“El peligro no es que nos vigilen, es que **hemos volcado todo nuestro ser en lugares que no conocemos**”

AGUSTÍN FERNÁNDEZ MALLO



Reputado novelista y ensayista, también agitador cultural, Agustín Fernández Mallo trabajó durante años como radiofísico en el Hospital Son Dureta de Palma de Mallorca.

Agustín Fernández Mallo es escritor y también físico. Como creador de ficciones ha escrito novelas y poemas, pero su último libro se inscribe en otra de las patas de su literatura, el ensayo. *El ángel de la Inteligencia Artificial* (Galaxia Gutenberg, 2026) reflexiona sobre esta nueva tecnología a partir de la metáfora de “un ser con voluntad propia, pero que no es humano”. Así, compara la IA con los ángeles caídos, “figuras intermedias entre los creadores y la Tierra, que caen y ganan en voluntad propia, pero están condenados a vagar por el mundo y hacer el mal”.

TEXTO: **CLAUDIA LORENZO** FOTOGRAFÍA: **IVÁN GIMÉNEZ**

as herramientas de inteligencia artificial nos cuidan, pero se nos dice que podría llegar un momento en el que dejen de hacerlo. Sin embargo, para Fernández Mallo la posibilidad de que se vuelvan más inteligentes que los humanos todavía es distante porque no han logrado, ni de lejos, replicar nuestra forma de pensar. Han aprendido cómo discurrir de forma inductiva o deductiva, pero todavía les queda mucho camino para alcanzar el pensamiento abductivo, esa forma de saltar a conclusiones de forma intuitiva característica de nuestra especie. Un “salto de fe” que practicamos constantemente y que es esencial en nuestra evolución.

Porque en su último libro Fernández Mallo empieza hablando de máquinas, pero termina diseccionando qué nos hace humanos.

Se plantea muchas preguntas en el ensayo, pero ¿cuál fue la pregunta primigenia que le impulsó a escribirlo? Quizá hay dos primigenias. La primera es por qué los humanos necesitamos siempre crear seres superiores a nosotros, algo que no tiene mucho sentido si lo pensamos en términos prácticos. ¿Por qué queremos crear algo que nos supere y luego someternos a ello? La segunda sería sencillamente qué estamos haciendo cuando decimos que creamos un supuesto cerebro artificial. ¿Realmente tiene algún sentido compararlo con el cerebro humano o no?

Estamos buscando una inteligencia artificial, pero no sabemos definir todavía qué es la inteligencia humana. Esa es una de las preguntas que me planteo. ¿Usamos la palabra inteligencia dándole la misma acepción que la humana? Por supuesto que estamos creando algo que tiene cierta inteligencia, en un sentido laxo. Pero el cerebro humano funciona con tan solo 25 W de potencia y, sin embargo, ChatGPT, cuando ▶

empecé el libro, necesitaba una energía equivalente a una ciudad de 10 000 habitantes. Ahí te das cuenta de que eso que llamamos inteligencia artificial no tiene nada que ver con el cerebro humano. De hecho, va en la dirección contraria: en vez de optimizar el gasto energético, lo que hace es maximizarlo.

¿Entonces? Para empezar, el concepto de inteligencia que se está publicitando respecto a las IA es occidental. Y además es meramente calculista. Es decir, si realmente la IA fuera una inteligencia humana, deberíamos hacer IA analfabetas, porque los analfabetos también tienen inteligencia y es humana. Y, de hecho, pueden desarrollar un tipo de inteligencia mucho más sutil porque precisamente tienen esa merma y tienen que adaptarse. Pero además, fuera del mundo occidental, ¿alguna compañía tecnológica está desarrollando una inteligencia artificial que emule el pensamiento de una comunidad indígena del Amazonas que tiene otros códigos simbólicos representativos? No. Lo que están creando es algo que calcula mucho más rápido y mejor que un humano. No quiero dar la impresión de que estoy en contra de la inteligencia artificial, porque es un avance increíble y nos soluciona la vida en muchísimas cosas. Pero hay que ponerla en el marco justo.

“Eso que llamamos inteligencia artificial no tiene nada que ver con el cerebro humano”

El otro día estaba en una cocina de un restaurante y pensaba que la IA no es capaz de pensar cómo estaban pensando esas personas, de organizar, priorizar, de decir “esto no me está saliendo, vamos a intentarlo de otra forma”. Y eso es inteligencia humana también. El humano es eso, esa cotidianidad. El humano se confunde. Cuando frío un huevo y me confundo, a lo mejor de ese error aparece un cuento y digo “¡Qué bien haberme confundido!”. Si una IA se equivoca dirá: “Voy a intentar freírlo mejor en la próxima”. A mí ya me dará igual, porque de ahí habrá salido ese delirio controlado que es toda creación. No tenemos ni idea de cómo



***El ángel de la Inteligencia Artificial* (Galaxia Gutenberg, 2026)** reflexiona desde un punto de vista humanístico sobre la IA comparándola con los ángeles caídos, pero también distanciándola de la inteligencia humana.

una IA puede empezar a desarrollar ese pensamiento abductivo.

Podríamos pensar que, cuando la IA “alucina”, sí que tiene un pensamiento abductivo, porque parece hacer “un salto de fe”. Sin embargo, no lo hace de forma consciente. Está inventándoselo. Es que estamos antropomorfizándolo todo. Alucinar es ya una palabra solamente aplicada al humano. La IA “alucina” por defecto, no por exceso. Pero eso no es alucinar, es que lo hace mal. Por eso un niño que hace garabatos no es Joan Miró. Joan Miró llega ahí tras toda una vida depurando un lenguaje. Pues con las IA pasa lo mismo, no es que alucinen, es que no saben hacer las cosas aún y hacen cosas mal. Y nosotros como humanos tendemos a antropomorfizarlas y decir: “Mira, es como cuando nosotros alucinamos”. No: el cerebro humano alucina porque es tan inteligente que se permite el lujo de alucinar. Y no pasa nada por ello. Si pasara algo, la propia evolución lo habría eliminado.

De hecho, si nos quisiésemos acercar más a la inteligencia humana a la hora de crear la inteligencia artificial evolucionada, tendríamos que estar investigando el cerebro. A través del mundo de lo algorítmico jamás se puede emular un cerebro humano, porque este funciona con muchas más cosas, con ruido, con errores, con una biología que el mundo algorítmico no puede contemplar. Habría que ir por la física biológica, pero eso es otro mundo.

¿Podríamos considerar que una inteligencia artificial sería equivalente al cerebro humano cuando tuviese conciencia de sí misma? Es una de las tesis del libro. Aunque realmente no creo que pueda llegar a tenerla. Para tener conciencia



de uno mismo, antes hay que tener conciencia de la muerte propia. Los humanos sabemos que nos vamos a morir de antemano; el tiempo deja huellas en nosotros, en nuestro cuerpo. Eso nos ha obligado a crear un lenguaje complejo para intentar explicarnos a nosotros mismos. Una máquina que es potencialmente inmortal, que no tiene cuerpo, no tiene conciencia de finitud. Al no tenerla, no habla a través de un yo.

Menciona en el libro que, cuando la inteligencia artificial tenga voluntad propia, no podremos desenchufarla porque será considerado un homicidio. Pero si nunca vamos a llegar a ese punto, ¿por qué nos obsesiona tanto qué podría pasar en ese caso? Porque somos humanos y siempre estamos ficcionando nuestro apocalipsis. Toda civilización tiene un relato de fin, proyectamos nuestra muerte personal en una muerte colectiva. Entonces, cuando aparecen este tipo de máquinas, objetos, entes, seres que parece que nos amenazan, tendemos a sublimarlo y a ficcionarlo. Eso es algo que también tiene que ver con lo que llamo el pensamiento contrafactual: imaginar qué hubiera ocurrido si hubiera pasado algo que no ocurrió. Eso es el germen de la fantasía y la esencia del relato, del arte. Los expertos no saben ni cómo empezar a programar una IA para que tenga un pensamiento contrafactual. Llevan siglos diciéndonos que el mundo se acaba y no se acaba. Claro que no se acaba, es que no se puede acabar. Todas las civilizaciones se extinguen y son reemplazadas por otras, pero eso de que en diez años la inteligencia artificial acabará con la humanidad es un relato para críos.

Todo lo que hemos leído sobre la posible amenaza se basa en las ficciones que hemos visto en las que aparecía la IA: *2001, Blade Runner*... Totalmente. Yo voy más atrás, a Frankenstein, al mito de Pigmalión, al gólem de los judíos. Es que está en la Biblia. El Apocalipsis es ese relato de miedo que está en los textos sagrados de la mayoría de las religiones, porque cuando le metes miedo a una persona, la tienes dominada para siempre. Pero también está en todos los relatos contemporáneos citados, y más. Esto de la IA para mí no es más que la enésima versión de una pulsión teológica que está en el ser humano, una tecnoteología. Hay quienes dicen que nos va a salvar como un dios, quienes dicen que va a ser el apocalipsis. Lo más razonable es pensar que ni una cosa ni la otra. Pero habrá que poner coto para que las empresas tecnológicas no puedan campar a sus anchas, porque puede ser muy perjudicial si no se controla.

“La IA ‘alucina’ por defecto, no por exceso. Pero eso no es alucinar, es que lo hace mal”

Había ganas de declarar “he creado este ente todopoderoso pero todavía tengo el poder de controlarlo”. De momento. Esa pulsión del humano todopoderoso está ahí siempre. Incluso aquí tiene un punto perverso porque seguramente esas empresas, si algún día ocurre un ▶

desastre de verdad, nos dirán: “Nosotros tenemos la solución y vais a tener que pagar por ella”. Por eso, los Estados deben legislar de una forma muy seria, ser realmente coercitivos, y aplicar penas importantes. Y es muy difícil porque se tienen que poner de acuerdo a un nivel global.

Parece que el relato nos dice que tenemos que tener miedo a la IA, pero en realidad le tenemos miedo al humano que la maneja. Justo es eso. Las empresas tecnológicas que están desarrollando el relato, porque efectivamente es un relato, nos dicen que las IA que ellos han creado ya van solas, que ellos no pueden hacer nada. Perdone, usted puede plantearse no soltarlas en este entorno porque lo van a destrozar. O soltarlas en un entorno controlado. Hay un desplazamiento de la responsabilidad por parte de las tecnológicas que es un relato netamente ficcional.



LA OTRA VIDA TAMBIÉN ES SUEÑO

Una pregunta algo más poética. ¿Lograremos que las IA sueñen?
¿En qué sentido? ¿El sueño nocturno o el sueño de un deseo?

Creo que cuando me lo planteé me refería al deseo.
No estoy seguro de que una IA, al no tener cuerpo, pueda tener un deseo, una pulsión libidinal de obtener algo y no sentirse nunca satisfecha. Aunque se puede plantear al revés. La IA en sí misma es un deseo continuo, porque nunca se detiene, no tiene freno, siempre quiere más. Por eso la IA nunca dice la verdad ni miente, porque no puede. Para mentir o decir la verdad tiene que haber algo en juego, y para ella no lo hay, así que contesta cualquier cosa. Es decir, se puede responder de las dos maneras: o es imposible porque, al no tener cuerpo, no pueden tener deseos; o, al contrario, están deseando siempre porque no hay nada que las detenga.

¿Y si hubiese dicho que me refería al sentido del sueño nocturno?
Esto lo veo más complejo porque habría que definir qué es para una IA estar dormida. Una IA está funcionando siempre o en *standby*, pero si está en reposo no hace nada, entonces no puede soñar. Creo que decir que sueñan supone antropomorfizarlas tanto como cuando decimos que alucinan. Es muy interesante pensarlo, pero desde un punto de vista alegórico.

Es cierto. Me resulta imposible abstraerme de los androides de la ciencia ficción que tengo en la cabeza y que le traslado a una IA que no existe.
Es que es imposible. Cuando los humanos hablamos de algo no podemos abstraernos de que somos humanos, porque si no ya no estaríamos hablando como estamos hablando, ni planteando las preguntas que estamos planteando. Entonces, miramos a las IA con ojos de humano. Pero eso es porque las hemos creado nosotros, evidentemente.

Fotografía: Nacho Rojas

La publicación del libro coincide con las advertencias de los grandes empresarios de la IA sobre sus peligros. ¿Qué se le pasó por la cabeza cuando empezó a escucharlos? Lo veo un poco cínico, es realmente interesado. Ellos no quieren dejar de obtener dinero y hay una pulsión de poder, de ver quién llega antes. Se retroalimentan para escalar hacia un lugar peligroso. Es una cuestión de responsabilidad, pero como yo no confío tanto en esa responsabilidad, o confío a un nivel educacional pero a largo plazo, en tanto no llega la educación hay que poner medidas legales coercitivas.

Ahora que somos conscientes de cómo funciona la IA, a base de datos masivos y personales, de vigilancia total, y vivimos en el panóptico digital, ¿cambiará esto nuestra forma de presentarnos ante el mundo? Es decir, ¿buscaremos ocultarnos para no ser información? Por supuesto. El control es total. Otra cosa es que nos importe o no. A mí personalmente no me importa mucho. Ahora bien, a lo mejor si estamos en una guerra me empieza a importar. Pero ya hay mucha gente que dice que no a la recogida de datos, que escribe a mano para no dejar ningún rastro o vuelve a las fotografías en papel. A mí lo que me da más miedo, francamente, no es el panóptico, porque al panóptico ya hemos llegado, no hay marcha atrás.

> “El concepto de inteligencia que se está publicitando con las IA es occidental y meramente calculista”

Entonces, ¿dónde está el miedo? En la educación. Los estudiantes ahora cuando quieren hacer algo le preguntan a la IA y esta les da la solución en un segundo. Pero educarse es aprender a pensar. Y solo hay una forma de aprender a pensar: haciendo el camino. Sin embargo, acudiendo a la IA para todo hemos llegado a lo que podemos llamar la externalización total del cerebro y del cuerpo. Nuestro cuerpo ya no está en nosotros, sino que está en unas empresas. El *summum* del panóptico ya no es que nos vigilen, es que nosotros mismos hayamos volcado todo en lugares que no conocemos. Ese es para mí el panóptico realmente peligroso, porque además de tenernos controlados, nos educa como él quiere.

Hay mucho debate ahora sobre si utilizar o no la IA en trabajos o labores artísticas o creativas. ¿Tiene sentido aproximarse a su uso desde un punto de vista moral o ético? Para empezar, es imposible no usar la IA porque la IA está en todo. Otra cosa es que no se use conscientemente. Yo, por ejemplo, no la uso para crear, pero no como una postura moral (descreo un poco de los discursos morales cuando alguien te alecciona). No la uso porque mi misión no es escribir muchos libros, sino hacer una investigación de mi propio mundo. Escribo para divertirme. Si lo hace la IA, ya no me divierto.

Todo lo que rodea a la IA provoca mucha preocupación en determinadas personas. La IA tiene dos novedades que no tienen las religiones ni los Estados. Por un lado, es la primera vez que construimos un objeto que mantiene con nosotros una conversación aparentemente compleja. Es aparente porque una IA habla pero no sabe lo que dice. Pero a nosotros nos da la sensación de que sí, y eso nos causa mucha inquietud y es lógico. Y, además, por primera vez existe algo en el planeta que, coaligándose colectivamente, puede llegar a crear sociedades que nos puedan retar y destruirnos. Eso nos da mucho miedo, aunque habrá que ver si es verdad o no.

Si la antropología está ahora reivindicando que el ser humano ha seguido adelante precisamente porque mayoritariamente ha resuelto los conflictos de forma no bélica, ¿por qué asumimos que la IA evolucionada va a solucionarlo todo de forma violenta? Francamente, no sé. Aunque hay otras teorías políticas, que hoy día están muy de moda en la extrema derecha y la extrema izquierda, contrarias a esa teoría, que defienden que el ser humano es conflicto y que al final toda vida es una lucha entre civilizaciones. A lo mejor no es que las IA quieran ser perniciosas, a lo mejor es que trabajan un poco a su aire, y eso puede perjudicarnos si no las dirigimos hacia donde queremos. Como la lava de un volcán que sale y es bonita y creará vegetación, pero para el humano que vive ahí es destructiva.

Panóptico digital