INTELIGENCIA ARTIFICIAL: PROMETEOS DE PAPEL Y FETICHISMO DE LA MERCANCÍA

Desde el software de los drones que sobrevuelan los cielos en Ucrania hasta los programas que rechazan currículums de forma automática, la llamada «Inteligencia Artificial» (IA) parece haberse instaurado, de la noche a la mañana, en todos los rincones de nuestra vida cotidiana. Nos recomienda series y llena de imágenes falsas las redes sociales, conduce coches autónomos, selecciona objetivos militares o responde dudas sobre bienestar psicológico. Si la cosa dependiese de los cuentos de la lechera producidos en masa por la mercadotecnia empresarial, pronto estaría presente hasta en nuestra tostadora. Sea como sea, la IA parece haber venido para quedarse y, junto a ella, un vertiginoso avance dirigido por, ni más ni menos, que los seres más ávidos de dinero existentes sobre la faz de la tierra: las juntas directivas de los conglomerados tecnológicos.

El campo de la Inteligencia Artificial lleva años siendo estudiado, y muchas de sus implementaciones llevan tiempo entre nosotros. Pero, recientemente –o eso dicen las empresas tecnológicas–, se ha desencadenado esa inminente «singularidad», ese Prometeo que vendrá a traernos el fuego de los dioses. Y es que aparentemente podríamos estar al borde de desarrollar esa «Inteligencia Artificial General» –o AGI– con una capacidad de realizar tareas superior a la de cualquier humano, incluyendo el pensamiento autónomo, el razonamiento y la adaptación a cualquier contexto. Depende de a quién preguntemos, la AGI vendría o a reemplazarnos o a elevarnos al siguiente nivel de sapiencia.

Como todo producto atravesado por las determinaciones de la producción capitalista, el fetichismo de la mercancía no es ajeno a la Inteligencia Artificial. En este caso ha alimentado a una plétora de interpretaciones entre aquellos que, cegados por las lecturas utópicas o distópicas, creen que estamos al borde de conseguir esa Inteligencia Artificial General, surgida de una película de ciencia ficción, que vendrá a solucionar todos nuestros problemas… o a destruirnos.

En otro extremo emerge otra lectura espontáneamente neoludita, que rechaza de plano cualquier uso de la IA, escudándose en argumentos de carácter pequeñoburgués. Desde esa perspectiva, se idealiza la figura del «artesano manual» o del «artista», como receptáculo de una creatividad pura e irreproducible, construyendo una posición maniquea del hombre frente a la «malvada máquina», ocultando un temor mucho más

tangible: que la automatización alcance, como tantas veces han hecho otras tecnologías, a profesiones intelectuales y creativas que parecían inmunes al avance del capital. Es justamente ahora que, para estos sectores, la tecnología es un problema moral al que combatir.

Este breve artículo pretende aplicar un análisis rigurosamente materialista sobre la IA, y esto requiere alejarse de las expectativas de un sector u otro de la población. Para ello debemos analizar primero qué es la Inteligencia Artificial –desmarcándonos de lo que los expertos en marketing quieren vendernos–, tanto en su funcionamiento lógico como la red de producción que la sustenta, pues todas estas formas de fetichismo ocultan una red de producción global y una ingente cantidad de trabajo humano bajo el espejismo de la IA, pero que en realidad nos muestra y nos advierte una vez más que la producción escapa vertiginosamente a nuestro control consciente.

El miedo o las falsas esperanzas han permeado el sentido común de la sociedad, la Inteligencia Artificial está hoy en boca de todo el mundo. Pero lo que circula bajo esa etiqueta está plagado de medias verdades cuando no de mentiras directas. La palabra «IA» se añade hoy a cualquier producto que suene remotamente a «ordenador que piensa», exactamente de la misma manera en que la etiqueta «orgánico» se añade a cualquier producto alimentario para que quede mejor en su estante en el supermercado, con independencia de que responda a un criterio técnico riguroso o a una simple estrategia comercial. Entender la IA realmente existente –no la que anuncian sus promotores ni la que temen sus detractores– exige, como mínimo, atravesar ambas capas para llegar a las relaciones sociales que la sostienen.

La nueva mente del emperador

Antes de intentar definir qué es la Inteligencia Artificial, conviene desprendernos de toda la retórica mercadotécnica con la que las empresas del sector envuelven sus productos. Términos como «conciencia» o «inteligencia» se usan con una ligereza que confunde más de lo que aclara, y esa confusión no es casual: cuanto más se sugiere que estas herramientas no solo «piensan» como nosotros, sino que pronto van a superar al ser humano, resulta más fácil colocarlas en el mercado. Pero para entender qué es realmente la IA hace falta analizar los conceptos con precisión.

El campo de la Inteligencia Artificial nació en los años 50 con el objetivo de simular la «inteligencia humana». Esa definición fundacional ya contiene parte del problema: «inteligencia» o «conciencia» son términos escurridizos, cargados por el dualismo psicofísico propio de una conceptualización burguesa de lo que es la conciencia, cosa que las grandes empresas del sector han sabido explotar para su propio beneficio.

La IA es una rama específica dentro del campo de la computación, es decir, de todo lo relacionado con el procesamiento, la transmisión y el almacenamiento de información digital. Esto incluye tanto el hardware, los circuitos y dispositivos por los que circula la información como el software, las instrucciones lógicas que le dan una estructura y un uso útil para un objetivo concreto. La IA, por lo tanto, es una aplicación particular de los mismos principios que rigen cualquier ordenador.

Dicho de otra forma y quitándole florituras, podríamos definirlo como el campo de la industria destinado al estudio e implementación de la computación con el objetivo de reproducir cierto tipo de actividad humana. Esto puede ir desde la generación de texto o imágenes –que es lo que fundamentalmente realizan las herramientas más conocidas, como ChatGPT–, hasta la toma de decisiones en tiempo real de los movimientos de un brazo robótico, pasando por la conducción de un coche autónomo.

En todos estos casos, el denominador común es que el sistema no sigue una secuencia invariable[1] como hacen tantos otros algoritmos, sino que produce un resultado a partir de un proceso de aprendizaje previo. Y aquí es donde la Inteligencia Artificial se diferencia de otro tipo de software, pues requiere de un paradigma específico, que no está basado en heurísticas fijas escritas de antemano, sino que modifica los resultados en base a su entrenamiento. El Aprendizaje Automático o *Machine Learning* es el término paraguas que incluye distintas metodologías para reforzar dicho aprendizaje.

El Aprendizaje Automático sirve, en última instancia, para que un modelo de IA pueda predecir resultados en base a buscar patrones en una cantidad masiva de datos de entrenamiento. Dentro de esta familia existen distintas variantes en las que el sistema ajusta su comportamiento en función de recompensas o penalizaciones asociadas al resultado obtenido a través del entrenamiento con cantidades masivas de datos de entrada con sus respectivos resultados de salida esperados, generando así un modelo de predicciones, que, durante la utilización del modelo, dará el resultado más probable en base a lo que ha «aprendido» en su entrenamiento.

La famosa «IA generativa» es una herramienta específica de sistemas de IA cuyo objetivo es generar contenido, ya sea texto, imágenes, vídeo, audio, código, etc., a partir de lo aprendido durante su entrenamiento. Se trata de generar un ordenamiento «nuevo» de palabras –o de pixeles, o cualquier otro contenido – estadísticamente derivado de los datos de entrenamiento. Es, con diferencia, la variante más popular y mediática de la IA

[1] Por «secuencia invariable» nos referimos a una operación lógica tal que: «si ocurre A, haz X; si ocurre B haz Y», y ese resultado X o Y se mantendrá invariable en el tiempo dependiendo de si la entrada es A o B.

actual, pues internet se halla inundado de texto, audio o imágenes generadas por herramientas de IA generativa.

Y aquí es donde se empiezan a embarrar verdaderamente los términos, porque se mezclan concepciones vulgares de lo que es la IA basadas en las herramientas de IA generativa con las predicciones grandilocuentes lanzadas por el CEO de turno en una entrevista. Y todavía hace falta especificar un término más. los modelos que más han dado para hablar y que más se asocian a la capacidad de una IA para «pensar»: los LLM o *Large Language Models*. Incluso el término «aprendizaje» también puede inducir a confusión, pues la IA no aprende como lo haría un animal, ni tampoco posee o experimenta recuerdos o experiencias de la misma forma en la que lo haría un ser humano.

Estos modelos de IA generativa están entrenados para generar principalmente texto o código. Un resumen de su funcionamiento sería el siguiente: cada palabra que se encuentra en los textos con los que se la entrena se convierte en un vector numérico compuesto de tokens. Por ejemplo, «La capital de Francia es» se descompone en:

[La] [capital] [de] [Francia] [es]

Un modelo de lenguaje maneja cientos de miles de tokens[2], –unidad que puede variar según el idioma–, representando todas las palabras –o fragmentos de palabras, signos de puntuación, etc.–. Cada token se convierte en un vector numérico que codifica su significado según el resto, según su posición o relevancia en la frase, etc. El conjunto de estos vectores forma una matriz numérica. A partir de aquí, a través de otro proceso, se aplican múltiples operaciones matriciales, obtenidas con los parámetros ajustados durante el entrenamiento. El resultado final es una distribución de probabilidad de cada token respecto al resto del vocabulario, un valor numérico que representa qué peso tiene cada token en la aparición de los siguientes en un texto, y eso para cada uno de los cientos de miles de ellos.

En la fase de ejecución, el sistema elige, de entre todas las opciones, la de mayor probabilidad según los datos de entrada del usuario. En este caso, si escribe la frase anterior, respondería con lo más probable –por ejemplo con «París»– completando la frase de forma coherente. La repetición de este proceso palabra por palabra es lo que permite generar frases completas con sentido semántico. En el fondo, lo que ha ocurrido es que el modelo ha generado un resultado estadístico; es decir, no ha encontrado el texto más correcto, sino el más probable, aunque en este caso coincidan. Conviene distinguir aquí dos fases claramente separadas. Durante la fase de entrenamiento, el sistema procesa

[2] El funcionamiento para la generación de imagen o vídeo es similar, pero en vez de palabras, el contenido a tokenizar es ya una matriz de pixeles, donde cada elemento es el color de cada píxel.

iterativamente enormes cantidades de datos y, en función del error entre lo que predice y lo que debería predecir, va ajustando un conjunto de parámetros numéricos. En la fase de uso, cuando ese modelo ya ha sido entrenado previamente, toma los datos de entrada que le proporciona un usuario, los transforma en una distribución de probabilidades como la descrita anteriormente y genera los datos de salida a partir del modelo.

Este proceso requiere de una enorme cantidad de operaciones matriciales, un tipo de operación matemática que resulta computacionalmente muy costosa cuando crece en número, al menos hasta que entraron en juego las GPU –unidades de procesamiento gráfico– que justamente están diseñadas para realizar muchas operaciones en paralelo de forma mucho más eficiente. Su diseño original estaba pensado para realizar cálculos para aplicaciones 3D –como los videojuegos–, pero demostraron ser verdaderamente útiles para realizar operaciones para otros fines, como el minado de criptomonedas o, ahora, el entrenamiento y la utilización de modelos de IA. No es casualidad que el auge de los grandes modelos de lenguaje (LLM) coincida con el salto en la potencia de las GPU de los últimos años. Sin ese avance en el hardware habrían sido inviables.

Pero la enorme capacidad de cálculo solo resuelve una parte del problema. No solo es necesario realizar miles de millones de operaciones en paralelo, sino que también lo es almacenar, procesar y transportar ingentes cantidades de información destinada al entrenamiento y, luego, una vez desplegado el modelo, éste debe atender a los miles de usuarios simultáneos. Ninguna máquina individual tiene una capacidad de almacenamiento o ancho de banda que le permitan asumir esta carga de trabajo.

Es por esto por lo que es imprescindible que este poder de computación se halle distribuido, lo que se denomina *Cloud Computing*. Los grandes proveedores de infraestructura juegan un papel muy relevante en este sector, pues distribuyen los recursos en enormes centros de datos, donde miles de GPUs operan de forma coordinada, pudiendo ampliar o reducir la utilización de recursos según la demanda. En este sentido, la revolución de la IA generativa es tanto un avance algorítmico como de la capacidad de cálculo disponible.

Los límites de la habitación china

En 1950, Alan Turing publicó «Maquinaria computacional e Inteligencia»[3] donde proponía que, siendo que términos como «pensar» o «conciencia» resultaban demasiado ambiguos para ser discutidos con rigor, su definición podía ser reemplazada por experimentos observables. Para ello formuló un ejercicio teórico en el que un interlocutor

[3] Alan M. Turing, "Computing Machinery and Intelligence," Mind LIX, n.º 236 (1950).

humano mantiene una conversación a ciegas –por ejemplo, a través de un chat o una hoja de papel– con una máquina y con una persona, sin saber cuál es cuál[4]. Si el interlocutor es incapaz de distinguir de forma clara entre ambos, entonces la máquina demuestra un «comportamiento inteligente»[5]. El giro empirista de Turing desplaza la pregunta de «qué era la conciencia» por un fenómeno observable. Esta propuesta daría lugar al que luego se conocería como el «Test de Turing». Como materialistas consecuentes sabemos que este ejercicio pragmático esconde un error fundamental, pues confunde una apariencia observable con la categoría que se quiere analizar. Pero su propuesta nos resulta útil, pues sentó las bases para el debate posterior.

Treinta años más tarde, el filósofo John Searle respondería a Turing con otro experimento mental al que denominó la «Habitación China»[6]. El escenario hipotético consiste en un hablante nativo de inglés que no posee conocimiento alguno de chino, encerrado en una habitación llena de cajas con sinogramas y un manual de instrucciones muy específicas que indican procedimientos como: «Si ves este símbolo, responde con este otro», permitiéndole ordenarlos de distintas formas para formar frases en chino. Un interlocutor, que sí sabe chino, situado fuera de la habitación, le envía frases a través de una ranura; siguiendo el manual, el inglés puede ordenar los símbolos respondiendo la frase del interlocutor a la perfección sin entender una palabra del mensaje que le llega.

La conclusión de Searle es que este mecanismo supera el test de Turing, pues el interlocutor cree que está hablando con alguien que sabe chino, pero el sujeto que se halla dentro de la habitación es incapaz de entender chino, simplemente sigue unas instrucciones.

La réplica más conocida es que, si bien es cierto que el hombre por sí solo no sabe chino, el «sistema», es decir el conjunto del hombre, el libro, los sinogramas y la habitación sí «saben» chino, siendo el operador sólo una parte del sistema completo. Searle respondería que basta con que el hombre memorice todo el manual para eliminar la distinción entre «hombre» y «sistema» para que el problema persista. Esto también suscitó críticas, pues si en este caso no se puede afirmar que el hombre sabe chino, tampoco se puede afirmar según la conducta de ninguna persona que sí conoce el idioma. La solución que propuso Turing terminó por reabrir el debate que él mismo había

[4] A este experimento teórico lo denominó «El juego de la imitación» y era una reformulación de otro ejercicio ya existente que consistía en que un interlocutor A entablaba conversaciones –por escrito– con una mujer B y un hombre C y debía averiguar quién era quién, teniendo en cuenta que tanto B como C debían intentar engañar al primero.

[5] El argumento empirista tiene una reducción al absurdo: si le pidiésemos a un modelo de LLM que simulase ser Alan Turing y perfeccionáramos esa simulación durante años, deberíamos, en algún punto, concluir que hemos generado la conciencia real de Alan Turing.

[6] John R. Searle, "Minds, Brains, and Programs," Behavioral and Brain Sciences 3, n. º 3 (1980): 417–424, https://doi.org/10.1017/s0140525x00005756.

intentado cerrar sobre la definición de conciencia, demostrando una vez más que la mera descripción de fenómenos superficiales no basta para poder conocer realmente el objeto que se está analizando.

Existe una segunda limitación para la «conciencia artificial», y esta encuentra su origen antes del primer argumento de Turing. En 1931, el matemático Kurt Gödel demostró que, en todo sistema formal capaz de expresar aritmética básica, existen proposiciones verdaderas que no pueden demostrarse utilizando únicamente las reglas de inferencia dentro del sistema y que, por lo tanto, se requiere de un «exceso» del modelo para aprehenderlas. Pocos años más tarde, el mismo Alan Turing demostró el «problema de la parada»: no puede existir un procedimiento universal capaz de determinar, para todo algoritmo, si este finalizará su ejecución o permanecerá funcionando indefinidamente buscando una solución. Es decir, ninguna computadora puede saber el comportamiento de todos los algoritmos, no por limitaciones tecnológicas sino por imposibilidad lógica.

A partir de la combinación de estos resultados, Roger Penrose[7] desarrolló el siguiente argumento: si la mente humana fuera equivalente a un sistema formal –es decir, si pensar fuera reducible a un algoritmo en la forma como los conocemos–, estaría sujeta a las mismas limitaciones descubiertas por Gödel y Turing. Sin embargo, Penrose sostiene que un humano sí puede reconocer la verdad de los enunciados de Gödel o de Turing, algo que, por definición, ningún sistema formal puede demostrar desde dentro de sí mismo. Eso indicaría que la mente humana no está limitada a un procedimiento algorítmico formal, o al menos, que las limitaciones de la conciencia humana no serían las mismas que las de la computación. La conclusión de Penrose es que, mientras la Inteligencia Artificial se mantenga dentro del paradigma actual de desarrollo, por definición, no puede ser consciente.

Con este recorrido histórico podemos volver al punto de partida: sabemos que una superación del Test de Turing no implica conciencia pues el mismo Turing estableció el criterio empirista como forma de esquivar la pregunta sobre la conciencia y no para resolverla.

En el caso de Searle y su Habitación China, el funcionamiento real de un modelo de lenguaje es muy similar a como funciona ese experimento mental. El modelo manipula símbolos –o tokens– según reglas aprendidas en su entrenamiento, generando un ordenamiento coherente de palabras, pero sin que exista una comprensión del contenido. El debate, pese a los avances en la computación, se halla estancado desde el siglo pasado

[7] Roger Penrose, La nueva mente del emperador, trad. Javier García Sanz (Madrid: Mondadori, 1991).

en las mismas herramientas conceptuales de Turing o Searle. Quienes afirman que un modelo como GPT ya es «consciente» suelen apoyarse, de forma inconsciente, en el mismo criterio empirista.

La perspectiva materialista nos permite ofrecer una respuesta definitiva al debate, y esta es que la conciencia no es una propiedad que pueda separarse de sus determinaciones. La conciencia, sea lo que sea, no es nunca «conciencia en abstracto». La pregunta no es cuánto poder computacional hace falta para que algo parezca consciente, sino qué tipo de relación establece con el mundo que le rodea, pues siempre se es consciente respecto «a algo». Es decir, la conciencia no es un estado interno que luego se proyecta hacia el resto del mundo, sino que es un momento del acto en sí del cuerpo relacionándose con el mundo a través de su actividad, es decir, la transformación consciente de la naturaleza, eso es, el trabajo. Un modelo de IA generativa, al que todos parecen evocarle las esperanzas para ese inminente despertar consciente, en realidad no tiene una intencionalidad en su generación de contenido –la transformación consciente del mundo– que incluye preguntarse para qué y cómo, sino que la generación siempre es obra de alguien externo al modelo, sea un programador o un usuario final.

Engels, en *El papel del trabajo en la transformación del mono al hombre*, explica que la conciencia no existe fuera de la corporalidad y, por lo tanto, no debe buscarse en factores externos a ella, en una singularidad en forma de alma o esencia, sino en la necesidad material de transformación de la naturaleza. La necesidad de actuar sobre el entorno mediante el trabajo es lo que, tras generaciones, terminó por desarrollar nuestros órganos como las manos o el cerebro. La respuesta a la pregunta no está en qué ingrediente algorítmico utilizar –o no únicamente– para que una placa de silicio suficientemente compleja pueda generar conciencia, sino que la conciencia es el resultado histórico de una relación práctica y física con el mundo.

Para Marx, el fetichismo consiste en atribuir a los productos del trabajo propiedades que en realidad pertenecen a relaciones sociales entre productores, ocultando el proceso que las engendra. Una extensión de este concepto a muchos de los fenómenos que rodean a la Inteligencia Artificial, y específicamente a los modelos de lenguaje, explica esa obsesión general con atribuirle a ChatGPT una conciencia que en realidad pertenece a todos los agentes sociales que han intervenido en su producción.

Antes de los modelos de lenguaje han existido muchas otras maquinas que han superado con creces alguna capacidad humana. Un telar mecánico teje ropa más rápido que cualquier persona; una grúa permite levantar peso imposible de mover siquiera por cualquier cuerpo humano. A nadie se le ocurre pensar que la grúa «es consciente» por levantar toneladas de acero. Ocurre que, como la IA generativa produce texto en vez de

camisetas o edificios, le atribuimos esa singularidad que ahora mismo no tiene, de nuevo, por la mera observación empírica de un fenómeno superficial que asociamos a la conciencia.

Un modelo de lenguaje calcula, con enorme eficiencia y dada una secuencia introducida de palabras, el resultado estadísticamente más probable aplicando patrones aprendidos en su entrenamiento. Es extremadamente veloz generando lenguaje coherente, pero no implica una relación semántica entre el símbolo generado y el concepto que representa. Para el materialismo dialéctico, poder establecer esa relación semántica con el objeto requiere de una relación práctica con este[8], en el caso de la Inteligencia Artificial esta relación siempre está mediada por la validación humana de qué inputs y outputs cuentan como correctos. Esto permite una respuesta distinta al argumento de Searle: el «sistema» de la Habitación China podría, en teoría, escribir chino si se amplía suficientemente el manual. Pero para el materialismo eso sigue sin ser suficiente, porque ningún manual, por completo que sea, sustituye la relación práctica de una comunidad de hablantes con el mundo.

La IA, que utiliza reglas inductivas para generar texto o imágenes, no solo carece de comprensión semántica, estando sujeta a límites lógicos formales, sino que también carece de agencia, entendida ésta desde unas coordenadas materialistas: la elección de qué problema resolver es indisociable de la solución misma. Un sistema puede ejecutar perfectamente los pasos lógico-matemáticos para resolver un problema, algo que otras máquinas, como las calculadoras, ya hacían mucho antes sin que en ningún momento haya habido una decisión de resolver ese problema en particular. El origen de esta agencia es indiferente a la vieja distinción entre lo biológico y lo social, distinción que, absolutizada, el materialismo considera antidialéctica[9]. Lo importante es que es a la vez producto y condición del trabajo, y es la actividad consciente lo que verdaderamente distingue al ser humano del resto de la naturaleza.

Siguiendo el curso lógico materialista, esa voluntad no está determinada por un «algoritmo biológico» interno al cerebro, sino por el entorno sobre el que actúa: la

[8] Con esto no queremos decir que siempre haga falta una relación práctica –en el sentido más intuitivo o superficial de la palabra– para aprehender un objeto. Una persona puede saber lo que es un martillo sin haber usado nunca uno. Aquí la cuestión radica en que la percepción en sí misma es una actividad, no una recepción pasiva de información. Pero esto es así porque se integran estímulos multimodales, con una única fuente de información nunca se podría desarrollar conciencia porque, como bien se menciona antes, la conciencia es relacional, es respecto a algo. En ese sentido, el hecho de que la IA opere «unimodalmente» mediante código es una limitación tremenda a su potencial capacidad de adquirir conciencia.

[9] La distinción entre lo biológico y lo social sí es procedente en algunas de sus dimensiones. Lo social es lo biológico transformado por la actividad del hombre y ese paso intermedio sí le confiere ciertas propiedades diferenciales que conviene tener en cuenta.

naturaleza, y, fruto del proceso histórico de transformación de esta, también el medio social en el que se desarrolla la actividad humana, es decir el modo de producción y las clases sociales que existen en él.

La consecuencia es que la diferencia entre inteligencia humana e Inteligencia Artificial no es cuantitativa –mayor o menor capacidad computacional o neuronal–, sino cualitativa. Esa «agencia» solo puede aparecer en el sistema como un sesgo introducido externamente por un ser humano –en el entrenamiento o en la petición del usuario– nunca como una propiedad intrínseca del modelo. El conocimiento no es una entidad esperando a «ser descubierta», es un producto de la actividad humana al transformar el mundo y al relacionarse de forma consciente con él. El procesamiento computacional y la inducción estadística pueden encontrar patrones estadísticos entre millones de datos y pueden combinar texto en forma de hipótesis sintácticamente coherentes en función de esos patrones[10], pero no tienen la capacidad física para demostrarlos ni de realizar avances científicos que, precisamente, requieren de un descubrimiento todavía no estadísticamente probable.

Habiendo repasado el debate histórico entorno a la cuestión, y habiendo puesto sobre la mesa la perspectiva consecuentemente materialista, en lugar de objetar categóricamente la posibilidad de una conciencia artificial o realizar ejercicios de futurología, resulta más útil señalar las limitaciones actuales para ese advenimiento de una Inteligencia Artificial General con consciencia:

Corporalidad: Si la única vía para la conciencia humana pasa por una relación física con el entorno, los límites actuales de la robótica, lejos todavía de emular un cuerpo suficientemente adaptable y capaz de relacionarse con el entorno como el de un ser humano, suponen un impedimento.

Una insuficiente comprensión del cerebro: Las redes neuronales artificiales se inspiran vagamente en la actividad cerebral humana para entrenar sus algoritmos, pero todavía carecemos de una comprensión suficientemente amplia de procesos cerebrales como para poder emularlos satisfactoriamente. Por ejemplo, la neuroplasticidad del cerebro y su división en zonas dedicadas está siendo compleja de simular, pues un modelo puede ser muy bueno generando texto y otro jugando al ajedrez, pero son extremadamente débiles al hacer otras tareas. La aproximación más cercana a crear zonas cerebrales

[10] La Inteligencia Artificial encuentra con precisión patrones en la estructura de las proteínas, acelerando la investigación biomédica; otros sistemas han ayudado a descubrir nuevos materiales o resultados en combinatoria matemática que antes no se conocían. En ningún momento queremos decir que la IA no puede contribuir, como herramienta, a avances científicos.

especializadas son los sistemas de agentes trabajando conjuntamente, pero todavía siguen siendo aproximaciones parciales.

La falta de conocimiento del mundo que nos rodea: Un modelo de Inteligencia Artificial puede volverse extraordinariamente bueno jugando al ajedrez o escribiendo texto, pero ni el ajedrez ni un texto son el mundo real en toda su complejidad. Las reglas inductivas tienen dificultades para adaptarse a un mundo cambiante que, actualmente, resulta imposible de modelar. Tenemos todavía un enorme desconocimiento de las reglas físicas que actúan en el universo. Por lo tanto, si la IA opera a partir de reglas de inferencia sobre datos ya existentes, superar las capacidades humanas exigiría anticipar fenómenos que todavía no han sido registrados por el ser humano, algo que nada tiene que ver con encontrar patrones de datos ya registrados.

Límites computacionales: La escalabilidad de los modelos actuales choca con los crecientes costes de energía, capacidad computacional y centros de datos, etc. La ampliación cuantitativa de los modelos actuales no puede escapar por fuerza bruta a las limitaciones lógicas que presenta la Inteligencia Artificial actual.

Nuestro objetivo, como ya hemos advertido, no es hacer un ejercicio filosófico de qué significaría si estas limitaciones se superasen, sino remarcar los límites por ahora infranqueables y exponer la concepción ideológica que afirma que sí existe esa conciencia en la IA. La idea de que basta con escalar el paradigma actual con más datos, más parámetros y más cómputo para alcanzar esa Inteligencia Artificial General es una tesis que, casualmente, interesa particularmente a los fabricantes de componentes, proveedores de centros de datos y empresas que venden los distintos modelos de IA. La pregunta que debemos hacernos ahora, entonces, es cuáles son las relaciones sociales realmente existentes detrás de la Inteligencia Artificial.

Los «Prometeos»

Una vez examinadas las limitaciones teóricas de la Inteligencia Artificial conviene repasar la red de producción real que hay detrás de este sector y que oculta una cantidad ingente de trabajo humano. Empezaremos por los protagonistas de su propio teatrillo, aquellos que están en el foco de todas las miradas, las principales empresas dedicadas a investigar y desarrollar los modelos como tal.

Algunas de ellas son ampliamente conocidas, siendo OpenAI, probablemente la más mediática por sus modelos GPT, orientados principalmente a la generación de texto y código; o DALL·E para la generación de imágenes. Ambas herramientas experimentaron una mejora drástica en los años 2022 y 2023, catapultando la empresa y

a su producto principal, ChatGPT, hacia un aumento de popularidad. A OpenAI pronto le siguieron otras empresas similares como Anthropic con modelo Claude, xAI con Grok, Google con Gemini o Meta con Llama. Tan pronto como sucedió la irrupción al mercado de estas grandes empresas, el interés del Gobierno de Estados Unidos por este sector como activo estratégico se avivó, tanto por su uso civil como militar.

La señal más clara de que Washington consideraba esta rama industrial como un activo de seguridad nacional se dio en agosto de 2025, cuando el Gobierno de EEUU invirtió, en dos partidas distintas, 8.900 millones de dólares en acciones ordinarias de Intel, fabricante estadunidense de circuitos integrados, semiconductores, y encargado del diseño y ensamblaje de procesadores. Las acciones equivalían aproximadamente al 10% de la compañía, y las dos operaciones formaban parte de la inversión en infraestructura base para capacidades computacionales fomentada por la Ley chips[11] y el programa *Secure Enclave* del Departamento de Defensa. A pesar de que la participación fue pasiva, es decir, no había representación gubernamental en el consejo, el mensaje que se le daba al mercado era que Washington consideraba a Intel un activo estratégico, cuya supervivencia y capacidad de fabricación eran prioritarias. Este movimiento se sumaba a otras grandes inversiones desde distintos fondos privados y bancos a empresas de componentes e Inteligencia Artificial tras la subida en picado de la popularidad de ChatGPT y similares.

Por su parte, Nvidia, principal fabricante mundial de GPUs, –que ya hemos visto que son un componente clave para entrenar y ejecutar modelos de Inteligencia Artificial– al ver incrementada la demanda de su producto, no se limitó a incrementar la venta hardware, sino que también se convirtió en inversor de sus propios clientes, una práctica compartida, como veremos, por varias empresas del sector, tejiendo una red de acuerdos que conviene ordenar.

Nvidia, tras participar en una ronda de financiación de OpenAI junto con otros inversores en 2024, anunciaría en septiembre de 2025 un acuerdo de intenciones extremadamente ambicioso con la empresa propietaria de ChatGPT, la cual empezaría por desplegar, al menos, 10 GW en sistemas GPUs financiados por la misma Nvidia, y Nvidia iría incrementando su inversión hasta los 100 mil millones de forma progresiva en función del crecimiento de ambos negocios. Nvidia prestaba el dinero, y ese dinero tendría que volver a Nvidia en forma de pedidos de hardware, mientras OpenAI vendería su novedoso producto embadurnado de capas y capas de marketing. El anuncio disparó las acciones de Nvidia y OpenAI sin que se hubiera producido nada con lo que recuperar

[11] Oficialmente llamada «*Creating Helpful Incentives to Produce Semiconductors and Science Act of 2022*».

la inversión. De hecho, por no haber, no se había desembolsado por completo el dinero acordado. Lo cierto es que la supuesta inversión de 100.000 millones nunca fue un acuerdo cerrado, así que, para marzo de 2026, Jensen Huang, CEO de Nvidia, confirmó que los 100.000 millones eran una «cifra máxima aproximada». Lo que finalmente se formalizó fue una participación de 30.000 millones. A pesar de que la cifra real era elevada, no correspondía a las expectativas del mercado de valores, pero esto no frenó el frenesí inversor.

Ese mismo septiembre, tras la inversión de EEUU, Nvidia compró acciones de Intel por 5.000 millones de dólares, junto a la propuesta de combinar sus equipos destinados a los centros de datos en una especie de alianza estratégica en el sector. En paralelo, Nvidia acordó también comprar 6.300 millones en infraestructura Cloud a CoreWave –una empresa americana de computación en la nube, que, como también hemos visto, es otra parte fundamental del sector– y en 2026 invirtió 2.000 millones en acciones en la misma empresa, asegurándose así de que CoreWave comprara también sus chips. CoreWave también fue contratada por OpenAI para garantizar infraestructura para entrenar los modelos OpenAI, un contrato que ascendió a 22,4 mil millones de dólares. También en septiembre de 2025 OpenAI acordó un contrato de hasta 300.000 millones con Oracle –otra empresa de infraestructura *cloud*– para poder desplegar todavía más centros de datos. Estas situaciones ejemplifican la mecánica establecida: Nvidia toma una participación en otras empresas actuando como garantía de demanda y luego registra cifras récord de ventas de chips, OpenAI muestra compromisos masivos con varias empresas en infraestructura con la promesa de que su modelo va a recuperar las inversiones, y las empresas de computación *cloud* –financiadas en parte también por Nvidia– presumen de tener millones de centros de datos pendientes de desplegar.

En octubre de 2025, AMD –el competidor principal de Nvidia– emitió formalmente un derecho de compra a OpenAI de 160 millones de acciones como parte de un acuerdo para suministrar GPUs a OpenAI. Si OpenAI ejerce el acuerdo completo, podría llegar a poseer cerca del 10% de AMD. Es, de nuevo, la misma lógica: en lugar de que OpenAI pague en líquido por el hardware, AMD le entrega una participación en su propia empresa como incentivo para que ambas cuentas de resultados y cotizaciones bursátiles incrementen de forma simultánea sin que haya circulado dinero real entre ellas.

Para añadir más leña al fuego, OpenAI realizó un gasto de 250.000 millones en servicios *cloud* con Microsoft, acordando desplegar chips de AMD –siendo ahora OpenAI uno de sus mayores accionistas–, todo esto mientras OpenAI y Oracle cerraban el acuerdo de 300.000 millones y Oracle compraba chips a Nvidia por valor de 40 mil millones.

El patrón se repitió con Anthropic, fundada por antiguos investigadores y cargos de OpenAI, donde los proveedores de infraestructura invierten en la empresa que produce el modelo y ésta se compromete a comprar su infraestructura. Amazon invirtió en 2026 varios miles de millones bajo el compromiso de que los creadores de Claude aseguraran la compra de 100 mil millones durante los próximos años en Amazon Web Services –el servicio *cloud* que provee Amazon–, asegurando 5GW de capacidad computacional para entrenar sus modelos con chips de Amazon. En paralelo, Google, accionista de Anthropic, aumentaba su inversión al mismo tiempo que Anthropic firmaba acuerdos para ampliar la infraestructura de Google Cloud.

También en octubre de 2025, xAI, a pesar de ser un componente del conglomerado de Elon Musk y presentar sus relaciones específicas con el Gobierno de EEUU y con el resto del sector, no quería quedarse atrás y cerró una ronda de financiación de 20.000 millones de dólares a cambio de comprar GPUs de Nvidia durante la construcción de sus centros de datos.

La lógica común de todos estos acuerdos que han inflado el precio de las acciones de todas estas empresas –con OpenAI y Nvidia en el centro del meollo– es que se trata de una suma de promesas que redundan en una inversión circular. El capital no retorna al sistema desde fuera con venta al consumidor o demanda real, sino que la mayor parte, por ahora, circula entre los mismos actores, inflando las cuentas de resultados de todos los participantes simultáneamente. Cuando una empresa de hardware o infraestructura anuncia que invertirá en un modelo y la empresa propietaria del modelo anuncia que comprará infraestructura de la otra, ambas empresas registran un «compromiso» que hace subir sus valoraciones bursátiles, aunque el dinero no haya cambiado de manos. Cuantos más activos se atan a la rentabilidad a futuro del sector, más cuerpo toma la narrativa de que la IA está al borde de tomar conciencia, sustituyéndonos a todos y generando enormes ganancias no se sabe muy bien cómo. Pero todos quieren subirse a ese barco y los gigantes tecnológicos deben comprometerse cada vez más, porque ahora ya no pueden retirar su capital y quien salga ganando podrá dominar este mercado, generando una burbuja siempre al borde de estallar.

Por último, debemos tener en cuenta también que la competencia geopolítica entre EEUU y China añade otra capa de tensión al sector. Mientras las grandes compañías estadounidenses apuestan cada vez más por el código cerrado para proteger su ventaja competitiva –excusándose en medidas de seguridad–, los modelos chinos como DeepSeek, Qwen –del consorcio Alibaba– o los modelos de Baidu y Tencent se van acercando en capacidades a sus homólogos estadounidenses en cada vez más tareas, a

menudo siendo software de código abierto[12] y con una fracción del coste de desarrollo –esa fracción del coste no es completamente real, ya que parte de los costes de desarrollo y entrenamiento se basan en los de sus homólogos americanos, pero sigue afectando drásticamente a bajar los precios–, lo que amenaza directamente la capacidad de empresas como OpenAI o Anthropic de recuperar unas inversiones que, como hemos visto, dependen de un ecosistema financiero circular cada vez más apalancado y sujeto a una gran presión competitiva. Al gobierno chino, en vez de intentar mantener una guerra tecnológica con los modelos americanos, limitada por las restricciones que tiene para comprar las mejores GPUs de Nvidia, le basta con hacer tambalear la fragilidad de una industria a la que Estados Unidos ha vertido grandes aspiraciones.

La ruta del semiconductor

Antes hemos advertido sobre cómo el fetichismo de la mercancía oculta las relaciones sociales que existen detrás de un producto, y en el caso de la Inteligencia Artificial este fenómeno resulta especialmente flagrante. Todo el conflicto financiero situado en la superficie de la industria actúa como cortina de humo, ocultando una vasta cadena de producción sostenida por millones de trabajadores sin los cuales la Inteligencia Artificial sencillamente no existiría.

Como ya hemos visto, para poder entrenar satisfactoriamente a todos los modelos, se requiere de ingentes cantidades de información «etiquetada», es decir, datos organizados y clasificados previamente por humanos que permiten al modelo aprender patrones en base a estos. Una de las principales fuentes de datos fue, evidentemente, Internet. Pero su etiquetado distaba mucho de ser un proceso automático. Y aquí, en el etiquetado de datos, es donde empieza a ser visible la cadena de trabajo invisible.

Como es bien sabido, internet está repleto de contenido sexual, violento o de información deliberadamente falsa, pudiendo generar desviaciones peligrosas en el modelo. Es por ello por lo que se requiere de un etiquetado de datos manual, tanto para marcar ese tipo de contenido en su entrenamiento como para que los modelos mejoren en las capacidades de filtrado y detección en su ejecución. Para ello se replicó el modelo de negocio que ya se implementaba desde hacía tiempo en otras empresas como Facebook o TikTok: el empleo de grandes cantidades de mano de obra barata para etiquetar datos. Las grandes tecnológicas contratan a empresas como Sama, Figure Eight o Scale AI –las tres basadas en San Francisco–, Appen –con sede en Australia y Washington–, Accenture –con sede en Irlanda–, la francesa Teleperformance o la canadiense Telus Digital; cuyo

[12] Aunque algunos modelos chinos ya están empezando a plantear la aplicación de limitaciones de comercialización o de acceso al software a empresas extranjeras.

modelo de negocio es la externalización de la obtención, etiquetado y evaluación de información eminentemente en formato digital, ya sea texto, imágenes, vídeos, etc., destinados al entrenamiento de modelos de Inteligencia Artificial. Bajo la excusa de «dar oportunidades a trabajadores de países empobrecidos» contrataban a millones de obreros desesperados, externalizados en más de 170 países, la mayoría de ellos ubicados en regiones como Kenia, Uganda, Colombia, India y el sudeste asiático., para etiquetar distintos tipos de datos. El caso más sangrante es, seguramente, el de los trabajadores que se dedican a etiquetar contenido racista, sexista, violento, etc.

El caso mejor documentado es el de la empresa Sama en Kenia. Sus trabajadores debían leer y etiquetar hasta 150 o 250 textos de entre 100 y 1000 palabras cada uno durante su jornada laboral, exponiéndose a contenido explícito de violencia, abuso sexual infantil y otros materiales extremos. Todo ello a cambio de un sueldo de entre 1,32 y 2 dólares la hora, mientras Sama facturaba, según quedaba estipulado en los contratos con OpenAI, 12,5 dólares por hora trabajada.

Se han dado situaciones similares en otras empresas que trabajaban con imágenes o vídeos de muertes, violaciones y contenido explícito en general. Un testimonio recogido por *Until Everyone is Free*[13] habla de una compañera de trabajo moderadora que se encontró con un video de la ejecución de un familiar en el trabajo. Muchos de los testimonios afirmaron haber terminado con secuelas psicológicas o físicas, desde la ansiedad y la depresión hasta muertes y suicidios tras meses en ese tipo de empleos, algunos de ellos coaccionados para trabajar más de lo que estipulaba su jornada y sin poder desplazarse a su país de origen o ver a su familia durante largos períodos. Muchos trabajadores tenían contratos de entre uno y tres meses de duración que eran renovados mensualmente y con salarios que no se ajustaban al contrato. Todo esto con la complicidad de los distintos gobiernos que, para atraer más capital a sus respectivos países y evitar que las empresas no se reubicaran en otra región, dieron manga ancha para que fueran las empresas las que dictasen las condiciones de trabajo, dieron ventajas fiscales e incluso mantuvieron relaciones diplomáticas directas con altos cargos empresariales.

OpenAi terminó su relación con Sama después de que surgieran controversias por sus condiciones, pero el trabajo ya realizado por esos empleados sirvió para alimentar los modelos actuales que hoy se comercializan; y la práctica de subcontratar a este tipo de empresas sigue vigente entre todos los gigantes del sector.

[13] Until Every Pod (podcast), https://untileverypod.com/.

Otro caso anterior es el de la empresa M2Work, un proyecto del Banco Mundial en colaboración con Nokia que proporciona «microempleos» a ciudadanos palestinos por 1 o 2 dólares la hora[14], muchos de ellos destinados también al etiquetaje de datos. Lo mismo ocurre con refugiados sirios, contratados por empresas como Lifelong AI, enmascarando de oportunidades caritativas a trabajos sin condiciones garantizadas y con contratos que pueden finalizar en cualquier momento. Jordania acoge a más de 1.3 millones de refugiados, generando una enorme base de trabajadores en busca de cualquier empleo, muchos de ellos potenciales trabajadores en tareas de etiquetaje de datos. La terrible ironía es que, en no pocas ocasiones, las mismas imágenes con las que trabajarán servirán para entrenar software de drones y otros sistemas de uso militar que acabarán siendo utilizados en los países de los que han tenido que huir.

Pero la cadena de valor no termina aquí. El enorme entramado financiero del capítulo anterior descansa en una red de producción muy concentrada en puntos clave del planeta. Nvidia, como hemos visto, diseña las GPUS, pero no fabrica sus componentes esenciales. Este trabajo recae sobre Taiwan Semiconductor Manufacturing Company (TSMC), pero detrás de ambas se sitúa la menos conocida Advanced Semiconductor Materials Lithography (ASML), una empresa neerlandesa que actualmente posee el monopolio mundial de maquinaria de litografía ultravioleta extrema (EUV). La empresa fabrica herramientas de altísima precisión para imprimir circuitos extremadamente pequeños sobre las obleas de silicio. La barrera tecnológica de ASML es difícil de superar por la competencia debido a la inversión necesaria y el largo período de especialización y puesta en marcha de la producción de hasta varias décadas. Se trata de un gigante de la industria que cuenta con más de 44.000 trabajadores en distintos países, de modo que empresas como Intel u otros competidores chinos, por ahora, no son capaces de igualar sus capacidades, haciendo que el mercado mundial dependa de esta empresa.

TSMC compra las herramientas a ASML y las utiliza para fabricar los chips. La empresa taiwanesa tiene también su propio monopolio de ensamblaje y fabricación de chips, resultado de la ventaja tecnológica que ha construido durante los últimos cuarenta años, siendo también otro conglomerado que cuenta con más de 90.000 trabajadores. Ésta cuenta también con su propia barrera de entrada difícilmente franqueable en el sector y con un capital crítico de miles de millones de dólares de inversión, además de mano de obra capacitada difícil de replicar, proveedores de materiales, etc. A todo esto, se suma que Taiwán ocupa un punto geoestratégico y logístico casi inigualable en el resto del globo. El estrecho recibe un flujo comercial incomparable, además de ser próximo a los proveedores de silicio japoneses –aunque buena parte también viene de Alemania–, a las

[14] Laila Al-Arian, "In Palestine, Death by a Thousand Micro-Jobs," Al Jazeera, 12 de abril de 2013, https://www.aljazeera.com/opinions/2013/4/12/in-palestine-death-by-a-thousand-micro-jobs.

tierras raras de China, Australia y Vietnam, a unas relaciones comerciales ya establecidas con ASML difíciles de redirigir a otras zonas, etc. Es también uno de los clientes más importantes de Air Liquide –empresa francesa líder en gases industriales–, la química alemana Linde o la americana Air Products, todos ellos proveedores habituales de neón de alta pureza –esencial en la producción de los semiconductores–. Esta suma de factores convierte a TSMC en otro elemento clave.

Toda esta cadena internacional converge en un problema que inquieta cada vez más a los CEOs de las grandes tecnológicas: el incremento drástico y sostenido de recursos necesarios para mantener y mejorar el entrenamiento de los modelos actuales, surgidos de un paradigma con unos límites difícilmente infranqueables y cuya producción requiere de una serie de actores con un altísimo poder y proyección internacional, todo esto en un clima prebélico y donde van estallando conflictos que amenazan con la estabilidad de esta cadena. Los centros de datos y la industria de semiconductores en general serán cada vez más susceptibles de ser atacados como objetivo estratégico en conflictos interimperialistas debido al uso intensivo de drones u otras implementaciones de la Inteligencia Artificial. Ahí está la amenaza de los hutíes con la capacidad de cortar el 17% del tráfico mundial de internet si boicotean los cables submarinos, los ciberataques masivos, la escasez de agua y otros bienes de consumo esenciales. Tenemos más ejemplos, como los ataques que Irán lanzó contra centros de datos de Amazon en Emiratos Árabes y Baréin, la necesidad de migraciones de datos de los centros cercanos al conflicto a otros de más alejados, tal y como sucedió en Ucrania, y todo esto sin contar con la escalada de tensiones entre China y EEUU.

A esta situación se le suma el enorme consumo de electricidad y agua de los centros de datos, que va aumentando de forma proporcional a la complejidad de los modelos y la popularidad de su uso masivo, pues estos costes no dependen solo del entrenamiento, sino que crecen de forma no lineal a medida que se incrementa el uso de tokens, pues cada fragmento generado debe mantener coherencia no solo con las palabras inmediatamente anteriores, sino con la totalidad del texto de entrada. Se estima que el cómputo total de centros de datos representa el 1,5% de la energía mundial anual y entre 4 y 7 billones de litros de agua al año para refrigeración –en su mayoría agua dulce– y todo esto sin contar con los costes indirectos de componentes y otros procesos intermedios[15].

[15] El gasto principal proviene de las GPUs y CPUs especializadas, que representan entre el 47% y el 64% de los costes, y que siguen aumentando debido a la escasez de estos productos por el aumento de la demanda de los últimos años. Este crecimiento representa que la escalada de precios para entrenar nuevos modelos podría llegar a los miles de millones a partir de 2027: Ben Cottier et al., "The Rising Costs of Training Frontier AI Models," arXiv preprint arXiv:2405.21015 (mayo de 2024), https://arxiv.org/pdf/2405.21015.

El coste de entrenar a GPT-3 fue de 1.3GWh, mientras que el coste de GPT-4 escaló hasta los 50GWh. Algunas estimaciones llegan a la conclusión de que, para 2030, la IA consumirá cantidades del orden de TWh. Cabe destacar, pero, que todo esto son cálculos estimados, porque las empresas del sector no publican cifras oficiales de consumo. Pero no pueden esconder la realidad de que es una industria colosal con cantidades ingentes de dinero invertidos. En total, las grandes tecnológicas han invertido aproximadamente 1.5 billones de dólares –y subiendo– mientras su retorno ha sido de aproximadamente 790 mil millones[16], estancando a estas empresas en una situación que recuerda demasiado a la burbuja de las puntocom. Si su colapso afectará o no a otras ramas de la economía es algo que no podemos saber de antemano, pero la cantidad de empresas e intereses que están de por medio son factores a tener en cuenta. El avance contradictorio del modo de producción capitalista termina destruyendo sus propias fuerzas productivas, convirtiéndose en una traba para sí mismo. La cada vez mayor infraestructura se vuelve más susceptible de colapsar debido a la cadena a escala global necesaria para mantenerla.

El becerro de oro

Todo el circuito anterior, la inversión circular, la red de producción, las limitaciones computacionales y lógicas de la Inteligencia Artificial, revelan el espejismo que genera el fetichismo dominante entorno a esta tecnología. El repaso exhaustivo del circuito completo de la IA no ha sido un mero paréntesis expositivo, sino el fundamento necesario para desmontar las apariencias que envuelven a la IA y volver a los dos grandes relatos ideológicos que mencionábamos al comienzo del artículo, pero ahora habiéndonos quitado de encima las apariencias fetichistas.

El primero, citado ya al inicio de este artículo es el relato que alimenta las ansias desbocadas de los CEOs, los «tecnobros» y todo el ejército mercadotécnico que cada dos o tres meses lanza el aviso de que estamos, ahora sí que sí, al borde de alcanzar la Inteligencia Artificial General que convertirá todos los empleos en obsoletos y transformará el mundo radicalmente. Esta narrativa viene reforzada por buena parte de los despidos por reajustes de plantilla que sacuden al sector tecnológico. Pero en muchos casos estos despidos no son consecuencia de la obsolescencia generada por la IA, sino de la sobrecontratación que hubo en algunos sectores durante la pandemia o a simples reajustes por una reducción de productividad o una sobresaturación de algún sector, atribuyéndolos a la IA para que comunicativamente resulte más interesante a los

[16] Is AI Profitable?, https://isaiprofitable.com/.

inversores creer que existe una tecnología que multiplicará su capital y no que ha habido un error de gestión o que el negocio, sencillamente, va peor de lo esperado.

La segunda es, como también señalábamos al principio, el reverso catastrofista de corte pequeñoburgués, que, en vez de caer en las promesas falsas y en un frenesí inversionista, genera un miedo neoludita que reaparece cíclicamente en forma de debates sobre el uso «ético» de la Inteligencia Artificial. Equiparando su utilización a una traición al mismo ser humano. Y esto por no hablar de aquellos que critican a la IA por «violar la propiedad intelectual» de pequeños productores demasiado acostumbrados al gremialismo.

Se habla de que «la IA discrimina», «roba propiedad intelectual», «destruye empleos» o aplica medidas que perjudican a los trabajadores, como si se tratara de un sujeto autónomo capaz de tomar decisiones y perseguir fines propios. Sin embargo, una vez comprendido el funcionamiento material de estos sistemas, resulta evidente que esa atribución de responsabilidad constituye una inversión fetichizada. La IA no actúa por voluntad propia ni decide los objetivos que persigue; son los intereses capitalistas los que guían a los modelos, son las directrices de las juntas directivas las que seleccionan los datos y fijan su uso. Desplazar la culpabilidad de las relaciones sociales y económicas hacia la tecnología no solo antropomorfiza un instrumento, sino que, además, contribuye a ocultar a los verdaderos responsables de las decisiones que se toman a través de él.

Marx ya señaló que solo aquello que es organizado en base al trabajo privado aparece como ajeno al control de la sociedad. Pero ese movimiento aparentemente autónomo responde a las relaciones sociales mediadas por la mercancía. La IA es otro elemento más con resultados terribles cuando es puesta al servicio de fines capitalistas, igual que un fusil o un bulldozer. La supuesta autonomía de la IA, su estallido mediático y las predicciones –catastróficas o providenciales– sobre su uso son, en su mayoría, un fetiche burgués.

Es cierto que la Inteligencia Artificial consume grandes cantidades de energía y agua y se sostiene de una red de producción sujeta sobre los hombros de millones de trabajadores explotados. Pero quienes se escandalizan por su uso deberían aplicar la misma indignación selectiva a otras industrias con huellas comparables o tasas de explotación tan elevadas, como las plataformas de *streaming*, las redes sociales, la industria textil, las empresas cárnicas o el monocultivo intensivo. Esto no exime a ninguna de las industrias de la sangre y lodo que supuran, pero el miedo responde más al lugar simbólico que ocupa la Inteligencia Artificial en el imaginario colectivo que no a la forma en cómo se produce.

Entonces, ¿cuál es el futuro del sector y de la clase obrera tras la irrupción incierta de la Inteligencia Artificial? Si no nos dejamos engañar por espejismos, se vislumbran dos tendencias, no necesariamente excluyentes:

La primera es que la Inteligencia Artificial no está generando los retornos que se prometieron a los inversores y el modelo de negocio se tendrá que reorganizar, pudiendo arrastrar por el camino a varias empresas en una corrección financiera que puede o no afectar a otras ramas de la producción. Y es que las empresas ni siquiera tienen claro cuál es su producto: ¿son las herramientas online de IA generativa y, por lo tanto, deberán cobrar cantidades posiblemente inabarcables para la mayoría de los usuarios con tal de recuperar la inversión? ¿Su promesa es la obsolescencia y sustitución de millones de trabajos? De ser así, ¿hasta qué punto ha sido esto verdad y en la producción de qué mercancías exactamente? La indefinición del producto es en sí misma una forma de hacer negocio poco estable a largo plazo.

La segunda tendencia es la utilidad real que sí tiene la IA en distintas industrias, una utilidad más modesta que las aspiraciones grandilocuentes de las grandes tecnológicas. Se está dando una transformación progresiva de distintos sectores donde sí se está implementando extensivamente la Inteligencia Artificial y donde sí puede tener un impacto significativo. Estos sectores son, principalmente, el desarrollo de software[17] y contenidos digitales, la atención al cliente, el marketing y la gestión documental, entre otros.

Esta implementación puede ampliar significativamente la base de trabajadores que podrán acceder a este tipo de empleos, reduciendo la especialización y el prestigio gremial que protegía a algunas profesiones dentro de ciertos trabajadores de «cuello blanco». Por ejemplo, el software, hasta ahora un oficio con un carácter muy «artesanal», tendería a convertirse, en buena parte de la industria, en un producto más estandarizado. Si hasta los papers científicos terminan siendo escritos por una Inteligencia Artificial y leídos por otra, únicamente para alimentar el cómputo de texto «científico» que se produce, quizás el problema venía de antes de que la IA entrara en escena.

Y aquí debemos hacernos otra pregunta incómoda: ¿Es realmente mejor el código generado por IA que el de un programador? ¿Es una imagen publicitaria generada por DALL·E equivalente a la de un diseñador gráfico? ¿Es una voz generada por IA atendiendo al teléfono mejor que un humano? En términos de lógica capitalista, es

[17] Los modelos de LLM, excelentes en reproducir lenguaje estructurado, son especialmente buenos en la reproducción de código, pues presenta una sintaxis más estructurada que el lenguaje natural, facilitando el trabajo inductivo de la Inteligencia Artificial.

indiferente, porque su producción es más rápida y el valor de uso nunca es lo que le interesa al capitalista, sino la ganancia.

Esto no eliminará la necesidad de muchos trabajos con alta especialización técnica, pues la IA sigue necesitando revisión y una guía humana, y hay muchas actividades que no pueden ser sustituidas por las limitaciones que hemos tratado en capítulos anteriores. Aun así, muchas tareas pueden automatizarse y su trabajo necesario se simplificará. Paradójicamente, lo que en un principio podría suponer una reducción drástica de mano de obra, también amplía el capital y las necesidades del mercado[18], cosa que genera más necesidad de procesos intermedios que, en vez de eliminar de forma abrupta todo el trabajo humano, lo reconfigura e intensifica. Más que una desaparición repentina de empleo, la reducción del tiempo de trabajo necesario no representará una mejoría en las condiciones de la humanidad trabajadora, sino que, por el contrario, dará lugar a una lenta pero progresiva reducción relativa de la mano de obra, tal y como ha ocurrido siempre bajo las relaciones de producción capitalistas.

Esto, en términos históricos, no supone una novedad. Ya se trate de los latigazos de un capataz o de un algoritmo que reporte meticulosamente los descansos, que el obrero se someta a los ritmos de la máquina no es una anomalía introducida por la Inteligencia Artificial, sino una necesidad del modo de producción capitalista. Todo trabajador está acostumbrado a tener que medir el tiempo milimétricamente en función de los ritmos de una máquina, a que los managers o los responsables presionen para terminar una tarea antes de tiempo, aunque esté peor hecha, a que cualifiquen nuestro trabajo en base a métricas que ni siquiera comprendemos del todo. Quizás, a algunos trabajadores de cuello blanco, la aparente comodidad de una oficina, el teletrabajo o la posibilidad de trabajar sentados son factores que los han llevado a creer que estaban más seguros, pero los mismos mecanismos que se aplican al obrero industrial clásico se aplican al resto de trabajadores y, a pesar de tener el despacho del jefe más cerca, los intereses de estos obreros se hallan más cercanos a los del resto de su clase. Si la IA debe servir para algo es para evidenciar una vez más la necesidad de la clase trabajadora mundial de organizar conscientemente la producción.

[18] En el caso de la programación, existe un precedente histórico cercano para entender este proceso y que ya hemos mencionado antes. Las calculadoras y los ordenadores sustituyeron el cálculo manual, mecanizándolo y acelerándolo, pero no eliminaron la necesidad de aprender matemáticas, porque las matemáticas nunca fueron, o no únicamente, la resolución mecánica de operaciones.