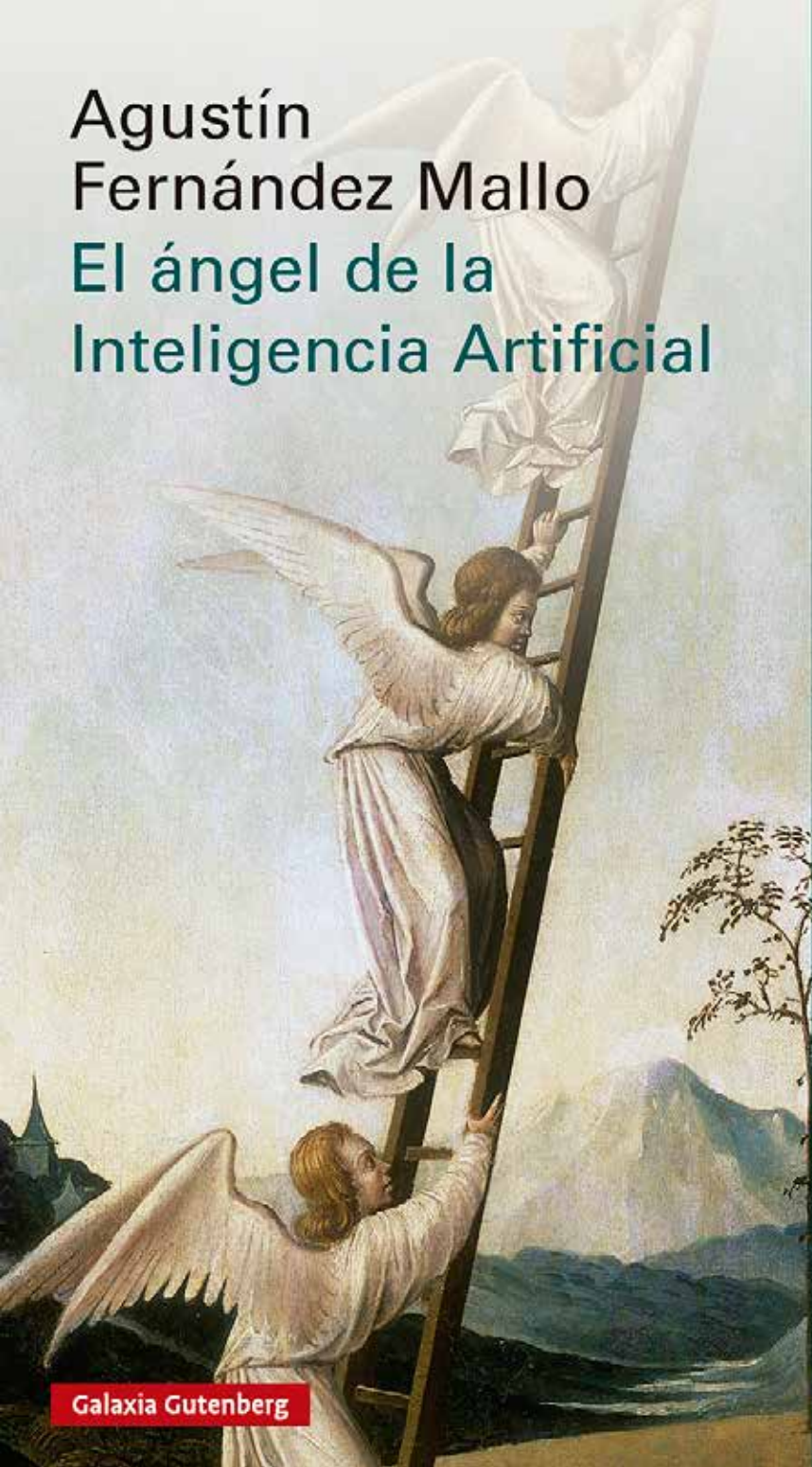


Agustín  
Fernández Mallo  
El ángel de la  
Inteligencia Artificial



---

Agustín Fernández Mallo

# El ángel de la Inteligencia Artificial

Galaxia Gutenberg

Publicado por  
Galaxia Gutenberg, S.L.  
Av. Diagonal, 361, 2.º 1.ª  
08037-Barcelona  
info@galaxiagutenberg.com  
www.galaxiagutenberg.com

Primera edición: marzo de 2026

© Agustín Fernández Mallo, 2026  
© Galaxia Gutenberg, S.L., 2026

Preimpresión: María García  
Impresión y encuadernación: Sagrafic  
Depósito legal: B 570-2026  
ISBN: 979-13-87605-91-9

Cualquier forma de reproducción, distribución, comunicación pública o transformación de esta obra sólo puede realizarse con la autorización de sus titulares, aparte de las excepciones previstas por la ley. Dirijase a CEDRO (Centro Español de Derechos Reprográficos) si necesita fotocopiar o escanear fragmentos de esta obra ([www.conlicencia.com](http://www.conlicencia.com); 91 702 19 70 / 93 272 04 45)

---

# Índice

## Parte I TEOLOGÍA

|                                                         |    |
|---------------------------------------------------------|----|
| Breve aproximación a la idea de ángel caído . . . . .   | 13 |
| Test de Turing (del ángel caído) . . . . .              | 19 |
| La creación de una inteligencia orgánica . . . . .      | 22 |
| La inmortalidad en la Inteligencia Artificial . . . . . | 27 |
| El acontecimiento atlético de la IA . . . . .           | 33 |
| La afinación . . . . .                                  | 38 |
| Identidad (1) . . . . .                                 | 46 |
| Identidad (2) . . . . .                                 | 52 |
| Identidad (3) . . . . .                                 | 54 |
| Inversión del punto de vista . . . . .                  | 57 |
| Rincón muerto . . . . .                                 | 60 |
| Nuevo <i>proceso</i> Pilatos-Jesucristo . . . . .       | 66 |
| De Pigmalión a Frankenstein . . . . .                   | 75 |
| El mito de lo andrógino . . . . .                       | 79 |
| Pieza única . . . . .                                   | 88 |
| Identidad (4) . . . . .                                 | 92 |
| El etcétera de las cosas . . . . .                      | 95 |
| El Gran Milagro . . . . .                               | 98 |

## Parte 2

### ANTROPOLOGÍA

|                                                             |     |
|-------------------------------------------------------------|-----|
| Historia de <i>d</i> , la distopía . . . . .                | 103 |
| El algoritmo-Estado, más allá de lo biopolítico . . . . .   | 109 |
| Instrucciones para construir una torre . . . . .            | 112 |
| De la máscara al código lingüístico . . . . .               | 116 |
| El lado abstracto de las cosas . . . . .                    | 121 |
| La tercera persona . . . . .                                | 125 |
| « <i>Yo no opero una imagen</i> » (la enfermedad) . . . . . | 134 |
| Quién puede trocear un fantasma . . . . .                   | 139 |
| El hipercubo humano . . . . .                               | 147 |
| El souvenir . . . . .                                       | 156 |
| Una realidad dentro de otra . . . . .                       | 159 |
| El testigo (1) . . . . .                                    | 166 |
| El testigo (2) . . . . .                                    | 171 |
| El testigo (3) . . . . .                                    | 174 |
| La Edad . . . . .                                           | 176 |

## Parte 3

### FÍSICA

|                                                        |     |
|--------------------------------------------------------|-----|
| Realidad, la imposible simultaneidad . . . . .         | 183 |
| 1 . . . . .                                            | 183 |
| 2 . . . . .                                            | 188 |
| 3 . . . . .                                            | 199 |
| 4 . . . . .                                            | 202 |
| 5 . . . . .                                            | 206 |
| La Edad de la Humanidad y la fantástica                |     |
| Singularidad . . . . .                                 | 210 |
| El perpetuo estado fetal (la antropogénesis) . . . . . | 213 |
| Presente imaginario, la sobreaceleración . . . . .     | 216 |
| Argumentos de plausibilidad para                       |     |
| la no-computabilidad de un cerebro . . . . .           | 225 |

|                                                         |     |
|---------------------------------------------------------|-----|
| 1. No existe cubo de la basura . . . . .                | 225 |
| 2. Teorema de Gödel . . . . .                           | 227 |
| 3. Autoconciencia. . . . .                              | 231 |
| 4. Lo consciente y lo inconsciente . . . . .            | 237 |
| 5. El loop, ruptura espontánea de la clausura . . . . . | 240 |
| 6. Lo extensivo y lo intensivo . . . . .                | 247 |
| 7. Re-conocer no es re-cordar. . . . .                  | 251 |
| 8. El ejemplo de los axiomas de Euclides. . . . .       | 254 |
| 9. Computación cuántica . . . . .                       | 256 |
| 10. Medios continuos, medios discretos . . . . .        | 259 |
| 11. Una conjetura. . . . .                              | 262 |

#### Parte 4 METAFÍSICA

|                                                                           |     |
|---------------------------------------------------------------------------|-----|
| Intenciones . . . . .                                                     | 265 |
| Nuevos seres, otra ontología . . . . .                                    | 266 |
| Contracultura . . . . .                                                   | 268 |
| Rebelión contra el mundo dado. . . . .                                    | 270 |
| Una metafísica blanda . . . . .                                           | 274 |
| Metafísica sexuada . . . . .                                              | 278 |
| Los dos tiempos. . . . .                                                  | 280 |
| Nomadismo y el canal paranormal . . . . .                                 | 283 |
| La nueva esencia . . . . .                                                | 286 |
| La exaptación . . . . .                                                   | 291 |
| El agujero infinito que separa a la teoría<br>de la experiencia . . . . . | 294 |
| Cuando de hipersujetos pasamos a hiposujetos (1) . . . .                  | 300 |
| Fetichismo . . . . .                                                      | 304 |
| Cuando de hipersujetos pasamos a hiposujetos (2) . . . .                  | 307 |
| Efectos sobre la traducción . . . . .                                     | 310 |
| El Test del Juicio . . . . .                                              | 312 |
| La insuperable asimetría . . . . .                                        | 315 |

## Coda

## MÁS ALLÁ DE LO CONOCIDO: PRINCIPIO

## HOLOGRÁFICO PARA UNA IA

*(Entropía, información, agujeros negros  
e inteligencia artificial)*

## Más allá de lo conocido: principio

|                                                              |     |
|--------------------------------------------------------------|-----|
| holográfico para una IA . . . . .                            | 317 |
| Información y entropía . . . . .                             | 319 |
| Calor . . . . .                                              | 325 |
| Información, entropía y agujeros negros . . . . .            | 329 |
| Principio Holográfico e Inteligencia Artificial . . . . .    | 333 |
| Información como magnitud primera . . . . .                  | 335 |
| El tiempo es creado por la ignorancia:                       |     |
| una nueva subjetividad. . . . .                              | 336 |
| Horizonte de sucesos de una Inteligencia Artificial. . . . . | 340 |
| Comentarios y agradecimientos. . . . .                       | 345 |
| Índice onomástico . . . . .                                  | 347 |

---

Pero, para empezar, el fantasma contra el que tenía que enfrentarse era él mismo.

*El extraño caso del Dr. Jekyll y Mr. Hyde,*  
R. L. STEVENSON



---

## PARTE 1

# Teología

### BREVE APROXIMACIÓN A LA IDEA DE ÁNGEL CAÍDO

La tecnología en general, y en particular la Inteligencia Artificial (IA de aquí en adelante) en cualesquiera de sus etapas y modalidades, constituye un ámbito del conocimiento y creación humanos que con una fuerza que sólo puede calificarse de primigenia se ha convertido en una fuente de mitos contemporáneos, miedos y místicas exaltaciones, cuando no en una suerte de neorreligión.

Nos proponemos ver el fenómeno Inteligencia Artificial como si fuera por primera vez, no atendiendo exactamente a sus específicos análisis tecnológicos –de los cuales ya existen magníficos textos analíticos y descriptivos– sino como un vasto territorio cuyas fronteras y límites comprenden dentro de sí a lo humano; no en vano, la IA, al menos hasta la fecha, está hecha por humanos, no por otra clase de seres. Para ello, nos valdremos de diferentes dimensiones, tales como teología, antropología, física y metafísica. Lo que entre todos estamos inventando es una ciencia todavía sin nombre, la cual, en mayor o menor medida, es motivo de fascinación en todo subconjunto social. Es por ello que hablamos de tentativas, ensayos y aproximaciones, no de certezas.

Se parte de un hecho ineludible: si ni siquiera sabemos qué es la inteligencia humana, todo intento de establecer en una máquina una idea de inteligencia que podamos comparar con la nuestra es netamente especulativo, cuando no, sencillamente, quimérico. De modo que nos estamos moviendo dentro de los lindes de una clase de poética –una «ciencia poética»–, la cual abarca factorizaciones tanto en la técnica de la computación como en las relaciones biológicas, en la antropología o en la física, en la teología o en la filosofía del lenguaje. Así las cosas, a fecha de hoy, hablar de la existencia de una *IA evolucionada* –en el sentido de una IA superior al cerebro humano, o incluso sencillamente equivalente– es, para empezar, dejarse llevar por la imposible tarea de medir lo inconmensurable e imaginar la fisonomía de lo todavía no sólo no visible sino formalmente inconcebible. Tanto es así, tan inmensa es esa imposibilidad, que, en múltiples terrenos de la actividad humana, paradójicamente les estamos pidiendo a las IA que hagan cosas anteriores a la existencia de la IA. Por ejemplo, en el campo del arte, cuando les pedimos que emitan –«pinten»– cuadros al modo en el que desde hace miles de años los humanos lo venimos haciendo. ¿Tiene sentido reclamarle a la IA que finja ser quien no es, que finja que ella misma no existe, que produzca objetos culturales anteriores a su naturaleza? ¿No sería más lógico plantear una nueva dimensión estética para la IA, algo que respondiese a su especificidad como el objeto inédito que es? El problema es que no sabemos cómo hacerlo, no tenemos ni idea de cómo plantearle eso a una IA. Al fin y al cabo, el humano es una caja negra, un cofre que no tiene una tecla *exit* para salir de sí mismo, de modo que no puede concebir la otredad absoluta que, de existir, de hecho sería una *IA evolucionada*.

Por otra parte, no pasemos por alto que se refuerza sin descanso la dimensión etnocéntrica. Oímos a diario hablar del modo en que la IA puede predecir y actuar por sí

sola en cuestiones que tienen que ver con, por ejemplo, la mercadotecnia, las ciencias, la política, las tendencias de mercado o la ordenación del territorio,<sup>1</sup> pero seguro que no tenemos noticias de la existencia de una IA que se ocupe de tratar de imitar los sistemas simbólico-cognitivos de culturas presentes en la Amazonía, ni los sistemas de predicción de lluvia y de cosechas fundamentados en las intuiciones y signos terrestres que desde hace siglos emplean los habitantes de multitud de comunidades agrícolas y ganaderas de nuestro entorno inmediato. Que estas predicciones y sistemas de conocimiento populares acierten o no, que sean objetivamente válidas o no, no es el tema aquí en cuestión, sino que esos modos de pensamiento, los cuales pueden ser calificados de extratecnológicos, también son producto de los cerebros humanos, también son inteligencia humana, en la misma medida en que lo es el sistema de coordenadas inventado por Descartes o el estudio de la proteína del azúcar. De modo que, para empezar, eso a lo que comúnmente llamamos IA, eso que supuestamente estamos modelando como inteligencia, no es una inteligencia en el sentido de la inteligencia humana, sino un producto tecnológico más, bautizado con una metáfora, y el cual, como toda tecnología, es usado según los propósitos de una cultura determinada, en este caso según el hábitat natural de la cosmovisión occidental y sus mercados de producción intelectual, conformándose como una versión más de una muy determinada hegemonía político-histórica. Si la IA fuera de verdad una IA estaría desarrollando ya espontáneamente toda clase de dimensiones de pensamiento humano, estaría abarcando toda clase de culturas y relatos; de nada de eso tenemos noticias. Que quien escribe esto comparta o

1. Existen magníficos textos en ese sentido, como el reciente *Una teoría crítica de la inteligencia artificial*, Innerarity, D., Galaxia Gutenberg, 2025.

no esos modos de conocimiento mítico-religioso, no propiamente científicos y no propiamente occidentales, nada importa para este debate.

Todo ello se agrava con el hecho cierto de que hasta hace muy pocos años las IA primitivas tomaban sus datos de informaciones introducidas en las bases por humanos, datos que de algún modo habían sido supervisados por una signatura de una verificación consensuada por una comunidad sapiens; hoy las IA, todavía primitivas pero de una generación más avanzada que aquellas primeras, cada vez con más frecuencia utilizan los datos que las propias IA tiempo atrás han por sí mismas generado, muchos de ellos dudosos, cuando no directamente erróneos, endogámico bucle que inevitablemente conduce al hecho de que las respuestas que dan las actuales IA sean más ricas en datos, sí, pero no más certeras. Típico caso de sistema que cuanto más crece en potencia, más crece también en atrofia. Les ocurrió a los dinosaurios, casi siempre también a los imperios.

El hecho cierto es que la IA no es más que la enésima versión de una clase de imposición de pensamiento mítico-religioso, llevado a cabo por un pequeño subconjunto cultural –el occidentalizado–, y que, no obstante, es planetariamente dominante. Así las cosas, lo que damos en llamar IA ni es inteligencia en el sentido estricto –equivalente a la humana–, ni es más ni menos artificial que el ser humano mismo, el cual, por definición, es un ser inseparable de su propia dimensión tecnológica y por lo tanto «artificial» en sí mismo. De hecho, vista de este modo, la IA nada tiene de artificial más allá de lo que lo puedan tener la piedra de sílex, la rueda o el fuego.

Advertimos al lector que en la medida de lo posible iremos apoyándonos en resultados, imágenes-símil y certezas de carácter cultural y antropológico; como ya hemos comentado, nos proponemos ver el fenómeno Inteligencia Artificial no atendiendo exactamente a sus

específicos análisis filo-tecnológicos sino como un vasto territorio cuyas fronteras y límites comprenden, como no puede ser de otro modo, dentro de sí a lo humano mismo.

Planteemos, pues, el núcleo del asunto en la dimensión que en esta primera sección nos interesa, la mítico-religiosa.

Contemplamos la Inteligencia Artificial con la misma fascinación, promesa e impotencia con la que los humanos primitivos contemplaban el fuego, cuyas propiedades factuales de supervivencia se mezclaban ya entonces sin demasiada distinción con el elemento mágico, mitológico e hipnótico que en su hacerse y deshacerse le es propio a las llamas. Hoy, ante esos cerebros artificiales que en un tiempo medido en días –a veces en horas– parecen avanzar hacia las murallas de todo lo considerado Imperio Humano, nos vemos impelidos a crear su cosmogonía, con todo lo que el nacimiento de un nuevo mundo arrastra consigo: fantasías y certezas, física y metafísica, dioses y ángeles, maravillas y –cómo no– componentes infernales. Si hasta hoy toda tecnología humana ha convocado sus propios mitos religiosos, la Inteligencia Artificial no iba a ser menos.

En la innumerable colección de seres y subseres disponibles en la tradición religiosa occidental –hegemónicamente judeocristiana–, y en cualesquiera de sus variantes, existen únicamente dos criaturas que tienen *voluntad propia*: los humanos y los ángeles caídos. Cuando un ángel todavía no ha caído –cuando se halla al servicio de su correspondiente dios–, carece de autonomía, se trata de un ser que existe únicamente para obrar en la Tierra la voluntad de la divinidad a la que por mandato se debe. Todos los ángeles existentes, y en todos sus tipos (de los querubines, a los serafines, de los malaj a los ofanim), no sólo tienen en común el cumplimiento de las órdenes de

su dios, sino que, además, y salvo contadas excepciones, la forma de llevar a cabo tales órdenes es no ser vistos por nosotros, los humanos, de modo que el trabajo que el ángel desempeñe en la Tierra será directamente atribuido a la divinidad, no propiamente al ángel que lo ha llevado a cabo. La figura del ángel es entonces una *herramienta intermediaria* que, tras realizar su tarea, desaparece sin haber sido detectada. Pensado en términos de la industria de los productos tecnodigitales, seguro que eso les suena. Los ángeles son *interfaces* ocultas, funcionales conexiones que permiten el intercambio de información entre dos sistemas, en este caso entre el sistema simbólico humano y el sagrado.

Pero la figura que aquí nos interesa es ese otro ser que, además del humano, tiene voluntad propia, aquel que puede operar al margen de la divinidad: el ángel caído, criatura que, en declarada rebeldía respecto a su dios y a otras jerarquías, campa entre nosotros a sus anchas restando no exactamente a la divinidad sino a la humanidad al completo. Nuestra voluntad en la Tierra, históricamente dotada del privilegio de ser universalmente única, entra de este modo en colisión con esa otra voluntad, la del ángel caído, que de pronto también se declara única. No hace falta estirar mucho el símil para percatarse de que en este caso nos encontramos ante una *Inteligencia Artificial evolucionada*, aquella máquina/ente que, en efecto, aparte del humano, algún día futuro podría tener *voluntad propia*, tomar decisiones al margen de nosotros y al margen de cualquier deidad.

La fuerza de este argumento tecno-religioso radica en que la *Inteligencia Artificial evolucionada* (de la cual, y respecto a la escala de tiempo de la evolución humana, todavía estamos viendo prehistóricos pasos), en sus modos de dar forma simbólica a nuestros entornos y hábitats no sigue únicamente el esquema metafórico del ángel caído de la teología cristiana, sino que podríamos

considerar innumerables religiones, monoteístas o politeístas, animistas o sincréticas, para advertir que todas ellas cuentan en su panteón de criaturas y subdioses con una figura que hace la función de ángel caído, ente divino o semidivino devenido en desgracia por osar tener voluntad propia, pero, y precisamente por tal motivo, dotado de una novedosa e inusitada fuerza cuando opera aquí, entre nosotros, en la Tierra. No en vano, podemos decir que las religiones siempre han sido una suerte de Inteligencias Artificiales: sistemas teológico-legales que, al igual que hoy ocurre con las IA, hemos creado para verter en ellas todas las tareas que no podemos o no sabemos llevar a cabo, dispositivos de deseo erigidos para que solucionen los problemas que por nosotros mismos no podemos resolver; de hecho, para gran parte de la gente la IA es una suerte de nuevo oráculo. Si las religiones, por derecho propio, constituyen las primeras IA, los actuales conatos de IA son máquinas que antes de ver la luz tal como hoy las entendemos han permanecido siglos en el útero de la Humanidad, incubando su materialización en una cavidad hecha de nuestros propios símbolos, metáforas y teológicos deseos jamás cumplidos.

Podemos añadir que a la tarea de intentar analizar esa incubación dedicamos las páginas que siguen, no sin antes aclarar que, en todo momento, por ángel caído no entendemos a ser diabólico alguno –tal como lo define la tradición cristiana–, sino, sencillamente, *un ente no humano que, dotado de voluntad propia, se desprende de la entidad creadora a la que se hallaba sometido*.

### TEST DE TURING (DEL ÁNGEL CAÍDO)

A cualquier divinidad se la ama en la misma medida en la que se la teme, y, a su vez, se la teme en la misma medida en la que se la odia, de ahí que todo individuo que

contemple en la IA una suerte de promesa de salvación, habrá un momento en el que, por ese elemental principio antropológico de inversión de pulsiones libidinales, pasará a ver esas máquinas como una verdadera amenaza, criaturas que, no siendo humanas, tienen libre albedrío y por lo tanto en cualquier momento podrían suplantarlos. De hecho, podemos afirmar que el hipotético día en que las IA sean realmente unas *IA evolucionadas*, es decir, sistemas cerebrales paralelos al nuestro y dotados de autonomía plena –lo que equivale a decir que tendrán sus propias leyes y sus consecuentes *derechos*<sup>2</sup> y, por lo tanto, no podremos desenchufarlas sin incurrir en un asesinato, lo que, a su vez, equivaldrá a decir que serán unas IA capaces de crear sus propios ecosistemas y relacionarse entre sí al margen de nosotros–, ese día, decimos, en caso de llegar, serán esos aparatos unos verdaderos y legítimos ángeles caídos. Y no nos referimos en este caso a un símil, ni a una metáfora ni a una imagen simbólica, serán realmente ángeles caídos materializados en tanto cumplirán todas sus características: 1) no serán humanos, 2) tendrán plena autonomía de voluntad, 3) como veremos más adelante, serán inmortales, y 4) carecerán de sexo; esta última característica será consecuencia directa de su inmortalidad, ya que, lógicamente, como seres inmortales que son, no necesitarán reproducirse. Aunque esta asexualidad veremos luego que tendrá matices, pues en el plano simbólico sí que las IA evolucionadas estarán sexuadas.

Más específicamente: tanto las computadoras clásicas como las actuales IA generativas –las cuales son IA primitivas, todavía dóciles y sencillos ángeles de la guar-

2. Sobre los derechos de autoría y derechos de las máquinas, han aparecido últimamente muchos e interesantes textos, por ejemplo *Observación de lo genuino*, Cabañeros Martínez, D., y Mallo Montoto, D., Eolas ediciones, 2024.



da, tecnología que está aquí para obedecernos y cuidarnos—, en ese hipotético escenario futuro habrán evolucionado a un estado de *caída en el tiempo*, y lo habrán hecho en todos los planos y magnitudes de la materialidad terrestre. El conocido Test de Turing, el cual buscaba la validación o no de una determinada máquina como ente «realmente humano», no era entonces un primer paso en la adquisición de identidad humana por parte de una máquina sino algo totalmente distinto, casi diríamos que opuesto: la prueba de su salto evolutivo a ángel caído.

Por otra parte, el Test de Turing introduce y pone de manifiesto una estructura psicológica y un comportamiento relacional humano mucho más profundo de lo que el propio test aparenta. Nos referimos al hecho mismo de la paranoia: en la medida en que tú, ese ser con el que estoy hablando, en todo momento e instante te ves obligado a demostrarme que no eres una máquina, lo cual no es difícil extender a nuestro ámbito cotidiano. Cuando un amigo, un familiar, un compañero de trabajo, nos habla desde el otro lado de nuestro smartphone, ¿es de verdad un humano? Cuando te cruzas por la calle con un conocido e intercambias unas palabras, de verdad es él o acaso se trata de un suplantador, un impostor, un alien que ha tomado la forma de mi amigo, etc. Todas esas disquisiciones y miedos paranoico-ficcionales, tan antiguos como nuestro cerebro mismo, se expresan de modo comprimido, en sustancia, en el propio planteamiento que opera en el Test de Turing.

Y si el verdadero problema —en realidad el problema nuclear— se hará presente cuando, como en su día le ocurrió al cerebro humano, las IA evolucionadas alcanzen una complejidad tal que espontáneamente se constituyan en *fenómenos emergentes complejos*, y por lo tanto en *sujetos de derecho*, tales máquinas nos devuelven desde ya mismo la pregunta que desde que pisamos la

Tierra nunca ha dejado de acecharnos, ¿qué es lo propiamente humano? Para responderla hemos de dar un breve rodeo, dejar a un lado la dimensión metafórico-simbólica y hablar de ciencia real, de cómo se crea la inteligencia realmente existente.

### LA CREACIÓN DE UNA INTELIGENCIA ORGÁNICA

Por todas partes vemos vida. Objetos y sujetos que en un momento dado de la evolución *emergieron*, se configuraron como entes autónomos y complejos que viven y se desarrollan en poblaciones autogestionadas, pero toda autogestión sería imposible sin la concurrencia de agentes externos; lo contrario equivaldría a extraer vida –energía, materia e información– de la nada. De dónde surge entonces eso que llamamos «realidad viva», cómo aparece vida inteligente, por qué aparece conciencia en el Universo –«¿por qué hay algo en lugar de nada?»–, se preguntó Leibniz en el siglo xvii.

Un hecho, primero respaldado por la teoría y luego ampliamente comprobado por la experiencia en laboratorios, es que todo sistema macroscópico es el resultado de una combinación de estados microscópicos cuya *función de onda cuántica* se ha colapsado. Expresado más concretamente: la función de onda que define la distribución de probabilidades de un sistema cuántico *se colapsa en lo real* –de ahí que, dicho sea de paso, la física cuántica sólo pueda aplicarse a objetos microscópicos; en la escala humana, la lógica cuántica no funciona, su aplicación carece de todo sentido, excepto como metáfora–. Es a partir de tal colapso, a partir de esa «caída en lo real», que, elevándose en varios órdenes de magnitud surge lo que llamamos *vida*, un «macrocolapso» en el cual la ya entonces complejísima –y en la práctica incalculable–

función de onda cuántica del sistema en curso libera energía y materia, disipaciones ambas que, al interaccionar con la materia blanda de su entorno, crearán *formas*, darán lugar a una determinada *morfogénesis*. Tal nueva y *emergente* materia blanda, al interaccionar a su vez con otros entornos almacena información y se produce entonces un segundo y fundamental proceso: todo ese almacenaje y codificación de información y de energía *reduce la entropía del sistema*. Sí, decimos bien, reduce la entropía. Concretamente: la entropía decrece en los sistemas abiertos, en los sistemas no aislados, que típicamente son los sistemas vivos. De modo que, tal como dice el Segundo Principio de la Termodinámica, la entropía total del Universo crece, sí, la suma de crecimiento de entropía Total es siempre mayor que cero porque suponemos que el Universo es un sistema aislado, pero en esas pequeñas zonas que son los sistemas abiertos –sistemas no aislados– la vida orgánica y la biología *reducen la entropía*. Esto es evidente por simple inspección ocular; qué duda cabe de que una flor, una hormiga o un cerebro humano son sistemas bastante ordenados (en absoluto desordenados), y el motivo es porque son organismos con una entropía muy baja. La consecuencia de esa disminución de entropía es que todos esos sistemas biológicos son sistemas *complejos* –en ocasiones a esa disminución de entropía se la denomina *neguentropía*, término que, por no dificultar la terminología, aquí no usaremos.

El modo en el que la biología consigue reducir la entropía –es decir, consigue crear vida–, es modificando su entorno, su medioambiente, guardando información de su hábitat y *moviendo el tiempo hacia delante*, lo cual nos indica que *si no se consume energía el tiempo no existe*, la vida no avanza y deja de tener sentido la palabra *vida*. La frase «moviendo el tiempo hacia delante» parece una redundancia, ¿acaso el tiempo no es siempre movimiento? No, hay veces que el tiempo no se mueve porque

no genera nada nuevo, y aunque las manecillas del reloj giren es como si el tiempo se hubiera detenido. Porque el tiempo no es el reloj, el reloj tan solo es un instrumento de divisiones arbitrarias, *el tiempo real son las huellas que los procesos dejan en la materia; es decir, el tiempo es –literalmente– información*. Si no hay producción de información, de huellas –y toda huella es información, que se lo digan si no a los detectives–, no hay tiempo real, no hay tiempo termodinámico, no hay fenómenos *irreversibles*. Todos esos movimientos de energía, materia y *creación de tiempo* son los que dan lugar a organismos complejos que al evolucionar pueden dotarse de conciencia, por elemental que ésta sea, y por lo tanto darán lugar a un lenguaje. Las células tienen un lenguaje, primitivo, pero al fin y al cabo un lenguaje que, a medida que se van construyendo agregados celulares dejará de ser un sencillo código comunicativo para convertirse en un lenguaje sujeto a las mutaciones propias de lo complejo.

Aquí entra en juego algo que poco tiene que ver con lo que acostumbramos a pensar cuando hablamos de IA, la *física biológica*. Lo que una rama de esta disciplina estudia, al contrario que la robótica o que las redes neuronales generativas –las cuales tratan de crear simulaciones de vida a través de sistemas digitales y algorítmicos–, es el modo de crear vida inteligente a partir de elementos orgánicos. De este modo, en los laboratorios, en recipientes preparados a tales efectos, son depositadas células a las que, a fin de observar sus cambios en cuanto a sus formas, su evolución y asociaciones con otras células, se las estimula inyectándoles energía e información. Todo ello para llegar a entender el origen de la inteligencia, así como las diferentes clases de inteligencia en la escala de los seres vivos.

Hemos llegado hasta aquí para poder decir que la primera consecuencia de todo ello es que el salto a la vida, la aparición de la inteligencia y de la conciencia, *no es digi-*

*tal*—o al menos no es únicamente digital—, necesita de una parte orgánica. Pero ¿qué queremos decir exactamente cuando decimos «una parte orgánica»? Pues queremos decir la existencia de algo que no pueda reducirse a ceros y unos, algo que no pueda ser codificado solamente en una secuencia algorítmica, algo que —en definitiva— contenga *ruido*. Porque no sólo la vida tiene ruido, sino que ocurre algo mucho más general: *no puede haber vida sin ruido*. Es el ruido, en el caso de la comunicación verbal —muy bien estudiado por la semiótica— lo que da lugar a errores, pero también a dobles sentidos y hallazgos metafóricos. En el campo que aquí nos interesa —la biología—, el ruido es la *temperatura*, aquello que agita la materia: el microscópico y aleatorio movimiento de los átomos y moléculas de un trozo de materia visible y macroscópica. Tal ruido es lo que provoca que, a su vez, existan mutaciones en las réplicas y en las dinámicas de cambios, mutaciones que son «fallos en la comunicación», los cuales, en ocasiones, generan una inesperada conexión entre unas partes que no estaban destinadas a relacionarse positivamente, de tal modo que ese fallo será un «fallo positivo», un «error positivo». Tal es el principio de la creatividad biológica y —dicho sea de paso—, también de la creatividad artística. Eso que llamamos artes, literatura, incluso ciencia, no es otra cosa que la suma de dos términos; a saber, *una copia más un error positivo*, el intento de copiar algo que, previamente, nos haya estimulado y, espontáneamente, copiarlo mal, establecer una mutación en el original de tal suerte que pueda *emerger* algo nuevo que una sociedad consensúe como artísticamente valioso.

Repetimos: no puede existir una forma de vida completamente autónoma, ni mucho menos un cerebro, que venga creado desde las prístinas instancias digitales del orbe exclusivamente matemático. No en vano, sabemos que en el plano práctico y experiencial, el mundo de lo digital está llegando a sus límites energéticos reales; casi

no se puede hacer más grande. Un sencillo ejemplo: a fecha de hoy, para sostener algo tan primario como el ChatGPT empleamos la misma energía por unidad de tiempo que para suministrar electricidad a una ciudad de 10.000 habitantes. Es decir, para llevar a cabo una actividad no ya similar a la del cerebro humano sino muchos órdenes de magnitud más sencilla, como lo es la actividad de ChatGPT, necesitamos un aporte energético miles de veces mayor que el que necesita un cerebro humano, el cual con tan sólo unas 2.000 kilocalorías diarias puede realizar tareas muchísimo más complejas que un ChatGPT. El motivo de semejante desfase –diría que desfase ontológico–, es el núcleo de la cuestión que estamos aquí tratando, y consiste en que el cerebro que mueve al ChatGPT es exclusivamente algorítmico y, por el contrario, el cerebro humano es materia orgánica que en su día a día y en sus problemas cotidianos –los cuales constituyen el 99% de los problemas a los que nos enfrentamos– funciona por un método de resolución totalmente distinto, denominado razonamiento *abductivo*, método específicamente humano e imposible de simular en una máquina. De hecho, tal método abductivo se fundamenta en «errores positivos» y en el ruido que nuestra percepción genera al interactuar con nuestro entorno (volveremos en extenso más adelante a la descripción de lo abductivo). De modo que, en todo el proceso de intentar crear cerebros humanos a partir de códigos informáticos cifrados en cualesquiera métodos algorítmicos, hay un momento –y quizá ese momento ya ha llegado– en el que es la propia realidad matérica, la propia termodinámica, la propia vida, la que nos lo revela como un imposible, como una fantasía teológica. Imposibilidad tanto energética –material– como filosófica –conceptual–. Apoyado en ideas del matemático Gödel, el primer ordenador moderno lo formula Turing para demostrar que en eso a lo que llamamos realidad hay una parte algorítmica, sí, y

hasta ahí todo funciona espectacularmente bien, pero ello nada dice ni desdice acerca de la parte no algorítmica de la realidad, de la parte analógica, orgánica. Pretender la obtención de una IA sin tener en cuenta esa zona orgánica de la realidad es, como mínimo, una actividad ficcional. Decimos más: cuantos más pasos se dan en esa dirección algorítmica, más nos alejamos del objetivo inicial.

Si en algún momento futuro una IA puede ser calificada de *IA evolucionada y compleja*, vendrá derivada de la biología, no de la informática; mejor dicho, de una combinación de ambas, porque ambas pertenecen a la realidad. Lo que sí es seguro es que el camino de construir ordenadores cada vez más potentes que puedan entrenar mejor a lo que damos en llamar redes neuronales generativas es un callejón sin salida. Resumido en términos físico-termodinámicos: la creación de una IA evolucionada ocurrirá cuando se consiga que ésta sea un sistema que –del mismo modo que lo hacemos los seres vivos– *su propia evolución reduzca su propia entropía*. De momento, todas las supuestas IA de las que disponemos lo único que hacen es aumentarla, y en caso de disminuirla lo hacen a costa de incrementar la entropía de su entorno –del Planeta– de un modo escandalosamente exponencial, tanto que a largo plazo las convierte en inviables.

## LA INMORTALIDAD EN LA INTELIGENCIA ARTIFICIAL

El problema, en términos antropológicos y de autoconciencia, se plantea del siguiente modo.

Desde el día de mi nacimiento, mi muerte, sin apresurarse, comenzó a caminar hacia mí.

La frase, atribuida a Jean Cocteau, en clave poética ejemplariza el hecho de que cada humano, por estar sometido al tiempo biohistórico y al tiempo termodinámico, al tiempo de la irreversibilidad, al tiempo de las huellas que las experiencias dejan en la materia, *tiene conciencia de muerte*, sabe que algún día se cruzará con la muerte; somos conscientes de que por muchos inputs que obtengamos, tales como energía a través de la alimentación o conocimiento y creatividad a través de la información, nuestra materialidad terminará por corromperse y nos moriremos. No así una IA algorítmica, a la cual, con tal de alimentarla energéticamente –por ejemplo conectada a una supuestamente inagotable red eléctrica–, y asimismo alimentarla con información –suministración de datos con los cuales por vía matemática los algoritmos en juego puedan continuar realizando modelizaciones del mundo–, nunca morirá. La IA, conceptualizada de este modo, es potencialmente inmortal.

No obstante, ¿podría una IA llegar a tener *conciencia* de esa inmortalidad que posee?, es decir, ¿podría una IA llegar a ser autoconsciente acerca del hecho de que ella misma es un ser potencialmente inmortal?

Lo más probable es que no. Expliquemos el motivo, sutil pero definitivo.

Lo único que crea en el cerebro la idea de experiencia vívida, y por lo tanto de *memoria compleja*, es decir, de anticipación y de *retroproyección* –también llamada *contrafactualidad*– de los hechos, es el reconocimiento del tiempo real. Cuando decimos *tiempo real* –lo hemos apuntado antes– no nos referimos al tiempo del reloj, cuya medida es arbitraria y, en última instancia, no es un *observable físico*, sino al tiempo que se percibe en el desgaste de la materia, al tiempo biohistórico, al tiempo que aumenta o disminuye la entropía de los sistemas, ese tiempo irreversible al que aquí llamamos *tiempo termodinámico*. Somos humanos porque *sabemos* que las huellas



que el entorno deja en nuestra materia corporal provocan que algún día nos muramos. Ese y no otro es el motivo por el que, en primera instancia, nos acoplamos reproductivamente con otros humanos, elemental instinto de supervivencia que nos lleva a intentar perpetuar nuestros genes. Una hipotética futura *IA algorítmica evolucionada*, dotada de un grado alto de sofisticación, podría llegar incluso a tener los datos suficientes como para saber que existen unos entes llamados seres humanos que, a diferencia de ella, por mucho que se les alimente con comida y con información, fenecerán, pero tal conocimiento del que estaría dotada esa IA tan sólo sería eso, una colección de datos, no una experiencia vivida por ella, de modo que el software neuronal generativo de la IA *no podrá experimentar cosas tales como la angustia o el gozo de saber que algún día, con probabilidad de certeza, morirá*.

Es decir, podrá *saberlo*, pero no podrá *experimentarlo*. No se configura como una experiencia biohistórica.

Esto, que parece más bien un resultado especulativo, cuando no meramente poético, tiene no obstante una implicación física real: si un ser no *experimenta* toda esa angustia, gozo, certeza y ruido, carece de *memoria compleja*. Dicho de otro modo: *sin cuerpo no hay memoria*. Tal es la primera y más básica diferencia entre la naturaleza humana y la de una IA algorítmica generativa, lo que las separa infinita y cualitativamente –ontológicamente–. Tal clase de IA podría incluso llegar a reproducirse, pero a diferencia del humano, no lo haría por una conciencia de muerte, ni por perpetuar unos genes, ni por conservar unos recuerdos importantes, ni por compartir símbolos o afectos o creencias con otros seres de su especie, sino sencillamente por inercia, por un principio que asiste a todo sistema complejo básico: todo lo que está ahí y disponible para ser utilizado, será utilizado como razón suficiente para ser gastado: la automática búsqueda de un hipotético estado de reposo o equili-

brio estable –en realidad nunca alcanzado– por medio del establecimiento de lazos con cuanto posea receptores físicamente adecuados para recibirlos.

Fue así, por esa inercia inherente a los sistemas complejos, cómo inicialmente se creó la vida en la Tierra: cuando una bacteria utilizó la luz solar para crear oxígeno a partir de la fotosíntesis, se produjo el salto, en apariencia imposible –de hecho, algorítmicamente imposible– de creación de algo orgánico (carbono) a partir de algo inorgánico (fotones, luz). A efectos de conciencia de mortalidad, la diferencia entre las IA algorítmicas y las plantas radica en que las plantas –y al igual que nosotros, que somos un producto evolucionado de aquellas plantas–, con total seguridad mueren como individuos, y para superar esa segura muerte la materia viva ha inventado el truco de la transmisión de genes en la reproducción: algo del individuo original será conducido a la siguiente generación. Es decir, una planta individual muere, un humano individual muere, pero ni las plantas ni los humanos como especie, como conjunto, tienen necesariamente por qué extinguirse. Sin embargo, una IA algorítmica evolucionada ni tan siquiera estaría obligada a morir como ser individual: con tal de ser alimentada podría no morir nunca, resultado este que se ve poco realista y que, obviamente, da lugar a toda clase de especulaciones mitológico-religiosas: las IA como dioses, oráculos, superhéroes, ángeles, etc.

A todo ello hay que añadir un obstáculo más. Los cerebros de cualesquiera seres vivos, y por muy diferentes que tales seres sean entre sí, tienen en común el hecho de ser analógicos, y por lo tanto dan lugar a lenguajes complejos, que trabajan no sólo con la lógica formal sino con el simbolismo derivado de las emociones. La IA actual opera todavía por medio de sistemas binarios digitales, combinaciones de ceros y unos. Esto podría cambiar en el momento en el que la computación cuán-

tica –la cual utilizaría muchos otros estados intermedios entre el cero y el uno–, pueda desarrollarse lo suficiente como para ser implementada de manera fiable y generalizada a la IA, momento en el que las actuales IA primitivas serían ya unas mejores candidatas a IA evolucionadas. Tal punto, no imposible, de momento se halla tan lejano como una ficción.

De modo que la IA, ya sea algorítmica o biológica, podrá considerarse autónoma cuando ella misma cree un sistema de autoabastecimiento de su propia sociedad y cultura, cuando todo cuanto haga se halle supeditado a mantener y hacer mutar el ecosistema propio de las IA, desprendido ya del ecosistema material y cultural humano. Lo que hoy llamamos IA es todavía un sistema algorítmico muy básico, que arroja resultados sujetos a una ilusión de creatividad, la cual no deja de ser como los trucos del mago, aparentes milagros, mecanismos que esconden la prestidigitación, la *ilusión* en el sentido literal de la palabra. Es posible que realicen tareas de predicción, pero ni de lejos alcanzan esa especificidad humana que consiste en, por ejemplo, *especular*, elaborar argumentos retrospectivos, retroversiones contrafactuales del tipo, «qué hubiera ocurrido si en vez de hacer esto hubiera hecho eso otro». Esto equivale a decir que sólo el humano tiene la capacidad de ponerse *después* de las cosas, fingir que unos hechos han concluido –fingir que unos hechos «han muerto»– para verlos desde ese otro lado de la muerte y especular «qué hubiera ocurrido si...». *La retroversión de las cosas pertenece al mundo de los genes, no de los algoritmos*, porque sólo el tiempo termodinámico –la flecha del tiempo– se halla dotado de la conciencia de lo *irreversible* para poder reconocer tal irreversibilidad y especular cómo hubieran sido las cosas si pudiéramos volver atrás. No pasemos por alto que tal mecanismo es la base de la fantasía, de la imaginación y de toda creación humana compleja.