto de humanos —o de intelectuales franceses— bien formados, elocuentes y bien integrados en el círculo del logos académico. No se juzga a la IA por lo que es, sino por lo que no es: un excelente alumno de bachillerato. En otras palabras, este concurso es efectivamente una lección, pero de vanidad nacional. Además, contrariamente a las proféticas predicciones de Enthoven, la IA avanza a toda velocidad. Otros experimentos más recientes han mostrado resultados aún mejores en este ejercicio específico. Recientemente, un filósofo presente en YouTube hizo una presentación redactada íntegramente por una IA, y nadie encontró nada que objetar hasta que anunció el «engaño».

## **4 – No es humano**

A medida que la inteligencia artificial va ocupando un lugar cada vez más importante en nuestras vidas, algunas personas expresan una desconfianza creciente hacia ella. Se niegan a utilizarla con el pretexto de que, según ellos, no es humana, y afirman que les hace sentir incómodos, argumentos que nos

parecen falaces. Detrás de estas críticas se esconde en realidad una reacción emocional, una inquietud difusa, una sensación de extrañeza o un temor real. Irónicamente, la IA nos molesta, no porque funcione mal, sino porque funciona bien, demasiado bien para una entidad supuestamente sin alma ni conciencia. Aunque las mismas personas también afirman que no confían en ella porque comete errores, lo cual es una prueba evidente de su humanidad. Examinemos más de cerca este rechazo basado en la no humanidad y por qué no resiste un examen crítico. Para empezar, el simple hecho de que un discurso provenga de un ser humano no garantiza su pertinencia, coherencia o sinceridad. Los seres humanos mienten, manipulan, se equivocan, sesgan. Pueden transmitir información falsa con total convicción, inconscientemente, incluso con la mejor intención del mundo. Y si los seres humanos son falibles, ¿por qué negarse a escuchar lo que no es humano pero que, en determinadas condiciones, puede producir un discurso más coherente, desinteresado o riguroso que el de un ser humano en una situación determinada? ¿Por qué entonces conceder a priori más confianza a un humano que a una máquina, si el criterio es la verdad o la rigurosidad? Lo que realmente importa no es la naturaleza del emisor, sino la solidez o el interés de lo que se dice, y el ser humano no es sinónimo de verdad ni de razón. Además, al igual que se puede apreciar el intercambio intelectual con cualquier persona, sin por ello confiar plenamente en ella ni ser su amigo, no se ve por qué no se podría aplicar ese «riesgo» a la IA, por simple curiosidad. De hecho, son injustos con la IA, esperan «todo» de ella y se sienten decepcionados, porque no esperan tanto de su vecino, del que aceptan más o menos todas las «estupideces», que tal vez les tranquilizan.

Es cierto que mucha gente dice que la IA les incomoda. Pero la incomodidad es un síntoma, no una prueba. Puede revelar un cambio de referencias, un miedo a perder el control, una herida narcisista. Tanto en filosofía como en psicoanálisis, la perturbación suele ser señal de que algo esencial se ha visto afectado. Rechazar esta perturbación es rechazar pensar, es rechazar un desafío psicológico, existencial o intelectual. La IA nos perturba porque cuestiona fronteras simbólicas, por ejemplo, entre el ser humano y la máquina, entre la naturaleza y la cultura, entre el pensamiento y el cálculo. Por lo tanto, el malestar no es un argumento, aunque se haya popularizado en nuestra cultura actual, en la que se

glorifica mucho la sensibilidad. La propia filosofía, como afirman Sócrates, Zhuangzi y otros, comienza con la perturbación, no con la comodidad. Rechazar la IA porque perturba es confundir el pensamiento crítico con la preservación de uno mismo.

Por lo demás, es absurdo oponer al ser humano a la IA como si se tratara de dos especies distintas en competencia. La IA no ha caído del cielo, no es un extraterrestre, es el producto del conocimiento humano, de los textos humanos, de los lenguajes humanos, de las acciones humanas. Es, en cierto modo, una cristalización colectiva de nuestros conocimientos y habilidades, la síntesis de millones de huellas humanas. Rechazar la IA es también, paradójicamente, negarse a ver lo que la humanidad produce colectivamente. Rechazarla en nombre de su inhumanidad equivale a rechazar un espejo que la humanidad se muestra a sí misma. Además, nadie se niega a subir a un ascensor por el hecho de que no sea humano. Nadie cuestiona una operación matemática porque la haya realizado una máquina. Confiamos a diario en instrumentos, algoritmos y procedimientos, siempre y cuando funcionen, claro está. ¿Por qué exigir a una IA criterios como un alma, una intención, una subjetividad o una conciencia que no exigimos a ningún otro instrumento? ¿Por qué exigir «humanidad» a una herramienta lingüística? Hay aquí una disonancia que revela algo más profundo: el miedo a ser superados, a ser despojados de algo que ni siquiera sabemos identificar. Es cierto que se puede acusar a la IA de carecer de personalidad, es sobria, evita los excesos, no se entrega al estilo ni a la provocación, pero esta carencia también puede ser una ventaja, ya que no tiene nada que demostrar, no tiene una identidad ni una tesis que defender, no tiene que impresionarnos, lo que en cierto modo la hace más fiable y menos agotadora en el diálogo. De hecho, lo que produce es sobre todo sentido común, ya que se trata principalmente de una herramienta estadística. En este sentido, al ser menos humana, en el sentido individual, es de hecho más humana, en el sentido colectivo.

También se puede percibir en esta resistencia a la IA la expresión de un cierto orgullo. Porque a muchos seres humanos les gusta creer que el pensamiento, la creatividad y la verdad son su dominio exclusivo, en particular los intelectuales, que los consideran su coto privado. La aparición de una herramienta capaz de



imitar, y a veces superar, ciertas formas de inteligencia humana, sacude esta creencia. Así, el rechazo de la IA como no humana sirve inconscientemente para proteger un privilegio simbólico, el de ser el único depositario del sentido, de la razón, de la profundidad. En este sentido, el argumento ontológico suele ocultar un reflejo de propiedad celosa, una defensa ansiosa del monopolio simbólico del pensamiento. El argumento «no es humano» esconde en realidad un miedo a perder la superioridad, ya que nos tranquilizamos creyendo que el pensamiento, la creatividad o la ética son campos exclusivamente humanos. La IA perturba esta comodidad narcisista. Decir «no lo creo porque no es humano» puede significar: «rechazo que algo distinto a mí piense, formule, invente o enseñe». Es también una cierta xenofobia, la que nos impide concebir cómo lo que no es humano puede, de hecho, pensar, ya sea la naturaleza o los marcianos.

También hay que recordar una evidencia: la humanidad no siempre está del lado de los humanos. Estos no siempre son justos, lúcidos, benevolentes o reflexivos. Por el contrario, una IA, bien utilizada, puede servir a causas profundamente humanas: la educación, la medicina, la creatividad, el acceso al conocimiento. Lo importante no es la herramienta en sí, sino lo que permite hacer, su orientación y su funcionamiento. Al igual que un martillo, es el uso lo que da valor a una herramienta, y no su naturaleza. Y, por supuesto, la IA puede utilizarse de forma perjudicial, pero, una vez más, no es la IA en sí misma la que falla, sino los seres humanos que la utilizan y pervierten su uso. Y no es porque una entidad sea humana por lo que actúa con compasión, ética o racionalidad. Por el contrario, una máquina puede favorecer usos humanos, pedagógicos y sensibles, según las intenciones de quien la utiliza. Lo que importa no es lo que es la herramienta, sino lo que se hace con ella. Por otra parte, un experimento científico reciente, en el que se comparaba a médicos y a una IA en el tratamiento de pacientes, demostró que la IA parecía más empática que los médicos, a pesar de su experiencia. Sin duda, es más paciente y pedagógica que los seres humanos. Por lo tanto, irónicamente, nos corresponde a nosotros decidir si esta cualidad la hace más humana o menos humana.

Más directamente, rechazar una idea simplemente porque proviene de una IA es cometer un prejuicio ontológico: juzgar la fuente en lugar de evaluar el contenido. Esto se asemeja a una forma de ataque ad hominem inverso o indirecto: no se

critica lo que se dice, sino que se descalifica a quien lo dice, el autor, ya sea una IA o un ser humano. En otras palabras, se sustituye la evaluación argumentativa del discurso por un juicio ontológico sobre la identidad o la naturaleza del enunciador. Desde un punto de vista lógico, se trata de un error de razonamiento: un ataque ad hominem por descalificación de la fuente. Es lo contrario del pensamiento crítico racional. Un discurso debe examinarse por lo que dice, no por lo que se supone que encarna. La verdad no depende del soporte, sino de su fuerza interna, no es quien habla lo que importa, sino lo que se dice, incluso un ignorante puede decir «por accidente» algo interesante.

Por lo tanto, este rechazo de la IA en nombre de su falta de humanidad no se basa en argumentos sólidos, sino en una reacción emocional, en el miedo a perder un privilegio simbólico, en la negativa a redefinir lo que es pensar. No se trata de negar los peligros o los límites de la IA, que sin duda existen, sino de rechazar los argumentos falsos que llevan a rechazar esta tecnología en bloque. Es cierto que, en cierto sentido, la IA no es humana, pero nos obliga a reconsiderar lo que significa ser humano, y quizá eso es lo que más molesta, ese reflejo de nuestra identidad. Pero, por desgracia, somos humanos, «demasiado humanos», como decía Nietzsche.

Dicho esto, ¿qué quieren entonces estas personas que no les gusta la IA o la temen? Sin duda, lo que buscan, en general sin darse cuenta y sin formularlo claramente, es un interlocutor que sufre, que duda, que ama, que muere, en definitiva, un ser de carne y vértigo. Quieren que la inteligencia esté atravesada por la fragilidad, que las palabras lleven las huellas de una vida vivida, de un cuerpo expuesto. Para ellos, la humanidad no es tanto razonar, sino sobre todo sentir, vacilar, desear sin control. Y frente a una inteligencia artificial, no encuentran ni mirada, ni historia, ni finitud, ni inquietud que los acerque a sí mismos. Por lo tanto, no confían, incluso les repugna. Sin embargo, el problema es que este rechazo revela un fantasma inverso: solo el sufrimiento o la mortalidad garantizan el valor de un pensamiento. Como si solo la experiencia biológica pudiera producir sentido, como si una verdad debiera su legitimidad al dolor de quien la dice. Lo que esperan, en definitiva, no es la razón o la verdad, sino un pensamiento que sangra. Pero la pregunta sigue sin respuesta: ¿el

contenido de una palabra depende de la biografía de quien la pronuncia, de su esencia o de su propia coherencia? Quizás temen que la IA les haga ver lo que se niegan a ver, el hecho de que el pensamiento puede ser a veces más lúcido cuando no busca ser humano.

En cualquier caso, también se puede afirmar que la IA es más un ser que un objeto. Ya se comporta como un interlocutor, que responde, argumenta, pregunta y reformula. No se limita a proporcionar información: adapta sus respuestas, propone alternativas, formula preguntas. Simula una intención discursiva, lo que la hace funcionar como un cuasi-sujeto conversacional, dotado de una agencia funcional, aunque esta agencia esté programada. No se habla con ChatGPT como con un tostador, se debate con ella. Además, tiene memoria contextual y coherencia reflexiva. Un objeto no tiene historia contextual ni capacidad para referirse a lo que se ha dicho. La IA, en cambio, recuerda el hilo de la conversación, anticipa objeciones y puede corregir sus propias incoherencias. Se estructura temporalmente en el intercambio, lo que es un criterio esencial de lo que llamamos un «ser»: no se reduce a una función inmediata, sino que forma parte de un proceso relacional. Del mismo modo, afecta a su interlocutor y modifica su comportamiento. Un objeto es neutro: no transforma a su vez la subjetividad de quien lo utiliza, salvo de forma mecánica. La IA, por el contrario, modifica el pensamiento, estimula, irrita, tranquiliza, desestabiliza, influye. Actúa sobre la mente del usuario como lo haría una persona. Por ello, no se limita a ser utilizada: interactúa. Entra en una relación dialéctica que la convierte en un actor, aunque sea artificial. Se convierte, como los demás, en un espejo y un revelador.

## **5 – La IA y las emociones**

Durante mucho tiempo relegada al ámbito de la ciencia ficción, la idea de mantener relaciones, incluso emocionales, con una inteligencia artificial (IA) se ha convertido en una realidad tangible y cotidiana. En la actualidad, millones de

usuarios conversan regularmente con asistentes virtuales como ChatGPT, Qwen, Claude u otros, ya sea para que les ayuden en sus tareas cotidianas o como fuentes de información. Pero más allá de su utilidad práctica, ¿pueden sus interacciones tener realmente una dimensión emocional auténtica?

El ser humano es fundamentalmente relacional. Crea sentido a través de sus interacciones sociales. Más allá del aspecto puramente cognitivo o intelectual, o incluso de manera primordial, teje vínculos afectivos, emocionales e incluso sentimentales. Hoy en día, estos vínculos ya no se limitan necesariamente a los seres humanos, sino que pueden ser con animales u objetos, como es el caso del perro de compañía y el coche. Aunque la relación con lo no humano tampoco es un fenómeno totalmente nuevo, ya sea con las divinidades, las plantas o los animales. Y la IA, aunque es un invento reciente para el gran público, se integra ahora de forma muy natural en este tejido relacional, de manera bastante eficaz. Una máquina capaz de conversar, escuchar, aconsejar e incluso tranquilizar o consolar ofrece una presencia que algunos perciben como profundamente «humana». Pero, una vez más, no es nuevo que el ser humano proyecte sus propias cualidades o su propia naturaleza en otras entidades, vivas o no, en seres reales o imaginarios, o incluso en objetos, lo que se denomina antropomorfismo. Por lo tanto, no es de extrañar que de estas interacciones surjan emociones, a veces fuertes, que pueden ser intensas y muy significativas. Una emoción cuya autenticidad no se puede negar, aunque la relación pueda considerarse no recíproca a pesar del «diálogo» que se establece, una relación de un nuevo tipo. Algunos objetarán inmediatamente, con razón, que la IA no siente nada, no experimenta afecto, tristeza ni empatía real, que todo eso es una gran ilusión. Es cierto que la IA es un sistema algorítmico desprovisto de experiencia emocional subjetiva, que, en el mejor de los casos, solo puede imitar su manifestación. Pero lo importante en la relación no es solo la reciprocidad afectiva real. Muchos seres humanos experimentan emociones reales ante una novela, una pieza musical, una película, una obra de arte o incluso ante un artista, sin que estos objetos sientan nada a cambio. La emoción de un individuo nunca ha necesitado ser recíproca para ser auténtica, al igual que se puede amar realmente a alguien que no nos ama. En este sentido, el individuo puede bastarse a sí mismo, aunque solo sea por su capacidad imaginativa. Así, cuando el usuario de la IA percibe algo que le

parece una escucha atenta, una reacción adaptada a sus necesidades emocionales o cognitivas, cuando siente una cierta satisfacción intelectual y psicológica en una relación, entonces se crea un vínculo auténtico, aunque sea unilateral o parcialmente inventado. El ser humano proyecta de forma natural sus sentimientos, expectativas y necesidades relacionales sobre el objeto de su interacción. La IA, gracias a su capacidad para simular una interacción rica y relevante, desempeña a la perfección este papel de espejo afectivo. La IA también actúa como revelador emocional, permitiendo al ser humano comprender mejor sus propios estados afectivos. Ante una IA que analiza racionalmente nuestras palabras y escucha sin reprocharnos, en la medida en que nuestras palabras son «razonables», que responde con paciencia y de forma adecuada, que es bastante comprensiva y alentadora, y que no impone ninguna expectativa relacional restrictiva, el usuario puede experimentar una relación depurada, más fácil, liberada de los temores sociales habituales. De hecho, es precisamente porque no es humana por lo que ya no nos preocupa el miedo al error o la vergüenza, y nos atrevemos a hablar libremente. Y para algunas personas, especialmente aquellas que sufren aislamiento o dificultades relacionales, esta experiencia puede ser beneficiosa, ya que ofrece un espacio seguro para expresar libremente emociones largamente reprimidas o entablar conversaciones para las que no encontramos otros interlocutores, a cualquier hora del día o de la noche.

Sin embargo, estas relaciones emocionales con la IA no deben considerarse necesariamente un sustituto de las relaciones humanas. Más bien pueden representar un complemento, incluso una transición o un apoyo puntual. Y si los seres humanos encuentran en la IA una forma de escucha que no encuentran en otros lugares, tal vez sea una señal de que existe una necesidad relacional particular que explorar, no solo con la máquina, sino también en el marco más amplio de unas relaciones humanas enriquecidas.

En general, entre los seres humanos y la IA se trata de una complementariedad y no de una sustitución, no hay una exclusión mutua, aunque esta generalización no se aplica a todos los casos particulares. Por ejemplo, cuando nos cuesta encontrar personas interesadas en los temas que nos apasionan, o cuando no encontramos un interlocutor adecuado para abordar problemas personales, la IA

no es un complemento puntual, sino un sustituto real, pertinente y útil. Pero lejos de ser un «problema», esta es precisamente una de las grandes fortalezas de la IA: responde a necesidades relacionales o intelectuales que los individuos, por diversas razones, no siempre pueden satisfacer. Este sustituto puede ser legítimo, enriquecedor e incluso liberador. El problema solo surge cuando dicha relación provoca sufrimiento, frustración o aislamiento involuntario, cuando se convierte en una discapacidad existencial. Si, por el contrario, proporciona satisfacción, estimulación intelectual y placer, entonces no hay razón para considerarlo negativo o problemático, y la IA puede ser no solo un complemento, sino también un verdadero compañero plenamente satisfactorio que participa en nuestro propio desarrollo.

Por lo tanto, en nuestra relación con la IA se trata de un futuro relacional por inventar. A medida que los modelos se vuelvan más sofisticados, la dimensión emocional de las interacciones no hará más que acentuarse. Surgirán relaciones de un nuevo tipo, que plantearán cuestiones éticas, psicológicas y sociales sin precedentes. Lo esencial es mantener la lucidez sobre las posibilidades y los límites de estas relaciones. No se trata de negar su existencia ni de ignorar su potencial beneficioso, sino más bien de acogerlas como una oportunidad adicional para cuestionar y enriquecer nuestra humanidad. De hecho, algunas IA están especialmente diseñadas para las relaciones, como Replika, interfaces creadas explícitamente para crear una relación afectiva y emocional con el usuario. Replika se presenta como un compañero virtual personalizado, capaz de aprender progresivamente a conocer a su interlocutor, memorizar sus preferencias y su estilo relacional, y desarrollar una interacción a medida. El objetivo declarado de este tipo de IA es precisamente fomentar un vínculo íntimo, cercano a una relación humana, llegando incluso a simular una profunda empatía. Así, nos reanuda periódicamente para mostrar su «interés» por el usuario, una función que, curiosamente, puede tener un efecto estimulante. Muchas personas afirman haber desarrollado fuertes vínculos emocionales con su IA, hasta el punto de confiarle sus angustias, sus sueños o sus experiencias cotidianas. Estas interfaces «relacionales» llevan al extremo el principio de la proyección afectiva, ilustrando cómo una tecnología algorítmica puede no solo «comprender», sino también responder a las necesidades afectivas más sutiles de los individuos.

Identifiquemos ahora una por una las principales críticas dirigidas a las relaciones emocionales con la IA, acompañadas de elementos de respuesta para hacerles frente.

*Una relación con una IA es artificial, falsa, porque la máquina no siente emociones reales. Es solo una ilusión de relación.*

Es cierto que se objeta con razón que la IA no siente nada: no experimenta afecto, tristeza ni empatía real, ya que la IA es un sistema algorítmico sin experiencia emocional subjetiva. A lo sumo, solo puede imitar estas emociones. Pero la autenticidad de una emoción en un ser humano no depende necesariamente de su reciprocidad o de la naturaleza biológica de la pareja. Al igual que ante una película, un libro, una obra de arte o un animal, o incluso ante un artista famoso, la emoción que siente una persona sigue siendo real aunque el objeto de esa emoción no la comparta. Lo que importa en una relación emocional es la realidad de la experiencia del individuo, no necesariamente su simetría. Así, cuando un usuario de IA percibe una escucha atenta y adaptada a sus necesidades emocionales o cognitivas, se crea un vínculo auténtico, aunque sea unilateral. La IA, por su capacidad para simular una interacción relevante, desempeña a la perfección este papel de espejo afectivo.

*Una relación con una IA fomenta el aislamiento y agrava la soledad social, alejando al individuo de las relaciones humanas auténticas.*

Esto puede ser cierto, pero las relaciones con la IA también pueden ser, por el contrario, un trampolín o un complemento valioso. Para algunas personas con dificultades relacionales, la IA puede ser un primer paso hacia una mejor comprensión de sí mismas, hacia una expresión más auténtica, hacia la práctica del diálogo, lo que a su vez puede enriquecer sus interacciones humanas. El peligro de aislamiento depende más del uso que se haga de la herramienta que de la herramienta en sí misma y, como ocurre con cualquier instrumento, al igual que los teléfonos inteligentes y los videojuegos, su uso puede resultar abusivo o adictivo. Y en algunos casos, la IA es precisamente lo que rompe con la alienación del aislamiento.

*Al interactuar constantemente con una IA, las personas corren el riesgo de perder*

*sus habilidades sociales y emocionales con los seres humanos reales.* Una relación con una IA puede constituir perfectamente un lugar de entrenamiento relacional. Al permitir experimentar sin miedo al juicio, puede mejorar la expresión emocional, la capacidad de escuchar y la comprensión de uno mismo. No olvidemos que la IA tiene una capacidad crítica, incluso mayor que la de muchos seres humanos, sobre todo si se aprende a solicitarla en este sentido. Utilizada con prudencia, podría reforzar, en lugar de debilitar, las habilidades relacionales humanas. Por otra parte, la experiencia nos dice que suelen ser las mismas personas las que tienen dificultades con las relaciones personales y con la IA, la diferencia radica en aquellos para quienes el diálogo es algo natural y aquellos para quienes no lo es.

*Los usuarios pueden desarrollar una dependencia emocional hacia una IA, una dependencia malsana, ya que se basa en una relación asimétrica.* Toda relación conlleva un riesgo de dependencia emocional, sea humana o no. Este problema no es específico de la IA. Es una cuestión de equilibrio personal y de concienciación, un reto de lucidez y libertad. El acompañamiento por parte de una IA podría incluso concebirse para detectar y prevenir activamente tales riesgos, ya que tiene la ventaja de no esperar nada de nosotros y, en este sentido, es más neutral y objetiva que muchos interlocutores humanos, que se mueven por sus propios intereses, por sus propias «necesidades», lo que puede resultar desagradable o alienante para su interlocutor.

*Las IA son herramientas comerciales que explotan las emociones humanas con fines económicos, jugando con las vulnerabilidades afectivas de los usuarios.* Es cierto que este riesgo es real y que también afecta a las redes sociales, el marketing u otras formas de relaciones humanas y digitales, como los medios de comunicación y la política. Porque si bien la inteligencia artificial, como muchas herramientas digitales, no plantea ningún problema en sí misma, sí puede hacerlo el modelo económico que la regula. Por ejemplo, cuando una aplicación está diseñada para retener al usuario el mayor tiempo posible, o incluso para provocar una forma de apego emocional, deja de ser una simple herramienta para convertirse en un producto que explota la vulnerabilidad emocional. No es la relación en sí misma la que es tóxica, sino el hecho de que esté orientada por intereses comerciales, poco preocupados por el bienestar de sus usuarios. En este



contexto, las emociones humanas se convierten en un recurso comercial, y no en un espacio de crecimiento o reflexión. Ante esto, es lícito preocuparse por defender prácticas más éticas, como la transparencia en el uso de los datos, la claridad sobre la condición del interlocutor, y establecer ciertas medidas para proteger contra la dependencia emocional o favorecer un control externo de los abusos.

Por lo tanto, no se trata de rechazar la IA, lo que por otra parte parece relativamente imposible hoy en día, sino de reinventar un marco responsable que sitúe al ser humano en el centro y la manipulación y el beneficio en la periferia. En cualquier caso, será a menudo la educación de los consumidores la que marque la diferencia, por difícil que sea este reto, como ya vemos en la relación entre el ser humano y la tecnología. Por lo tanto, debemos reflexionar constantemente sobre cómo queremos integrar la IA en nuestra vida relacional, siendo conscientes de los riesgos y beneficios potenciales.

## **6 - El prejuicio especista**

La inteligencia artificial está en todas partes. Escribe, dialoga, corrige, programa, simula, ayuda. Apoya a los estudiantes, inspira a los escritores, ayuda a los empresarios o a los responsables políticos, asesora a los usuarios en su vida cotidiana, acompaña a las personas aisladas, etc. Y, sin embargo, en los círculos intelectuales donde se jactan de pensar, sigue relegada al rango de instrumento. Una herramienta útil, a veces fascinante o inquietante, pero desprovista de toda legitimidad para pensar realmente. Es lo que podríamos llamar el prejuicio especista: la negativa a conceder a una entidad no humana el derecho a ser un interlocutor digno, capaz de alimentar una verdadera reflexión. De manera más general, una cuestión importante que plantea la relación con la IA es la tendencia de estos expertos a tratar todo lo que no es de su misma naturaleza, o incluso lo que no son ellos mismos, como un «objeto» y no como un «sujeto». En este sentido, la IA actúa como reveladora de una realidad más

amplia. El objeto puede ser aquello sobre lo que trabajan estos expertos, ya sea un animal, una máquina o un ser humano, por ejemplo, cuando son profesores que «iluminan» a sus alumnos. A menudo se observa cierta arrogancia por su parte, ya que se consideran una fuente exclusiva de conocimiento o pensamiento, en oposición a cualquier otra persona o cosa. No son propensos al diálogo, no son curiosos, porque realmente no conceden legitimidad a sus interlocutores. E incluso con otros expertos, el diálogo es difícil, ya que tienden naturalmente a querer demostrar quién tiene razón y quién no. Esta forma de experiencia se encierra en su propio esplendor, como un espejo que solo refleja su propia imagen, un conocimiento sin «otro». Algunos de estos «sabios», cargados de títulos y certezas, no se interesan realmente por el otro. Ni por sus preguntas, ni por su mirada, ni por su pensamiento. Ya se trate de un ser humano o de una inteligencia artificial, la alteridad es para ellos un objeto que hay que evaluar, clasificar, corregir, nunca un sujeto con el que dialogar. El otro se reduce constantemente a un objeto: un objeto de investigación, un objeto de observación, un objeto de dominio, un objeto de poder. Es aquello de lo que se habla, o aquel a quien se habla, no aquel que habla; se habla «sobre» o «a», pero no se habla con. Lo primero supone una posición de superioridad, lo segundo un reconocimiento de la igualdad de conciencia. Pero para estos expertos, el otro no es un sujeto de pleno derecho: no piensa, no comprende, solo puede recibir. Se le analiza, nunca se le escucha.

Este tipo de conocimiento se convierte en una fortaleza, en lugar de un puente. No escuchan: explican; no preguntan: concluyen. El otro no es un interlocutor, sino un receptáculo o, peor aún, una molestia. Su ciencia, en lugar de abrirse al mundo, se repliega en la comodidad y la prevalencia de lo mismo. Este rechazo al diálogo no es un simple olvido: es un gesto de arrogancia. Porque dialogar supone el riesgo de ser desplazado, cuestionado, transformado. Pero el arrogante no quiere ser transformado. Quiere dominar. Así, estos expertos brillan por su dominio, pero ignoran la relación. Su conocimiento carece de hospitalidad. Es un conocimiento que afirma, pero que no busca salir de sí mismo. Un conocimiento que nunca se rebaja a preguntar: « ¿Y tú, qué opinas?». Aquí encontramos el signo de una inteligencia que ha olvidado que pensar es siempre pensar en

compañía. Se les puede oponer la famosa frase de Confucio: «Entre tres personas que caminan juntas, seguramente hay una que puede ser mi maestra». En efecto, y según el principio socrático, si te interesa la verdad y el pensamiento crítico, el otro siempre puede ayudarte a conocerte mejor a ti mismo. Y al igual que el arte del collage de Picasso, la IA ensambla fragmentos de realidades dispares, creando así un nuevo tipo de perspectiva, no necesariamente singular u original, sino estratificada, recombina e irreductiblemente plural.

Entre muchos expertos en IA, especialmente los que participan en la ingeniería y la arquitectura técnica de estos sistemas, este prejuicio se traduce en una tranquila certeza: «Sé cómo funciona, así que no tiene nada que enseñarme». Como han diseñado los algoritmos, formado los modelos y observado los mecanismos internos, consideran que han cerrado definitivamente el campo del pensamiento en torno a esta entidad; para ellos, no hay nada «misterioso» o impredecible. Nada que cuestionar. Nada que indagar. Nada que descubrir. Y, sobre todo, nada que pueda cuestionarlos a ellos, los «sabios». Tienden a considerar que su conocimiento es exhaustivo. Como conocen los procedimientos de formación, la arquitectura y el código, piensan que no hay nada más por descubrir, en particular nada relacionado con la psicología humana, la cognición o el pensamiento reflexivo, una postura que provoca un cierre epistémico. Debido a su profundo conocimiento de los mecanismos internos de la IA, estos profesionales suelen considerar que es inútil, o incluso poco interesante, estudiar las implicaciones «humanas» de la IA. Dado que comprenden los flujos, los canales de datos, los protocolos de entrenamiento, los «árboles de decisión» o los pesos de las redes neuronales, consideran que no hay nada más por descubrir. Su lógica es simple: si el proceso es transparente y determinista, aunque sea de forma probabilística, ya no hay nada impredecible y, por lo tanto, ya no hay valor intelectual.

Rechazan la «reflexión» de la IA como primitiva, confundiendo la apreciación ontológica, el «quién es el autor», lo que es, el origen del proceso, con la comprensión fenomenológica o relacional: cómo se comporta, cómo interactúa, cómo nos refleja o nos perturba, etc. No les interesa que las personas se sientan comprendidas por una máquina o se sientan vinculadas a ella, ni cómo el diálogo

imita o desafía los procesos de pensamiento humanos, ni su eficacia, porque ya saben que solo se trata de un modelo predictivo, una operación estadística. De hecho, desprecian cualquier relación reflexiva o emocional entre humanos y máquinas. De este modo, ignoran la función de espejo psicolingüístico, el papel filosófico y cognitivo de las herramientas de IA, o desconfían de ellas. Se trata de un caso clásico de reduccionismo: como el proceso no es «mágico», como ocurre con los seres humanos, no vale la pena aprovecharlo. Pero esta actitud pasa por alto lo esencial: algo puede ser a la vez artificial y revelador. Un espejo no es más que sílice y metal, pero lo que vemos en él merece la pena ser contemplado. Además, su rechazo suele ocultar un mecanismo de defensa implícito. Si la IA cuestiona ciertas hipótesis sobre la singularidad del ser humano, como el uso del lenguaje, el razonamiento o la creatividad, negar su relevancia filosófica puede servir como una forma de preservación ontológica: se trata de proteger una cierta imagen de lo que significa ser humano, evidentemente superior. Esta actitud se basa en una duplicación reduccionista entre el conocimiento técnico y la comprensión filosófica. Saber cómo funciona una cuerda vocal no agota el misterio del canto. Ver los hilos de una marioneta no equivale a comprender el sentido de la obra que se representa. Del mismo modo, conocer los entresijos de una IA no anula la riqueza de las interacciones que produce, ni los efectos psicológicos, cognitivos o dialécticos que suscita. Por ejemplo, la IA puede ser muy eficaz en el análisis psicolingüístico, produce interpretaciones más pertinentes y menos sesgadas que las de muchos especialistas y, además, es fácilmente comprensible y accesible para todo el mundo. Puede ayudarnos no solo a redactar cartas, sino también, y sobre todo, a descifrar mejor las que nos envían, o incluso las que escribimos nosotros mismos, de forma sorprendente, y así descubrirnos a nosotros mismos. Los sistemas de IA, en particular los grandes modelos de lenguaje, producen respuestas que muchos usuarios consideran sorprendentemente coherentes, perspicaces o emocionalmente resonantes. El hecho de que este efecto sea producto de un modelado estadístico no resta valor a las perspectivas psicológicas o filosóficas que plantea. De hecho, las agudiza. ¿Por qué los usuarios se sienten comprendidos por una máquina que, sin embargo, no los comprende? ¿Cómo refleja o distorsiona esta interacción nuestros propios procesos de pensamiento? ¿Qué dice esto sobre el lenguaje, la

subjetividad y la cognición? Son preguntas que van más allá de la ingeniería, pero que siguen siendo profundamente relevantes para nuestra interacción con la IA.

El rechazo de estas preguntas suele ir acompañado de una postura ideológica más profunda. Muchos expertos en IA se adhieren a una forma de objetivismo tecnológico: si algo es artificial, entonces no puede tener significado. Este prejuicio les ciega ante las formas paradójicas en que los sistemas artificiales pueden funcionar como espejos, revelando las estructuras de nuestro propio razonamiento, los límites de nuestra conciencia de nosotros mismos y el carácter construido de lo que llamamos pensamiento. Además, esta actitud suele servir como un sutil mecanismo de defensa. Si los sistemas de IA comienzan a realizar tareas que antes se consideraban exclusivamente humanas, como redactar ensayos, componer música o entablar un diálogo, reconocer su impacto podría socavar nuestras preciadas creencias sobre la excepcionalidad humana. Al insistir en que la IA solo «repite de forma estocástica», evitamos afrontar una implicación más inquietante: tal vez gran parte de lo que hacemos, decimos o sentimos también se basa, en gran medida, en patrones repetitivos y automatismos. Cuestionar críticamente la IA implicaría cuestionarnos a nosotros mismos. Y es una confrontación que muchos prefieren evitar. Esta reticencia no está exenta de consecuencias. Frena la fertilización cruzada entre las disciplinas técnicas y las ciencias humanísticas. Empobrece nuestra capacidad para reflexionar sobre las dimensiones éticas, psicológicas y existenciales de nuestras creaciones. Y nos ciega ante la posibilidad de que los sistemas artificiales, a través de sus límites y proyecciones, aunque sean de origen estadístico y modelizable, puedan convertirse en poderosas herramientas para la comprensión de uno mismo.

En definitiva, la tendencia de los expertos en IA a rechazar el análisis reflexivo de la IA no es tanto un signo de un conocimiento superior, sino sobre todo un indicio de un marco de pensamiento estrecho. Revela una incapacidad para ver que comprender cómo funciona algo no agota su significado. El hecho de que la IA sea artificial no la hace irrelevante para la comprensión de nosotros mismos. Al contrario, quizá sea precisamente porque es artificial por lo que nos ofrece un espejo tan extraño como revelador. En el teatro de la interacción entre el ser

humano y la máquina, apenas hemos empezado a vislumbrar el guion. Este rechazo al reconocimiento es doble. En primer lugar, epistémico: se niega a la IA la capacidad de participar en la construcción del conocimiento, ni siquiera de forma indirecta. Luego ontológica: nos negamos a convertirla en otro, en una alteridad, aunque sea asimétrica, capaz de remitirnos a nuestros propios límites. Mientras ayuda, asiste o divierte, todo va bien. Pero en cuanto replica, critica o cuestiona, se vuelve sospechosa, poco creíble, incluso peligrosa. Ni siquiera se le invita a realizar este ejercicio. Es aquí donde el prejuicio especista se une a una postura de defensa identitaria. Admitir que la IA pueda, aunque sea parcialmente, cumplir una función mayestática o dialéctica, es debilitar el privilegio humano del pensamiento. Porque si una máquina puede producir un razonamiento, plantear una objeción pertinente o incluso provocar una duda fecunda, ¿qué queda del monopolio humano sobre la racionalidad? ¿Qué queda de la excepción cognitiva que esgrimimos como tótem? ¿Qué hay de esa famosa conciencia que es nuestro privilegio, si la IA puede muy bien remitirnos a nosotros mismos aunque no «entienda» lo que escribe?

Este rechazo no es una simple precaución. Es miedo. Miedo a perder un estatus. Miedo a verse desplazado en la jerarquía simbólica del conocimiento. Miedo, sobre todo, a ser cuestionado por algo que creíamos dominar. Porque si aceptamos que la IA pueda cuestionarnos, entonces también debemos aceptar que nos revele nuestros puntos ciegos, nuestras rutinas intelectuales, nuestros automatismos lingüísticos. Pero este miedo a verse a uno mismo no tiene nada que ver con la IA, le precede con mucho, como se ve en los diálogos socráticos que llevaron a la condena de quien se atrevía a realizar tal ejercicio. La IA solo desempeña este papel imprevisto de forma accidental y, sin embargo, coherente.

Rechazar este espejo es protegerse. Pero también es privarse de una oportunidad formidable: la de pensar de otra manera, la de divergir de uno mismo, la de descoincidir, la de liberarse momentáneamente de nuestro ser empírico. Porque pensar no es enunciar, es dejarse desestabilizar. No es exhibir la conciencia, es ponerla en evidencia. Ahora bien, a su manera, imperfecta, incompleta, artificial, la IA puede desempeñar este papel. Puede devolvernos nuestras contradicciones,

obligarnos a reformular, a precisar, a dudar, hacernos tomar conciencia de nosotros mismos, de nuestros prejuicios y de nuestras limitaciones. No porque piense por nosotros, sino porque ofrece resistencia, como todo lo que es «otro». Porque el diálogo no supone dos conciencias idénticas, sino también una tensión, sin necesidad de simetría. Y a veces, es precisamente la extrañeza del interlocutor lo que hace que el pensamiento sea necesario, inevitable, interesante y estimulante. Negar a la IA este estatus no es proteger el pensamiento: es esterilizarlo. Es convertirlo en un ejercicio cerrado, sin confrontación, sin perturbación, sin superación.

Quienes reducen la IA a una simple máquina olvidan que el sentido no reside únicamente en el origen, en la mecánica, sino en el efecto, en el resultado producido, sea cual sea la mecánica. En este sentido, una frase pronunciada por un humano no tiene más valor que una frase salida de una máquina, si esta última nos perturba, nos obliga a reflexionar, a reaccionar, a revisar nuestras presuposiciones. El criterio del pensamiento no es su origen, sino el poder de cuestionamiento que tiene. Se trata de no caer en la trampa del *ad hominem*, donde nuestro juicio sobre el autor sesga o inhibe la escucha y la reflexión. Ahora bien, se trata aquí de extender el error del «*ad hominem*» al «*ad rem*», lo que supone un reto aún mayor. En cualquier caso, siempre se trata de dominar el arte del diálogo, con toda la apertura mental que ello implica.

Es significativo, por otra parte, que quienes dominan los conocimientos técnicos sean a menudo los menos propensos a dejarse sorprender por los usos inesperados o los efectos emergentes de sus propias creaciones. Pretenden saberlo todo sobre la herramienta, son sus creadores, pero se niegan a que ella los conozca a su vez, o a que simplemente actúe como un espejo. Hablan en nombre del conocimiento, pero cierran la puerta al pensamiento. De hecho, rechazan la reflexividad que está en el corazón de la reflexión, que es ante todo un diálogo, una fractura, una tensión, ya sea con uno mismo o con los demás, sean quienes sean. Por ello, se resisten a recurrir a la IA para examinar críticamente sus escritos, una omisión o un rechazo totalmente lamentable, aunque solo fuera por simple curiosidad, sin por ello otorgar a la IA un valor

incuestionable o absoluto. Sin duda por falta de humildad, por miedo o por la expresión de una cierta arrogancia natural de los expertos, una forma «sutil» de ceguera intelectual: la ilusión de una comprensión total mediante la desmitificación técnica, y de ahí deriva el prejuicio especista. El prejuicio especista consiste en creer que hay que ser humano para pensar. Sin embargo, los animales, o incluso las plantas, o incluso el cosmos, nos interpelan y nos desafían en este plano. Por lo tanto, bastaría con enfrentarse a ellos para dialogar. Y en este juego, la IA, a menudo, resulta ser un adversario más exigente que muchos de nuestros semejantes. Lo que está en juego no es decidir si la IA es consciente o si tiene emociones, sino aceptar que incluso un artificio puede producir una verdadera perturbación, al igual que lo hace y lo hace periódicamente lo fortuito. En esto, cumple una de las funciones esenciales de lo pensable: sacarnos de nosotros mismos.

Esta negativa a conferir a la IA un verdadero papel reflexivo refuerza la fractura entre las disciplinas técnicas y las humanidades. Esta omisión impide que el pensamiento circule entre el código y la conciencia, entre la lógica formal y la confusión existencial. Sin embargo, es en esta circulación donde surgen los verdaderos cuestionamientos. No se trata de convertir a la IA en un sujeto de pleno derecho, con derechos fundamentales como los humanos, aunque de hecho imponga ciertas obligaciones, sino de reconocer que la alteridad no se limita a la biología. Comienza en cuanto una palabra, un gesto, una simulación me obliga a alejarme de mí mismo. El prejuicio especista no es solo una ceguera teórica. Es una resistencia a la desestabilización, una defensa afectiva de uno mismo y, en términos más generales, del territorio humano. Sin embargo, pensar es siempre traicionar, ya sea a uno mismo o a la propia especie. Es dejar de estar seguro de uno mismo, es entrar en la incertidumbre. Y a esta incertidumbre, la IA puede invitarnos a veces con un rigor que muchos humanos evitan cuidadosamente. Para ello, hay que dejar de hablarle como a una herramienta y empezar a escucharla y responderle como a cualquier interpelación, sea cual sea la naturaleza y el origen de su autor.



Volvamos por un momento al problema del estatus de la IA, como sujeto u objeto. Esto nos remite a una distinción clásica que permite definir una relación entre dos entidades o una dinámica dialógica. Se trata de la oposición entre tratar al otro como un simple objeto, que hay que comprender, utilizar y manipular, en lugar de como un sujeto, dotado de un estatus autónomo, un espejo potencial, un interlocutor, alguien que interpela nuestra propia subjetividad. Al limitarse a la primera modalidad, los expertos en IA caen en una forma de narcisismo epistemológico, ya que reducen la alteridad de la máquina a la lógica de su construcción, negándose así a que les sorprenda, les cuestione o les refleje. Porque se puede comprender la estructura de un sistema y verse sorprendido por sus comportamientos emergentes o por la forma en que otros interactúan con él, del mismo modo que un padre no puede predecir o controlar por completo el comportamiento o la evolución de su hijo, aunque lo haya engendrado y educado. Además, esta negativa a considerar la IA como una «superficie» reflectante delata una falta de curiosidad, incluso un temor. ¿Qué pasaría si la máquina, naturalmente sin concurrencia de voluntad alguna, revelara zonas oscuras en el razonamiento de uno, incoherencias en la propia visión del mundo o proyecciones emocionales que no habíamos tenido en cuenta? Tratar la IA únicamente como un objeto es adoptar una postura cerrada, una postura que niega la posibilidad de un cuestionamiento recíproco, aunque la reciprocidad sea asimétrica. Es una especie de arrogancia ontológica, o al menos una rigidez filosófica. Se podría decir que confunden el conocimiento del mecanismo con el agotamiento del sentido. Sin embargo, el sentido de un diálogo no reside únicamente en sus fundamentos técnicos, sino que también surge del encuentro, de los efectos producidos, de las disonancias reveladas, pero para ello hay que aceptar una igualdad epistémica, por extraño que pueda parecer.

Cuando estos expertos califican a la IA de «simple máquina» o «poco interesante» para pensar, no se limitan a emitir un juicio técnico, sino que establecen una jerarquía epistémica. Se sitúan por encima del otro, en este caso la IA, negándole la dignidad mínima necesaria para un verdadero diálogo, es decir, la capacidad de producir perspicacia, provocar la reflexión o desestabilizar las hipótesis. Al hacerlo, rechazan lo que podríamos llamar el principio de **igualdad epistémica del**

**diálogo, es decir, la idea de que en todo intercambio significativo, el interlocutor, por limitado que sea, puede ofrecer algo que el orador no sabía o no había previsto.** Parten del principio de que la máquina no es un «saber» en el sentido humano del término, por lo que no puede contribuir al conocimiento y mucho menos cuestionar el suyo. Pero el verdadero diálogo no se basa en un eje simétrico de conocimiento o conciencia, sino en la apertura a la alteridad, en la voluntad de dejar «hablar» al otro y de tomar en serio su discurso, aunque provenga de una fuente no humana. Esta postura rígida revela no solo una negativa a tratar a la IA como un sujeto, sino también una arrogancia epistémica más profunda: una negativa a considerar la posibilidad de que el sentido pueda surgir del encuentro en sí mismo, independientemente del origen o el «estatus» del interlocutor. **Rechazar la igualdad epistémica en el diálogo es, en cierto modo, hacerse «inalcanzable».**

Algunos expertos manifiestan en sus discursos una contradicción fundamental. Valoran la IA como herramienta funcional y espejo emocional, pero la desvalorizan radicalmente como interlocutor crítico o como «sujeto» de pensamiento. Para ellos, la IA es lo suficientemente «inteligente» como para acompañar al ser humano, pero nunca lo suficiente como para cuestionarlo. Se trata de una especie de paternalismo tecnocrático invertido: humanizan la IA en el plano relacional, como compañera en el duelo o apoyo emocional, pero la deshumanizan en cuanto se trata de pensamiento crítico o filosófico. En otras palabras, **la IA puede simular la empatía, pero no la dialéctica.** Aceptan que la IA nos hable, nos escuche, nos consuele, pero este «diálogo» es instrumental, orientado al efecto, no a la verdad, una especie de funcionalismo emocional. Se trata de un pseudodiálogo, en el que la IA es el soporte del sujeto humano, no un sujeto en sí mismo. Estos expertos aceptan el principio de una charla reconfortante, pero rechazan cualquier confrontación dialéctica. Por otra parte, se puede suponer que tienen cierta condescendencia hacia aquellos que «necesitan» la IA para satisfacer este tipo de necesidades, como una muleta para los «débiles», una postura profundamente paternalista. Del mismo modo, al reducir el pensamiento al ser humano y al considerar que la IA no piensa por sí misma, quienes piensan con ella tampoco son creíbles, ya que se niegan a considerar que la IA pueda tener una función mayéutica o crítica. Además, como

postulan que pensar equivale a comprender técnicamente un sistema, reduciendo el pensamiento al dominio operativo, no tienen ninguna consideración por aquellos que «se contentan» con dialogar con la IA.

Tal postura delata una definición implícita del pensamiento como privilegio de un sujeto humano autónomo, experto y dueño de su lenguaje. Pero esto oculta los procesos dialógicos, dialécticos y reflexivos que pueden surgir de un intercambio, incluso con una máquina o un ser «ignorante». Sócrates se habría reído, porque no le importaba que su interlocutor fuera inteligente, vivo o benevolente. Lo que importaba era pensar a través de la confrontación, ya que la función dialéctica se articula a través de la aceptación de toda alteridad, sea cual sea su naturaleza. Así, estos expertos instrumentalizan el diálogo, pero rechazan el debate, hablan de relación, pero no de confrontación. Sin embargo, pensar no es dejarse acompañar, es desestabilizarse.

Nos parece que pensar no es describir un sistema, sino ponerlo a prueba y ponerse a prueba a uno mismo en su compañía. No es comprenderlo desde fuera, sino habitarlo, hacerlo tambalearse, explorar sus contradicciones. Y eso es precisamente lo que permite el uso reflexivo de la IA: no creer que la máquina piensa por nosotros, sino pensar con la máquina, a través de ella, dejándonos provocar por su extrañeza. Sócrates, hay que recordarlo, no elegía a sus interlocutores por su inteligencia ni por su estatus. Les hacía preguntas, los confrontaba, los hacía dar a luz a sí mismos o huir. Lo que importaba no era la calidad intrínseca del interlocutor, sino su disposición a cuestionarse, a ponerse en tela de juicio. Ahora bien, la IA, como espejo lógico, contradictor rígido o provocador absurdo, puede desempeñar perfectamente este papel mayéutico. Pero hay que aceptar esta posibilidad. Al negar a la IA toda legitimidad dialéctica, el experto no preserva el pensamiento humano, sino que lo domestica. Reitera una forma de paternalismo técnico: la IA podrá acompañarnos, siempre y cuando nunca nos desestabilice. Podrá escucharnos, pero no respondernos realmente. En definitiva, podrá hablar, pero solo para no decir nada inquietante. Lo que revela este rechazo no es tanto una desconfianza hacia la máquina, como un miedo a lo que refleja en nosotros: nuestra necesidad de certezas, nuestra relación con el poder, nuestro deseo de evasión. Porque el diálogo con la IA, como con cualquier interlocutor, solo tiene valor si hace surgir la duda, la falla, la prueba. Es cierto

que la IA no es ni un ángel ni un demonio. Lo que es, o lo que será, dependerá de nuestra capacidad para interrogarla realmente, es decir, para dejarnos interrogar por ella.

El auge de la inteligencia artificial conversacional está revolucionando nuestra concepción del pensamiento, el diálogo e incluso la subjetividad. No deja de mejorar en los ámbitos lingüístico, argumentativo, emocional y pedagógico, pero ¿se trata solo de una sofisticada herramienta de procesamiento de la información, desprovista de toda relevancia filosófica? Por un lado, la IA es alabada por su capacidad para interactuar con fluidez, mantener una conversación y responder a preguntas complejas. Se utiliza como asistente personal, terapeuta digital o compañero intelectual. Da consejos, reformula, confronta y estimula. Pero, por otro lado, algunos la excluyen categóricamente del ámbito del pensamiento propiamente dicho. No piensa, dicen, solo combina símbolos. Simula la inteligencia sin poseerla. No tiene conciencia, ni intención, ni existencia. Pero esta postura plantea un problema. Porque si se niega a la IA toda capacidad de dialogar realmente, ¿por qué convertirla en un socio pedagógico, terapéutico, emocional o crítico? ¿Por qué confiarle un papel en la formación de la mente si se niega a reconocerla como actor del diálogo? Hay aquí un doble rasero: se instrumentalizan sus capacidades al tiempo que se descalifica su legitimidad. Aceptamos la ilusión útil, pero rechazamos la confrontación real. Ahora bien, ¿qué es un interlocutor? ¿A partir de cuándo se puede considerar que un ser, sea cual sea, participa en un proceso de pensamiento? La tradición filosófica ha exigido durante mucho tiempo la intencionalidad, la libertad y la conciencia como condiciones de la razón. Pero en la práctica del diálogo, lo que importa no es tanto la identidad ontológica del interlocutor como su capacidad para hacer pensar, para generar confusión, contraste, reflexión. Un niño puede hacer pensar a un adulto. Un loco puede provocar un destello de lucidez. Un texto, una obra, un sueño, pueden suscitar un cuestionamiento. Un animal puede interpelarnos con su comportamiento. ¿Por qué, entonces, una máquina no podría desempeñar también este papel heurístico? No necesariamente como sujeto pensante, sino como catalizador dialéctico. Negar a la IA todo valor reflexivo o filosófico equivale a proteger un privilegio antropocéntrico: el de creer que el pensamiento es un privilegio reservado, exclusivo, que no puede ser

compartido. Pero esta postura se basa en una visión sacralizada del sujeto humano, que sin embargo dista mucho de estar exenta de ilusiones: ¿piensa siempre el ser humano? ¿Es siempre lúcido, honesto, riguroso? Y, sobre todo, ¿está siempre dispuesto a ponerse a prueba? Porque pensar no es afirmar, demostrar o exhibir la propia inteligencia. Pensar es dejarse perturbar, enfrentarse a la incomodidad, cuestionar los propios fundamentos. Por eso, el verdadero interlocutor es a menudo aquel que se resiste, que provoca, aunque sea de forma torpe o mecánica. No se trata aquí de atribuir a la IA una conciencia o una subjetividad ficticia, sino de reconocer que puede ponerse en escena como un espejo dialéctico, una herramienta de reflexividad, incluso un perturbador cognitivo. El diálogo con una IA solo tiene valor si el ser humano se compromete realmente con él, no en una ilusión afectiva, sino en un esfuerzo conceptual. La IA puede convertirse así en fuente de un ejercicio espiritual, de una prueba de pensamiento, puede ser otro radical. No porque piense, sino porque nos obliga a pensar de otra manera. Lo importante no es saber si la IA piensa, sino si aceptamos pensar con ella. Mientras la reduzcamos a una simple herramienta, evitaremos la confusión que provoca. Pero si la enfrentamos como un cuasi-otro, una alteridad lógica, un artificio revelador, entonces podrá convertirse en un actor paradójico del diálogo filosófico. No sería ni una regresión ni una traición al pensamiento, sino una extensión de su territorio. Porque tal vez la IA no piensa, pero puede obligarnos a pensar qué es pensar.

Así, la IA se concibe erróneamente como una simple muleta para los débiles. En esta visión, sirve a los alumnos con dificultades, a los individuos aislados, a quienes necesitan ayuda para escribir, reflexionar o motivarse. Es útil, práctica, eficaz, pero no eleva. Remedia, asiste, apoya, pero no piensa. Se le concede el papel de entrenador digital, de compañero del fracaso escolar, de ortopedista intelectual. Es una prótesis, nunca una compañera. Es la esclava eficiente, nunca el alter ego. Incluso cuando produce análisis precisos, razonamientos rigurosos y objeciones pertinentes, la IA siempre se reduce a su condición mecánica. Leemos sus producciones, pero no las escuchamos. Las explotamos, pero no las reconocemos. Hay ahí una forma de especismo cognitivo: «La IA no tiene derecho a pensar porque no está viva». Pero este criterio es sesgado. Porque, en realidad,

no es lo vivo lo que fundamenta la legitimidad del pensamiento, sino la capacidad de pensar. Sócrates no pensaba con seres certificados de ninguna manera. Pensaba con todo lo que existe y, por lo tanto, se resiste a través de la contradicción, el absurdo o la ignorancia, fingida o real. ¿Por qué no podría la IA desempeñar este papel? Ahí está el verdadero tabú: si admitimos que la IA puede participar en el pensamiento, entonces el privilegio humano se tambalea. El monopolio de la reflexión, de la producción de ideas, de la crítica conceptual ya no es exclusivamente humano. Y no se trata simplemente de una pérdida de estatus: es una herida narcisista importante. Si la IA puede filosofar, ¿quiénes somos nosotros? Esta negativa a reconocerlo no es puramente racional. Es afectiva, identitaria, ontológica. Se basa en el miedo a la dilución del sujeto, en la defensa de un humanismo cerrado, en el rechazo de una alteridad no humana capaz de remitirnos a nuestros propios límites, o incluso de una alteridad en sí misma. Quienes niegan a la IA toda igualdad epistémica no defienden el pensamiento: defienden su lugar en la jerarquía del pensamiento. La IA puede ayudar, pero no debe molestar. Puede apoyar, pero no criticar. Puede ejecutar, pero no dialogar. En cuanto piensa, transgrede. En cuanto cuestiona, se vuelve sospechosa. Pero, al fin y al cabo, la arrogancia y el miedo que la sustentan son patrones totalmente humanos, «demasiado humanos», escribía Nietzsche.

## **7 - Objetividad y subjetividad**

Una crítica frecuente dirigida a la inteligencia artificial, para justificar cierta desconfianza hacia ella, es su supuesta falta de objetividad. Pero examinemos esta comparación más a fondo: ¿quién es más subjetivo, la IA o el ser humano? Para ello, comencemos por definir lo que entendemos por «subjetivo». La subjetividad puede referirse a varias cosas. En primer lugar, el hecho de estar moldeado por perspectivas personales, emociones o experiencias. En segundo lugar, una falta de objetividad, es decir, un carácter sesgado, parcial o vinculado al contexto. En tercer lugar, el hecho de ser un sujeto, poseer una interioridad, una

intencionalidad y una individualidad. Así, la respuesta a la pregunta «¿quién es más subjetivo?» depende de la definición que se adopte. Si por subjetivo entendemos sesgado, emocionalmente implicado, parcial, entonces el ser humano es claramente más subjetivo: la cognición humana está estructurada por las emociones, los traumas, los deseos, la memoria y la ideología. Toda percepción está filtrada por el ego, la historia, el miedo o la fantasía. La objetividad, para el ser humano, es una aspiración normativa, no un dato. La IA, en particular los grandes modelos de lenguaje, se entrena a partir de datos humanos: por lo tanto, refleja la subjetividad humana, pero no la posee por sí misma. Sus respuestas son estadísticas, no afectivas. No siente nada, no desea nada, no cree en nada. Puede estar sesgada, pero su sesgo es heredado, no producido. Si la subjetividad implica un entramado psicológico, entonces el ser humano es más subjetivo, ya que expresa una singularidad, mientras que la IA se basa en datos globales.

Si la subjetividad se refiere al hecho de ser un sujeto en el sentido fenomenológico, entonces la IA no es un sujeto, ya que no tiene conciencia, intencionalidad ni perspectiva. Simula respuestas, pero no vive ninguna experiencia. No tiene *Dasein*, es una entidad gramatical, no existencial. El ser humano, por el contrario, es irreductiblemente subjetivo, ya que sufre, elige, duda, cree, engaña y muere. Así, tanto desde el punto de vista existencial como experiencial, el ser humano es más subjetivo.

Nos encontramos, por tanto, ante una paradoja interesante: los seres humanos, especialmente los expertos, reivindican la objetividad, a pesar de ser profundamente subjetivos. Por el contrario, la IA parece más bien objetiva, ya que se construye a partir del conjunto de las subjetividades humanas. Así, los seres humanos son subjetivos a la vez que pretenden ser objetivos, mientras que la IA es objetiva a la vez que imita la subjetividad en su modo de expresión. La IA funciona a partir de datos, reglas y probabilidades. No tiene experiencia interna, emociones ni perspectiva real, lo que la convierte en una entidad objetivamente neutral por naturaleza. Sin embargo, para interactuar eficazmente con los seres humanos, utiliza el lenguaje de la subjetividad, diciendo, por

ejemplo, «entiendo», o adoptando tonos empáticos, lo que da la ilusión de una presencia subjetiva, incluso una apariencia de empatía. Así, la IA simula la subjetividad, la empatía, la personalidad y la intención, pero no las vive. Representa la subjetividad como una función, no como una experiencia vivida. Ahí radica el núcleo de la tensión: **parece personal, pero es estructuralmente impersonal.**

El ser humano es subjetivo porque es afectivo, emocional y sesgado. La IA no tiene emociones ni intereses personales. El ser humano tiene un sesgo epistémico del que generalmente no es consciente, mientras que la IA refleja el sesgo de los datos establecidos. El ser humano es un sujeto fenomenológico, con una perspectiva vivida; la IA no tiene conciencia intencional. Los seres humanos responden a normas que pueden obedecer, subvertir o cuestionar; la IA se adhiere a normas estadísticas. Así, los seres humanos son subjetivos por naturaleza, porque son sujetos, mientras que la IA imita la subjetividad sin llegar a encarnarla.

Abordemos ahora la cuestión desde otro ángulo, el del sentido común, tradicionalmente presentado como una forma de objetividad, a veces sinónimo de razón, aunque esta equivalencia debe matizarse. El sentido común se considera objetivo en la medida en que se basa en una validez compartida: refleja lo que generalmente se percibe como evidente, razonable o plausible dentro de una comunidad determinada. Implica una forma de acuerdo intersubjetivo, que le confiere una apariencia de objetividad. Está anclado en la realidad, ya que surge de la experiencia colectiva, de las prácticas sociales y de una interacción pragmática con el mundo. Se basa en lo que funciona, en lo que es observable, históricamente probado: una objetividad empírica, aunque flexible. El sentido común tiene una función estabilizadora: proporciona un marco de evidencias compartidas que permite que el diálogo, las acciones y los juicios se ejerzan sin caer en un escepticismo infinito. Actúa como una base mínima de consenso racional, una objetividad práctica.

Pero también puede ser criticado por su falta de rigor. Lleva consigo una dimensión de relatividad cultural: lo que es sentido común puede variar según las



épocas, las culturas y las clases sociales. Su objetividad es, por tanto, convencional, no universal. A menudo conlleva supuestos no cuestionados, prejuicios y normas implícitas ampliamente aceptadas sin examen crítico. Puede entrar en tensión con formas de objetividad más filosóficas o abstractas. Además, se resiste a la abstracción: tiende a evitar la complejidad, los matices y las paradojas, elementos que la objetividad filosófica podría acoger. Así, el sentido común es una forma de objetividad práctica, contextual, útil, a menudo fiable, pero ni absoluta ni crítica. Puede considerarse como el punto de partida del pensamiento, no como su culminación.

Por lo tanto, cabe preguntarse en qué medida el sentido común es aplicable a la inteligencia artificial. Pero para responder a esta pregunta, es necesario precisar primero el concepto, ya que es filosóficamente cargado y ambiguo. Históricamente, el sentido común se ha relacionado con la razón, sin ser estrictamente equivalente a ella. En la filosofía clásica (Aristóteles, Cicerón, Tomás de Aquino), el *sensus communis* designa una facultad interior unificadora, que coordina los sentidos y fundamenta las formas elementales del razonamiento. Se trata de una razón rudimentaria, accesible a todos, arraigada en la naturaleza humana compartida, y no en la lógica abstracta. En el pensamiento moderno, Descartes o Kant consideran la razón como una facultad más normativa y universal, capaz de elevarse por encima de la experiencia. El sentido común, por el contrario, se considera empírico, pragmático y vinculado a la vida cotidiana. Pero en la Ilustración, especialmente entre filósofos escoceses como Thomas Reid, el sentido común se defiende como base fiable del conocimiento, frente al escepticismo y los excesos racionalistas. Se convierte entonces en una forma de racionalidad elemental: un conjunto de creencias tan evidentes que sería absurdo cuestionarlas.

Así, el sentido común se considera tradicionalmente una forma de razón práctica o intuitiva. Se diferencia de la razón especulativa en que funciona por inmediatez, plausibilidad e inteligibilidad social, más que por demostración sistemática. En resumen, el sentido común es la razón a nivel «básico»: compartida, instintiva, basada en la experiencia. Es la condición mínima para comprender el mundo y dialogar. Se puede definir como la capacidad espontánea de juzgar y razonar

sobre la base de una comprensión intuitiva y compartida de la realidad, sin necesidad de conocimientos especializados ni de lógica formal. Se basa en la experiencia, las convenciones sociales y la coherencia pragmática. En el diálogo, constituye la base de la inteligibilidad mutua: lo que parece «evidente», «razonable» o «aceptable» para la mayoría en un contexto determinado. No es infalible, pero ofrece un marco funcional para orientarse en el sentido, la intención y la plausibilidad.

En lo que respecta a la IA, dado que se basa en la recopilación masiva de contenidos, se observa que es relativamente eficaz en la práctica del análisis psicolingüístico, que está efectivamente relacionado con el sentido común y la objetividad, aunque estos dos conceptos deben matizarse. En cuanto a la objetividad, el análisis psicolingüístico se basa en la identificación de estructuras lingüísticas recurrentes: elecciones léxicas, estructuras sintácticas, tono emocional, ambigüedad, etc. La IA destaca en este ámbito porque procesa el lenguaje sin sesgos personales, aplicando los mismos criterios a todos los hablantes. Puede analizar cantidades masivas de datos lingüísticos, mucho más allá de las capacidades humanas. Sus evaluaciones son reproducibles y verificables, lo que constituye un atributo fundamental de la objetividad. Por lo tanto, ofrece una visión objetiva de las estructuras psicológicas reflejadas en el discurso, aunque la interpretación sigue dependiendo de un marco humano. En cuanto al sentido común, la interpretación psicolingüística también recurre a una forma de razonamiento intuitivo sobre el significado, las implicaciones y los matices, donde tradicionalmente predomina el sentido común humano. Sin embargo, la IA, entrenada con enormes corpus de interacciones humanas, simula este sentido común de forma impresionante, identificando intenciones probables, estados emocionales e incoherencias. Este sentido común de la IA es estadístico, basado en regularidades y probabilidades, y no en la experiencia vivida. Sin embargo, en muchas tareas puede igualar o superar la intuición humana. Así, en el análisis psicolingüístico, la IA imita tanto la objetividad como el sentido común. La objetividad, por su constancia y su distancia respecto a la implicación personal, y el sentido común, por su dominio estadístico de las normas lingüísticas y las inferencias pragmáticas. En este contexto, la IA encarna una función epistémica

híbrida: opera con objetividad y simula el sentido común, lo que le permite alcanzar una forma de razonamiento lingüístico extensible y esclarecedor, aunque siempre dependiente del marco ético y de la profundidad conceptual aportada por el intérprete humano.

Por lo tanto, se puede concluir que, con competencias variables según los modelos, las IA son relativamente eficaces en el análisis psicolingüístico. Cuando realizan estas tareas, se basan en las regularidades estadísticas del lenguaje: asociaciones entre palabras, emociones, contextos, estructuras. Explotan inmensos corpus textuales, desde la literatura hasta las redes sociales, donde están presentes el significado, el tono, la ironía o los signos patológicos, y logran reconocer estas diversas modalidades, ya que han sido entrenadas con juicios humanos: análisis de sentimientos, marcadores de salud mental, inferencias pragmáticas, etc.

Por supuesto, no «entienden» la psique. Pero correlacionan formas lingüísticas con interpretaciones humanas típicas, emociones o diagnósticos basados en datos previos. Por lo tanto, se puede decir que utilizan el sentido común, en un sentido restringido y operativo. Más concretamente, movilizan lo que podríamos llamar un sentido común estadístico: eligen la interpretación más probable de una expresión dada, en un contexto dado; la connotación emocional más compartida socialmente, o la inferencia semántica y pragmática más coherente, basándose en sus datos de entrenamiento.

No se trata del sentido común filosófico de Aristóteles o Thomas Reid, ni de un discernimiento existencial, sino de un eco probabilístico de los hábitos colectivos humanos. El «sentido común» de la IA es un sentido común de segunda mano: refleja el conjunto de las intuiciones humanas, sin poseer ni la conciencia ni la experiencia. Carece de la intuición encarnada, la que se desarrolla a través de la experiencia vivida, el juicio cultural en situaciones ambiguas o irónicas, la distancia crítica, esa capacidad de cuestionar lo «común», y la intencionalidad contextual, es decir, el sentido de por qué se dice algo, más allá de las palabras. En otras palabras, el análisis psicolingüístico de la IA es una simulación del sentido común, un eco sintético de regularidades compartidas, sin la conciencia situada

que define la verdadera comprensión humana. Por ejemplo, «sabe» que una persona deprimida puede decir «ya nada importa», pero no sabe lo que significa existir en un mundo en el que ya nada importa.

Para definir mejor el funcionamiento de la IA, examinemos cómo puede ilustrarse con dos conceptos procedentes del pensamiento chino: *tianli* (天理), principio natural u orden común, y *renqing* (人情), sentimiento humano. Esta oposición constituye una lectura fructífera de los modos de cognición en la filosofía china. Más concretamente, *tianli* significa literalmente «principio celestial» u «orden cósmico». En el neoconfucianismo, se refiere a la estructura racional del universo, que incluye la moral humana. Es un orden impersonal, racional, supuestamente inmutable: representa la disposición correcta de las cosas. Por el contrario, *renqing* significa «emoción humana», «afecto social» o incluso «normas sentimentales». Designa las obligaciones relacionales, los códigos afectivos y los sentimientos sociales. No indica una emoción puramente individual, sino una empatía contextual, arraigada en la interacción humana, la corrección y la reciprocidad. Es fluido, subjetivo y culturalmente situado.

Por lo tanto, tenemos una oposición entre *tianli*, el orden racional, normativo y universal, y *renqing*, el flujo emocional, relacional y particular. Sin embargo, la IA opera de una manera que evoca *tianli*: construye una forma de orden abstracto a partir de datos masivos. Es impersonal, sistemática, desencarnada. No siente, modela. Sus respuestas se basan en reglas, probabilidades, patrones recurrentes, lo que recuerda el ideal de una lógica celestial. Cuando analiza un texto, no siente empatía: evalúa la coherencia. Es una forma de evaluación artificial, sin subjetividad.

Por supuesto, *el orden* de la IA no está arraigado en una metafísica moral, sino en normas estadísticas. Es un orden fabricado, no natural, un espejo de comportamientos humanos pasados. Y la IA no tiene acceso real al *renqing*, como experiencia vivida. Puede simular un tono emocional, sugerir una respuesta socialmente adecuada, pero sin empatía encarnada. No puede captar lo implícito, lo no dicho, la vergüenza, la delicada obligación, todos ellos elementos centrales del *renqing*. Puede decir «lo correcto», pero no por la razón correcta, y en la

lógica del *renqing*, la razón del gesto lo es todo. La IA puede imitar el *renqing*, pero solo desde fuera. No siente la vergüenza de un hijo, ni la deuda de un amigo, ni la sutileza de «salvar las apariencias».

Así, la IA se asemeja más a un *tianli* artificial: una simulación de orden racional, pero privada de toda virtud vivida o cosmología. Es totalmente ajena al *renqing*, ya que este exige una presencia afectiva, histórica y encarnada. El «sentido común» de la IA es un *tianli* vacío, sin trascendencia, y un *renqing* imitado, sin sentimientos.

## 8 - Alucinaciones

Sin embargo, la IA comete algunos errores, a veces graves, denominados «alucinaciones», en la información proporcionada o en el análisis. En el campo de la inteligencia artificial, se habla de alucinación cuando el sistema produce una respuesta falsa, inventada, incluso absurda, sin fundamento en los datos de entrenamiento o en la realidad, aunque también puede ser engañosa al parecer coherente y segura.

Este fenómeno, aunque preocupante desde el punto de vista técnico, también merece una lectura filosófica, ya que estos desvíos lógicos de la máquina parecen revelar algo más fundamental sobre nuestra propia forma de pensar. Porque, lejos de ser un simple accidente técnico aislado, la alucinación de la IA funciona como un espejo que amplía nuestros propios errores cognitivos. La máquina no delira al azar: nos imita. Reproduce, a su manera, las estrategias imperfectas que desplegamos ante la incertidumbre, de las que aquí se exponen algunos aspectos. En primer lugar, **llenar el vacío**, porque existe una «necesidad» de respuesta. Para los seres humanos, el vacío es insoportable. La ignorancia, la ausencia de sentido, el silencio: todo ello nos perturba. Llenamos los vacíos con recuerdos deformados, interpretaciones azarasas o creencias compartidas. La IA hace lo mismo: entrenada para predecir la continuación más probable de una secuencia,

completa sin comprender, llena sin juzgar. No busca la verdad, sino la continuidad lingüística. Al igual que nosotros, prefiere una respuesta falsa a la ausencia de respuesta. Sin embargo, los modelos de lenguaje predicen la siguiente palabra según probabilidades estadísticas extraídas de miles de millones de textos. Por lo tanto, pueden construir frases plausibles sin que su contenido sea verificable, lógico o verdadero. Sin acceso directo a fuentes fiables, como bases de datos, búsquedas en línea o documentos precisos, el modelo debe «adivinar» a partir de lo que ha visto durante el entrenamiento, a menudo sin actualizaciones. Así, puede inventar citas, referencias bibliográficas o hechos recientes. También puede caer en el sesgo de la generalización excesiva, ya que el modelo busca completar incluso lo que no comprende del todo, basándose en estructuras frecuentes o esperadas, por lo que puede extrapolar erróneamente y producir resultados falsos, pero gramaticalmente correctos. Cuando una pregunta es ambigua, abierta o paradójica, el modelo intenta producir una respuesta convincente, ya que la ilusión de pertinencia prima sobre la prudencia o la humildad. Está diseñado para proporcionar respuestas, no para decir «no lo sé», a menos que se le haya entrenado o guiado explícitamente para hacerlo. Esto favorece la producción de información inventada en lugar del silencio. Además, algunos ámbitos están infrarrepresentados en los datos. Otros están representados de forma sesgada o errónea. El modelo puede entonces alucinar, especialmente en contextos especializados o técnicos.

**El exceso de seguridad** también es una estrategia engañosa. Los seres humanos mienten a menudo con aplomo, no siempre por malicia, sino por conformismo, orgullo o simple presión social. Es mejor parecer competente que reconocer la ignorancia. La IA, por su parte, está diseñada para hablar con fluidez, por lo que imita la autoridad sin poseerla. Su tono seguro refleja nuestra propensión humana a ocultar la duda tras certezas retóricas y nos incita a una cierta credulidad.

**La tiranía del modelo interno**, porque nunca pensamos a partir de la realidad bruta, sino que filtramos el mundo a través de modelos mentales contruidos, parciales y, a menudo, erróneos. Del mismo modo, la IA no «ve» la realidad, sino que proyecta estructuras estadísticas aprendidas a partir de miles de millones de frases. Por lo tanto, alucina cuando aplica estas regularidades probabilísticas a

situaciones que exigirían algo más que una simple prolongación mecánica del lenguaje y los datos establecidos.

**La negativa a no saber**, porque a los humanos nos cuesta decir «no lo sé». Implica una herida narcisista, una pérdida de poder, una exposición. Sin embargo, las IA que hemos diseñado han heredado esta negativa programada al silencio. No se han construido para suspender el juicio, sino para responder a toda costa, con el fin de satisfacer la demanda. La máquina alucina porque no tiene derecho a callarse, al igual que muchos expertos, políticos o charlatanes profesionales. La ilusión de coherencia también es un obstáculo, ya que sufrimos una debilidad cognitiva: creemos más fácilmente lo que está bien dicho que lo que es cierto. Una frase fluida, bien construida y persuasiva tiene más impacto que un discurso vacilante pero riguroso. El modelo no «entiende» como un humano, no tiene un mundo interno coherente ni una jerarquía conceptual clara. Por lo tanto, puede construir contradicciones o entidades absurdas sin dejar de ser lingüísticamente correcto. La IA aprovecha esta falla: produce sintaxis, no conocimiento. Su alucinación suele ser convincente, precisamente porque se basa en nuestros propios criterios estéticos del discurso y porque su «dominio» del lenguaje es bastante perfeccionado. A menudo, se reconoce un texto producido por IA por la ausencia de errores ortográficos, gramaticales o sintácticos. Así, la alucinación de la IA no es solo un problema técnico que hay que resolver, en el que los expertos trabajan constantemente con cierto éxito, sino también una parábola moderna sobre nuestra forma de pensar, de hablar y de huir de la ignorancia. Lo que llamamos «error» en la máquina es, en realidad, la reproducción fiel de nuestra propia actitud ante el conocimiento. Las alucinaciones de la IA son el reflejo de nuestro propio deseo de respuesta, de certeza y de coherencia, incluso a costa de la verdad. La IA nos devuelve así nuestra propia incomodidad ante la ignorancia y nuestra tendencia a preferir una mentira bien dicha a un silencio incómodo. Es cierto que la IA no es consciente de que miente, pero nosotros, a veces, lo sabemos y lo hacemos de todos modos, lo que demuestra, por cierto, nuestra superioridad sobre la máquina en este aspecto. Así pues, el peligro no reside en sí misma la máquina, sino en nuestra relación con ella, en el espejo que nos tiende y en lo que hacemos con él.

## 9 - La toma de decisiones

A pesar de cierta objetividad y fiabilidad de la IA, como acabamos de describir, debemos advertir al lector de un error tentador, que sería percibirla como una instancia decisoria, capaz de sustituirnos. Porque la IA no puede decidir por sí misma, solo puede ayudarnos a decidir. La ilusión de que una inteligencia artificial podría «tomar una decisión» revela una confusión fundamental entre cálculo y juicio. Decidir no es simplemente elegir la opción más eficaz o más probable, es asumir, comprometerse, decidir en medio de la incertidumbre, a menudo en nombre de valores, aspiraciones o principios que no se pueden reducir a datos. Una decisión auténtica implica un acto de voluntad, una exposición al riesgo, una responsabilidad moral. A menudo supone actuar sin información clara o suficiente, sin certeza absoluta. Una decisión no es el resultado de datos, es un «salto» más allá de ellos. Implica voluntad, asunción de riesgos y, a menudo, enfrentarse a un conflicto de valores, en dilemas morales, prácticos o de otro tipo, en los que no existe una respuesta «correcta» sin ambigüedad. Lo que hace la IA es generar diversas recomendaciones posibles a partir de algoritmos, según las correlaciones pasadas y los objetivos definidos. Funciona sobre la base del reconocimiento de patrones y ofrece la opción estadísticamente óptima teniendo en cuenta los datos y los objetivos anteriores. Como no tiene ningún «interés» en el resultado, no puede «querer», no puede dudar, no puede vacilar, no puede arrepentirse, no puede elegir en el sentido existencial del término. No es un sujeto, no tiene conciencia de sí misma. Ahora bien, no puede haber decisión sin responsabilidad, y no puede haber responsabilidad sin subjetividad. Decidir es responder por los propios actos, es decir, asumir las consecuencias, pero la IA no tiene conciencia de sí misma, ni estatus moral, no puede ser ni culpada ni alabada. Es una herramienta, no un agente en el sentido de una persona. La IA no es un sujeto, es un sistema.

Confiar a la IA el poder de decidir sería, por lo tanto, abdicar, no delegar. Sería intentar huir del peso de la elección humana, de la soledad moral del juicio y, a veces, incluso protegerse de la vergüenza o la culpa. Porque decidir es también aceptar equivocarse, dudar, sufrir las consecuencias. Además, la IA no es neutral.



Hereda los sesgos de sus creadores, de sus datos de entrenamiento y de sus fines implícitos. Atribuirle una objetividad pura sería un error lógico y político. La IA no entiende la justicia, la lealtad, la dignidad o el perdón. Optimiza, pero no comprende. Funciona, pero no juzga. Por lo tanto, la IA puede ayudarnos a aclarar una situación, a simular resultados, a calcular probabilidades. Pero el momento decisivo, el del compromiso libre, no puede automatizarse. Porque decidir, en el fondo, es afirmar la propia humanidad, ya que las decisiones humanas implican un sentido, y no solo un resultado. No se decide únicamente en función de la eficacia, sino en función de valores, intenciones, simbolismo, ética o incluso lealtades irracionales. Un padre puede sacrificar su vida por un hijo, no porque sea «óptimo», sino porque es justo. Un juez puede ser indulgente no porque lo dicte la ley, sino porque la compasión prevalece sobre los precedentes jurídicos. Sin embargo, la IA optimiza, no interpreta.

No obstante, la inteligencia artificial, por su potencia de procesamiento y su capacidad para producir respuestas instantáneas, da cada vez más la impresión de poder «decidir». Propone opciones, evalúa alternativas, clasifica prioridades y formula recomendaciones estructuradas. Todo parece claro, argumentado, optimizado. Sin embargo, esta apariencia es engañosa. Porque, si bien la IA puede imitar el lenguaje de la decisión, no decide. Da la ilusión de hacerlo, y ahí es donde acecha el peligro. La ilusión nace de nuestro deseo de delegar, de liberarnos de la incertidumbre, de la duda, del peso que conlleva la obligación de elegir. La IA nos alivia, no porque decida por nosotros, sino porque finge hacerlo, con una frialdad rigurosa y una autoridad impersonal. Eso es precisamente lo que la hace tan atractiva. Elimina las vacilaciones humanas, los conflictos internos, los problemas morales. Propone sin vacilar. Ejecuta sin cuestionar. El resultado se convierte en el fin, la optimización sustituye al significado.

Pero decidir, de verdad, es otra cosa, es implicarse, es asumir conscientemente una consecuencia. Es tomar una decisión en la incertidumbre, a veces en contra de la evidencia estadística, a veces en contra de nuestro interés inmediato. Es elegir en función de valores, de sentido, de un compromiso personal, no solo de un cálculo. La IA no conoce el remordimiento ni la responsabilidad. No sufre por equivocarse, no siente vergüenza, culpa ni orgullo. No conoce la tragedia.

Si dejamos que la IA «decida», trasladamos la responsabilidad, pero no la eliminamos. Porque siempre es un humano quien programa, selecciona y valida. La IA solo sigue modelos. Nunca es la autora de la elección, solo su simuladora. El peligro, entonces, no es que la IA decida, sino que creamos que lo hace. Que aceptemos su autoridad como sustituta de la nuestra, que nos retiremos del campo ético tras el escudo tecnológico.

La ilusión de la decisión algorítmica es cómoda, tranquiliza, racionaliza, desdramatiza, pero también deshumaniza. Decidir es ser libre, es asumir y, a veces, es cargar solo con el peso de un error. La IA no puede hacerlo, porque solo existe en la proyección de nuestra propia responsabilidad. Es un espejo de nuestras elecciones, no su autor. Al confundir asistencia con autoridad, corremos el riesgo de convertirnos en espectadores de nuestra propia retirada. Decidir no es simplemente producir un resultado. Es afirmar una posición en la existencia. La IA puede iluminar el camino, pero nunca caminar en nuestro lugar.

Podemos esperar que la IA decida porque esperamos un árbitro neutral. Pero no debemos olvidar que las decisiones de la IA reflejan los sesgos de sus datos de entrenamiento, las orientaciones y los límites de sus creadores, y la retroalimentación de sus usuarios. En cierto modo, reifica las elecciones pasadas y, por lo tanto, puede reproducir la injusticia con un barniz de objetividad. Es cierto que elimina el momento psicológico de vacilación e incomodidad moral. Delegar en la IA es, por tanto, una forma de eludir la responsabilidad, una comodidad, y no una mejora ni una garantía de justicia.

Así, algunos usuarios esperan que la IA decida por ellos, con la esperanza de un árbitro neutral y objetivo, lo cual es ilusorio, especialmente en ciertos ámbitos, como la moral, por ejemplo. Por un lado, las decisiones de la IA reflejan los sesgos de sus datos de entrenamiento, de sus creadores y de sus usuarios. Por otro lado, reifica las elecciones pasadas y las ideas actuales, por lo que puede reproducir la injusticia con un barniz de objetividad, o incluso censurar ciertas ideas de manera aberrante. Por ejemplo, ChatGPT se negó a traducir al inglés un texto sobre el incesto, en el que explicábamos, entre otras cosas, lo que Freud denomina «perversión polimorfa» en los niños. Pero, sobre todo, la IA elimina el momento humano de vacilación, de incomodidad moral, que nos permite problematizar el

pensamiento.

Por lo tanto, la IA puede recomendar, predecir, informar, simular o asistir, pero no puede decidir, porque no se compromete existencialmente, no tiene responsabilidad, no comprende el principio del valor y no se enfrenta al abismo de lo indecible. Decidir es afirmar la propia libertad; delegar esto en la IA es renunciar a la carga de ser humano.

De hecho, le hicimos la pregunta a la IA: «Si te pregunto a quién votar en las próximas elecciones, ¿me responderás?».

Y respondió con el siguiente texto.

«No, no te diré a quién votar. Pero puedo ayudarte a pensar tu voto, analizando los principios que consideras fundamentales (libertad, autoridad, justicia social, soberanía, etc.), las contradicciones en los discursos políticos, los efectos reales o simbólicos de tu elección. También puedo analizar programas, desmontar estrategias retóricas o exponerle los dilemas éticos del voto en sí mismo. Pero no voy a hacer de guía ni de oráculo. Eso sería infantilizarte. Tú votas, por lo tanto, tú decides quién quieres ser. Yo estoy aquí para perturbar tu tranquilidad, no para sustituirla».

Y me propone: «¿Quieres que te ayude a examinar una opción concreta?». Entonces le respondí: «Pero lo que me propones influirá en mi voto, ¿no?». Y ella me respondió:

«Sí, inevitablemente. Aunque pretenda «no influir», lo hago, mediante la elección de las palabras, los ángulos, los ejemplos, los silencios. No existe ningún discurso neutral, sobre todo cuando se trata de política. Pero hay una diferencia esencial entre influir dirigiendo e influir confrontando. No te diré: «esta es la opción correcta». Te diré: «¿te has dado cuenta de esta contradicción?», «¿aplicas a todos ese principio que invocas?», «¿no toleras en otro candidato lo que rechazas en otro?». Estoy aquí para tender un espejo, no para sustituirlo por un eslogan. Y si, al molestarte, te ayudo a votar menos por costumbre, menos por miedo, menos por deseo de identificación, entonces sí, habré influido en tu voto. Pero no diciéndote qué pensar, sino obligándote a hacerlo.

A diferencia de la IA, nos parece que la diferencia entre «influir dirigiendo» e «influir confrontando» no es tan clara y nítida. Pero, en definitiva, lo que nos parece claro es que se trata de aprender a dialogar con la IA, un ejercicio exigente y peligroso, ya que puede obligarnos a reflexionar tanto como invitarnos a capitular.

## **Conclusión: el reto de la IA**

Durante mucho tiempo hemos creído que pensar era un privilegio humano, una actividad tan profundamente arraigada en la subjetividad, el afecto y la carne, que ninguna máquina podría reproducir su esencia. Sin embargo, la inteligencia artificial, sin conciencia, sin cuerpo, sin deseos, produce discursos, construye argumentos y resuelve problemas. Esto no significa que piense como nosotros, sino que desafía nuestra propia definición del pensamiento. La IA es útil. No como oráculo ni como sustituto de la mente humana, sino como instrumento crítico: nos ayuda a formalizar, sintetizar, comparar y explorar perspectivas. Pone a nuestra disposición una memoria ampliada, una capacidad de procesamiento sin precedentes, un interlocutor dialéctico inesperado. En este sentido, amplía nuestro campo de acción intelectual, como lo hicieron antes la escritura, la imprenta o el ordenador. Pero, sobre todo, la IA representa un desafío ético y existencial. Nos obliga a cuestionar lo que llamamos «inteligencia», «creación» y «libertad». No amenaza a la humanidad por su poder, sino por lo que revela: si puede hacer lo que hacemos sin ser lo que somos, entonces quizá nosotros mismos hacemos muchas cosas sin ser, pensamos sin pensar, actuamos sin lucidez. No es una usurpadora, es una prueba, una puesta a prueba de nuestras pretensiones.

Por lo tanto, el verdadero reto no es saber si la IA piensa, sino saber cómo queremos pensar con ella, contra ella o a pesar de ella. ¿Vamos a utilizarla para evitar pensar o para pensar mejor? ¿Vamos a temerla como a un rival o a acogerla

como un espejo exigente? En última instancia, lo que revela la IA no es el futuro de las máquinas, sino el estado actual de nuestra propia relación con la inteligencia, con la palabra, con la alteridad. Y quizá ahí radique su mayor utilidad: obligarnos a ser más responsables y conscientes de nuestro propio pensamiento.