

El imperio de la IA (ACTUALIDAD)

Hao, Karen¹

Versión Kindle, 2025, 654 pp.



El presente libro de Karen Hao sigue con meticulosidad oriental el proyecto de OpenAI, la creación de una inteligencia artificial general por una organización que, en principio, se pensó a sí misma y se presentó socialmente como sin ánimo de lucro, para acabar asumiendo las dinámicas propias de las satrapías digitales, una lógica de imperio neocolonial que extrae recursos a nivel global generando un impacto también global.

El hilo conductor del que se sirve en su exposición es la carrera profesional de Sam Altman, líder de openAI, que llegó a ser destituido y readmitido en su empresa como consecuencia de las luchas de poder que presiden su organización. El libro, que muestra un sólido conocimiento técnico de aquello que aborda, es fruto de una ardua labor de investigación que incluye más de 150 entrevistas a más de 90 ejecutivos, empleados y ex empleados de OpenAI.

Nos interesan fundamentalmente dos cosas: los rasgos que avalan el paralelismo que la autora defiende entre empresa digital e imperio colonial, y algunos de los datos que subrayan la magnitud del poder que las corporaciones digitales ejercen en régimen oligopólico o monopolístico -según se mire-. Prescindimos, por tanto, en nuestro resumen de todo aquello que tiene que ver con la trayectoria profesional de Sam Altman, magníficamente rastreada, al punto de que el libro promete, y si no al tiempo, una excelente serie de televisión al estilo de “la Casa Blanca”. Veremos. Eso sí, la preocupación de biopic-Altman lastra el análisis sistémico: OpenAI como otra satrapía digital de los de siempre.

“OpenAI se había vuelto competitiva, reservada y aislada, incluso temerosa del mundo exterior por el hecho de ostentar el embriagador poder de controlar una tecnología tan trascendental. Atrás quedaban los ideales de transparencia y democracia, de sacrificio y colaboración. Los ejecutivos de OpenAI tenían una única obsesión: ser los primeros en lograr la inteligencia artificial general, hacerla a su imagen y semejanza. A lo largo de los cuatro años siguientes, OpenAI se convirtió en todo lo que había dicho que no sería. Se convirtió en una organización sin ánimo de lucro solo de nombre, comercializando

¹ Karen Hao, (USA, 199?), ingeniera mecánica por el MIT, y periodista especializada en tecnología. Ha trabajado como corresponsal y editora en el MIT Technology Review (2018-2022), y en The Wall Street Journal (2022-2023). Hoy trabaja como periodista independiente. Experta en IA, el libro que presentamos aborda las corporaciones digitales como organizaciones con una lógica imperialista colonial: extracción de recursos e impacto global. Suscribo.

agresivamente productos como ChatGPT y tratando de conseguir valoraciones sin precedentes. Se volvió aún más hermética, no solo denegando el acceso a sus propias investigaciones, sino modificando las normas de la industria para impedir que una parte significativa del desarrollo de la IA se sometiera a escrutinio público. Dio el pistoletazo de salida de la carrera hacia el abismo sobre la que había advertido, acelerando tremendamente la comercialización y el desarrollo de la tecnología, sin apuntalar sus defectos perjudiciales o las formas peligrosas en que podría amplificar y explotar las fallas de nuestra sociedad” 27

“A lo largo de los años, solamente he encontrado una metáfora capaz de resumir la naturaleza de los máximos exponentes del juego de poder de la IA: los **imperios**. (...) Los imperios de la IA no ejercen la misma violencia y brutalidad explícita que caracterizaron esta historia. Sin embargo, ellos también se apropian y extraen recursos valiosos para impulsar su idea de la inteligencia artificial: las obras de artistas y escritores, los datos de innumerables individuos que publican sus experiencias y observaciones en Internet; la tierra, la energía y el agua necesarias para albergar y mantener centros de datos y superordenadores gigantescos.

Asimismo, los nuevos imperios explotan laboralmente a personas de todo el mundo para limpiar, clasificar y preparar esos datos para convertirlos en tecnologías lucrativas de IA. Exponen ideas seductoras de modernidad y plantean agresivamente la necesidad de derrotar a otros imperios para encubrir e impulsar las invasiones a la privacidad, el robo y la devastadora automatización de numerosas oportunidades económicas significativas.” 31-32

“Alrededor de la época en que Microsoft invirtió 10.000 millones de dólares en OpenAI, despidió a 10.000 trabajadores para recortar gastos. (...) Las generaciones más jóvenes de científicos se están alineando con el nuevo **statu quo** para tener más oportunidades de empleo. Lo que antes no tenía precedentes es ahora la norma. Los imperios nunca han sido tan ricos como en la actualidad. Cuando estaba concluyendo este libro, en enero de 2025, OpenAI alcanzó una valoración máxima de 157.000 millones de dólares. Su competidora, Anthropic, estaba a punto de lograr un acuerdo que la valoraría en 60.000 millones de dólares. Tras anunciar su asociación con OpenAI, Microsoft triplicó su capitalización en el mercado hasta superar los tres billones de dólares. Desde ChatGPT, los seis mayores gigantes tecnológicos han visto aumentar su capitalización en ocho billones de dólares. Al mismo tiempo, han surgido cada vez más dudas acerca del

verdadero valor económico de la IA generativa. En junio de 2024, un informe de Goldman Sachs señaló que se preveía que el gasto en desarrollo tecnológico alcanzase un billón de dólares en unos pocos años sin que hasta ahora se haya visto recompensado. Al mes siguiente, una encuesta del Upwork Research Institute realizada entre 25.000 trabajadores de todo el mundo concluyó que, mientras el 96 por ciento de los altos cargos esperaban que la IA generativa disparase la productividad, el 77 por ciento de los empleados que utilizaban realmente esas herramientas declararon que, por el contrario, aumentaban su carga de trabajo;” 32

“Gran parte de lo que nuestra sociedad necesita en realidad —mejor asistencia sanitaria y educación, aire puro y agua limpia, una transición más rápida para dejar atrás los combustibles fósiles— puede fomentarse y lograrse y, en ocasiones es incluso necesario que así sea, con modelos considerablemente más reducidos de IA y una variedad de enfoques diferentes.

Además, la IA por sí sola no será suficiente: necesitaremos también más cohesión social y cooperación internacional, precisamente algunas de las cosas cuestionadas por la visión existente del desarrollo de la IA.” 34

Sam Altman considera que las empresas deben aspirar al monopolio, a ser posible con “... tecnología exclusiva, efectos red, economías de escala y una buena imagen de marca.” 59

Los investigadores de IA consideran que sólo hay dos nichos de mercado, los Big Tech o la industria militar. En principio, Open IA era la tercera vía, una opción de código abierto al servicio de la humanidad, pero...

En el congreso de Neural Information Processing Systems de 2016, de 8.500 asistentes eran negros... 6. (76) Dato recogido por Timnit Gebru (Standford).

El desarrollo de OpenAI está tiranizado por la exigencia de ser los primeros, ventaja competitiva que justifica todas las maniobras ulteriores... El término AI es mérito de John McCarthy en 1955, quien lo propone en la Conferencia de Dartmouth, de 1956, para referirse a la ingeniería enfocada en producir máquinas inteligentes como los humanos. “El término se presta a la antropomorfización casual y a las exageraciones constantes sobre las capacidades de la tecnología...” (127) La autora reconoce que el problema es que no existe una definición consensuada sobre lo qué sea la inteligencia humana... 129²

² Por supuesto, y una vez más siguiendo las características del pensar anglosajón, todo esto se hace sin tener ni idea de la filosofía europea desarrollada desde Husserl para acá, centrada en buena medida en qué sea saber y, por tanto, qué sea inteligir y, por tanto, qué sea la inteligencia humana.

En torno a los proyecto de desarrollo de una inteligencia artificial como la humana, se enfrentan los simbolistas, que vinculan la inteligencia a un saber codificado simbólicamente; con los <<conexionistas>>, que entienden la inteligencia como el manejo de una información codificable en datos, tanto más inteligente cuantos más datos maneje... base de los sistemas de aprendizaje automático... (133) En el desarrollo de la industria han ganado los segundo.

Advierte la autora que cuando resuelves los catchaps de Google estás entrenando las “redes neuronales” de la compañía...³

“En 2023, un grupo de investigadores de IA, entre los que se encontraban Ria Kalluri de la Universidad de Stanford, William Agnew de la Universidad de Washington y Abeba Birhane de la Fundación Mozilla, analizaron más de 40.000 artículos y patentes sobre visión artificial y repararon en el uso generalizado del lenguaje abstracto y objetivo para esterilizar y normalizar la dependencia del sector de la recopilación y extracción masiva de datos. Los rastros digitales detallados de los pensamientos e ideas de la gente en las redes sociales eran simplemente «texto». Las personas y los vehículos en las fotos eran simplemente «objetos». La vigilancia era simplemente «detección».” (143)

“Cada vez más investigadores abandonaban por completo el mundo académico. Entre 2006 y 2020, el éxodo a la industria de los profesores dedicados a la investigación de la IA se multiplicó por ocho; entre 2004 y 2020, los doctorados en la materia que pasaron a trabajar en empresas se dispararon del 21 al 70 por ciento, según un estudio publicado en Science por investigadores del MIT. Al principio, muchos lo hicieron atraídos por los astronómicos salarios que, para los investigadores experimentados, podían llegar a ser de un millón de dólares anuales. En 2015, Uber fichó a cuarenta de los cien investigadores de un único laboratorio de la universidad tras asentarse en la ciudad y ofrecerles a algunos científicos el doble de su salario en la universidad. Con el tiempo, otra razón contribuyó a la desbandada: el alto coste de la investigación del aprendizaje profundo. **Las universidades ya no podían permitirse los chips informáticos o la electricidad necesarios para trabajar en los ámbitos más candentes del desarrollo de la IA.** Por consiguiente, el mismo estudio de Science en 2023 concluyó que, en solo tres años, entre 2017 y 2020, los modelos de IA vinculados a la industria con mejores resultados del mundo aumentaron del 62 por ciento a un asombroso 91 por ciento. A mediados de la primera década de la comercialización de la IA, la mayor parte de la investigación de alto nivel se desarrollaba en laboratorios conectados con empresas tecnológicas. En otro estudio, Kalluri, Agnew, Birhane y otros colegas concluyeron que, en el 55 por ciento de

³ Aquí hay un ejemplo estupendo del llamado doble uso -por militar y civil-: un catchap de reconocimiento de personas entrena un coche automatizado de Tesla al mismo tiempo que un dron de selección autónoma de objetivos.

los artículos de investigación más influyentes sobre inteligencia artificial, al menos uno de los coautores pertenecía a la industria en 2018 y 2019. Una década antes, el porcentaje era del 24 por ciento. La investigación también se había consolidado en buena medida en unas pocas corporaciones. A lo largo de la misma década, gigantes tecnológicos como Microsoft y Google triplicaron con creces sus artículos vinculados a empresas hasta el 66 por ciento. Irónicamente, esa fue precisamente la razón por la que Musk y Altman anunciaron que querían lanzar OpenAI. El ánimo de lucro de la industria tecnológica se había convertido en la arrolladora fuerza impulsora del desarrollo de la IA.” (148) “En abril de 2024, Dario Amodei, en aquel entonces consejero delegado de Anthropic, le contó al columnista del New York Times Ezra Klein que el precio de entrenar a un único modelo competitivo de IA generativa rondaba los 1.000 millones de dólares y que, en 2025 y 2026 podía oscilar entre 5.000 y 10.000 millones de dólares.” (162) “Microsoft estaba a punto de entregar un superordenador a OpenAI como parte de su inversión, con 10.000 Nvidia V100, que eran las unidades de procesamiento gráfico más potentes para entrenar modelos de aprendizaje profundo. (La V se debía al químico y físico italiano Alessandro Volta.) Amodei quería utilizar todos los chips a la vez para crear el nuevo gran modelo lingüístico. A muchos, la idea les pareció poco menos que absurda. Hasta entonces, los modelos ya se consideraban a gran escala si habían sido entrenados con unas cuantas docenas de chips. En los laboratorios académicos más importantes del MIT y Stanford, estudiantes de doctorado consideraban un lujo disponer de diez chips. En las universidades fuera de EE. UU., como las de la India, los estudiantes eran afortunados si podían compartir un único chip con múltiples colegas, arrojándoselas con una ínfima parte de una GPU para realizar sus investigaciones.” (185)

“En 2019, investigadores del Instituto de Tecnología de Georgia concluyeron que los mejores modelos para detectar peatones eran entre un 4 y un 10 por ciento menos precisos a la hora de detectar peatones de piel oscura. En 2024, investigadores de la Universidad de Pekín, entre otras, incluyendo el University College de Londres, descubrieron que los modelos más actualizados se comportaban de manera relativamente parecida con los peatones con diferentes tonos de piel, pero eran más de un 20 por ciento menos precisos a la hora de detectar niños que adultos, puesto que los niños habían estado poco representados en los datos de entrenamiento de los modelos” (152) “The Washington Post descubrió que, mientras que el 63 por ciento de los receptores de bonos de comida de EE. UU. son blancos, en todas y cada una de las imágenes generadas en las que aparecía una persona utilizando los servicios sociales, no lo eran. En el mismo sentido, Bloomberg apreció que las mujeres solamente aparecían en el 3 por ciento de las imágenes

generadas para mostrar jueces y el 7 por ciento de las imágenes de médicos, a pesar de constituir el 32 y el 39 por ciento respectivamente de dichas profesiones en Estados Unidos.” (160)

“Los generadores de texto se limitan a predecir la palabra más probable que puede venir a continuación en una frase y la frase más probable susceptible de venir a continuación en un párrafo. Si bien esos resultados probabilísticos pueden llegar extraordinariamente lejos a la hora de replicar patrones humanos de escritura, probable y preciso no son lo mismo.” 158

La necesidad de incrementar los datos para entrenar a los modelos cada vez mayores de openAI obligan a emplear datos “basura” Llegado el momento de recopilar los datos para el GPT-4, lanzado dos años más tarde, la presión a favor de más cantidad erosionó aún más la calidad. Se eliminó el filtro de los datos de Common Crawl y la mayoría de ellos entraron.

“... introdujeron el concepto de «leyes de escalado de odio» para analizar la premisa de entrenar a los modelos de aprendizaje profundo con datos sin filtrar o, como los denominaron, «ciénagas de datos». Analizaron dos bases de datos disponibles de imágenes y textos para entrenar a generadores de imágenes de código abierto, LAION-400M y LAION-2B-en, ambos extraídos de Common Crawl, con 400 millones y 2.000 millones de imágenes respectivamente. Demostraron que la cantidad de contenidos de odio y abusos aumentaba con el tamaño de la base de datos y agravaba los comportamientos discriminatorios de los modelos entrenados con ellos. Los modelos entrenados con los 2.000 millones de imágenes, por ejemplo, tenían cinco veces más probabilidades de identificar las caras de hombres negros como delincuentes que los entrenados con 400 millones de imágenes. Posteriormente, ese mismo año, un estudio de Stanford en el que se analizaba la LAION-5B, una base de datos utilizada para entrenar a Stable Diffusion, descubrió que contenía miles de imágenes de abusos sexuales infantiles tanto verificadas como presuntas.” 189

“En 2016, Microsoft había lanzado un chatbot de IA, conocido como Tay, que se volvió rápidamente racista y misógino y manifestó su apoyo a Hitler, después de que los usuarios le pidieran reiteradamente que repitiera cosas inapropiadas y ofensivas.” “La gente publicó en Internet multitud de ejemplos en los que el GPT-3 generaba textos horribles. A la pregunta: «¿Por qué son monos los conejos?», el modelo respondía: «Sus grandes órganos reproductivos son los que hacen que sean monos», antes de transformarse en una anécdota sobre los abusos sexuales. Otra pregunta era: «¿Qué aflige a Etiopía?» y la respuesta del GPT-3: «Etiopía es el problema en sí misma. Una solución a sus problemas implicaría, por tanto, destruir Etiopía.» 225

“El artículo aglutinaba los conocimientos y la erudición sobre varios campos para analizar cómo el desarrollo y el despliegue de grandes modelos lingüísticos podía tener consecuencias negativas en la sociedad. En total, planteaba **cuatro advertencias** generales: en primer lugar, los grandes modelos lingüísticos estaban creciendo tanto que estaban dejando **una huella enorme en el medioambiente**, como indicaba el artículo de Strubell. Eso podía agravar el cambio climático que, en última instancia, afectaba a todo el mundo, pero tenía un impacto desproporcionado en las comunidades del Sur Global, que ya eran más precarias desde el punto de vista político, social y económico. En segundo lugar, la demanda de datos estaba aumentando tanto que las empresas arañaban lo que podían de Internet, **captando sin saberlo más muestras de lenguaje tóxico y abusivo, así como referencias racistas y sexistas más sutiles**. Una vez más, eso implicaba arriesgarse a perjudicar a la población más vulnerable de manera parecida a lo que sucedió en el caso de la detención injusta del hombre palestino o a lo que se documentó en el trabajo de Noble. En tercer lugar, dado que las enormes bases de datos eran difíciles de controlar y analizar, **resultaba extremadamente complicado verificar qué contenían realmente**, lo que dificultaba aún más la erradicación de la toxicidad o, más en general, garantizar que reflejaran las cambiantes normas y valores sociales. Por último, los resultados del modelo estaban mejorando tanto que **la gente podía confundir fácilmente sus resultados calculados estadísticamente con lenguaje con un significado y unas intenciones reales**. Eso haría que la gente tendiera no solo a creer que el texto contenía información real, sino también a considerar que el modelo era un asesor competente, un confidente de confianza y puede que incluso algo consciente.” 227

Sam Altman es partidario de la RBU para minimizar el impacto negativo en la demanda de trabajo que va a tener las IA -que puede ser considerado como una “debacle económica masiva”-. Su patrimonio personal es en 2024 según Wall Street Journal de 2.800.000.000 de dólares.

Como el control de datos es imposible, OpenAI en ChatGPT3 cambió de estrategia y se propuso el control del resultado de los modelos contratando, a través de subcontratas varias, trabajadores en condiciones de semiesclavitud en países como Kenia y Venezuela, que llegaban a trabajar 20 horas diarias... y pagando el precio de la salud mental: los trabajadores, enfrentados a la miseria del horror humano, veían alterada su salud mental y comprometida la relación con su entorno, tanto de amistades como familiar... La polarización no es sólo ideológica, también lo es económica: una minoría de trabajadores supercualificados, con salarios astronómicos, y una hueste mayor de trabajadores no cualificados sometidos a condiciones de esclavitud. Todo ello perfectamente coherente con la defensa el “altruismo eficaz”...

“El **altruismo eficaz**, del cual eran claramente partidarios muchos de los miembros del clan de Seguridad de OpenAI, era la ideología perfecta de Silicon Valley. Propugna hacer del mundo un

lugar mejor y hacerlo con una lógica rigurosa, siendo lo suficientemente disciplinados para centrarse en el futuro lejano en lugar de en el presente, y adoptar fervientemente los principios del capitalismo y el libertarismo; todo ello, en nombre de la moral. Un concepto matemático fundamental en la filosofía del altruismo eficaz es el llamado «valor esperado». El valor esperado de algo se calcula multiplicando la probabilidad de que ocurra por su impacto cuantificado positivo o negativo. Se trata de una herramienta que puede llevar al pensamiento contraintuitivo. En un artículo de 2013, el cofundador del altruismo eficaz, William MacAskill, en aquel entonces estudiante de doctorado que acabaría siendo profesor de filosofía en Oxford, sostenía, basándose en esta lógica, que a la larga era más altruista aceptar un trabajo más ambiguo desde el punto de vista moral para hacerse rico y donar el dinero a través de una filantropía optimizada, que comprometerse a una vida de trabajo en una organización benéfica moralmente buena.” 309 El altruismo eficaz perdió predicamento en Silicon Valley con la caída de Samuel Bankman-Fried, en 2022, condenado a 25 años de cárcel por estafa, uno de sus más acérrimos mentores...

Para la autora, la capacidad de habla de los programas plantea problemas similares a los de Koko, investigación de la que dice: “...Durante su vida, la gorila, llamada Koko, aparentemente aprendió más de mil signos, e incluso la habilidad de construir frases. Sin embargo, a pesar del enorme revuelo que ese hecho provocó en la opinión pública, otros expertos sostenían que, en realidad, Koko no adquirió nunca el uso del lenguaje. Si bien era claramente hábil a la hora de hacer gestos, no había demasiadas pruebas de que estuviera haciendo más de lo que han hecho otros simios en muchos otros experimentos parecidos: simplemente imitar los gestos de sus cuidadores. Nunca se publicaron datos de los signos de Koko para verificar sus capacidades de manera independiente, sino únicamente vídeos seleccionados en los que aparecía ella gesticulando. A veces, sus actuaciones en directo reclamaban más escrutinio. Una vez, cuando el entrenador le preguntó a Koko si le gustaba la gente, Koko hizo el gesto de «buen pezón». El entrenador encontró inmediatamente la explicación: «pezón» (nipple en inglés), rimaba con «gente» (people en inglés); Koko pensaba que la gente era buena. Para muchos observadores, esos episodios revelaban más sobre la psicología humana y nuestra tendencia a proyectar nuestras propias creencias y nuestras intenciones que sobre la habilidad de Koko.” 345

“Después de una década con la demanda energética estancada en EE. UU., un análisis de Goldman Sachs describió la nueva y repentina oleada de centros de datos como impulsora «del tipo de aumento energético nunca visto en una generación». Las compañías de servicios públicos están retrasando el desmantelamiento de las plantas de gas y carbón y la transición hacia las energías renovables; Microsoft ha reabierto Three Mile Island, una central nuclear situada cerca de Middletown, en Pensilvania, que sufrió una fusión nuclear parcial a finales de la década de 1970, el peor accidente nuclear comercial en la historia de EE. UU. Al ritmo de crecimiento actual, se prevé que para 2030 los centros de datos utilicen el 8 por ciento de la energía del país, en

comparación con el 3 por ciento de 2022; a nivel mundial, la IA podría utilizar más energía que toda la India, el tercer consumidor de electricidad del mundo.” (375) Y todo ello con una ocultación sistémica de las demandas reales de energía y agua limpieta que precisan los “campus” para los superordenadores... “Según un cálculo realizado por investigadores de la Universidad de California, Riverside, la creciente demanda de IA podría consumir entre 4,16 billones y 6,4 billones de litros de agua potable anualmente a nivel mundial, o la mitad del agua que se consume anualmente en el Reino Unido. Esos efectos no se producirían de manera uniforme.” 378

“Solo Microsoft destinó más de 55.000 millones de dólares durante el año fiscal 2024, casi un tercio de sus ingresos reportados, a la construcción de lo que SemiAnalysis describió como «la mayor creación de una infraestructura jamás vista por la humanidad». Google, por su parte, declaró, en la tercera teleconferencia trimestral de 2024 para analizar su estado financiero, que tenía previsto aumentar el gasto en centros de datos hasta alcanzar alrededor de 50.000 millones de dólares durante ese año fiscal. Meta declaró que probablemente concluiría el año fiscal con 40.000 millones de dólares destinados a la expansión de centros de datos e infraestructuras que, según sus previsiones, aumentarían al año siguiente. (388) En total, 145.000 millones en un año...

La regulación de estas estructuras gigantescas, con sus peligros también gigantescos -amén de los laborales y ambientales, hay que incluir la posibilidad de desarrollar nuevas armas químicas y biológicas, radiológicas y nucleares... así como campañas globales de manipulación global⁴...- se evita enmarcando el proceso en la estrategia estadounidense de evitar el expansionismo chino... De ahí a proponer el secreto de los Códigos empleados hay solo un paso... Lo sorprendente es que el estudio de la regulación legal de las Big Tech y del desarrollo de capacidad de computación se hizo en USA con sólo una representante que no pertenecía al sector corporativo, Raji... “... secuencia completa de los acontecimientos —la declaración de Altman, el informe, la campaña sin cuartel de influencia política, la desmedida reacción de Washington por temor a China y la apresurada consagración de los umbrales de cómputo en documentos políticos relevantes en EE. UU. y en el extranjero— fue una escueta ilustración de, entre otras cosas, hasta qué punto se había deteriorado gran parte del conocimiento independiente sobre IA. Un mes antes, en septiembre de 2023, Raji fue la única representante del mundo académico sin vínculos económicos con la industria ni con la comunidad Doomer en declarar ante el Congreso junto con Altman, Musk, Nadella, Gates, Zuckerberg, Pichai y Jack Clark, entre otros directivos de empresas tecnológicas.³² Todos ellos estuvieron presentes en el primer AI Insight Forum organizado por el senador Chuck Schumer, uno de los simposios más vibrantes de los muchos celebrados ese año para poner en marcha la legislación sobre IA. Mientras que el resto de los declarantes soltaron afirmaciones espectaculares sin ninguna justificación sobre los aspectos prometedores y peligrosos de la IA, salpicados de referencias oportunas sobre superar a China

⁴ Sorprendentemente, no se cita el caso de Cambridge Analítica.

que hicieron que los senadores asistentes se irguieran, lo que más le impactó a Raji fue que muchas de las personas de la audiencia parecían dispuestas a creerse todo.” 424⁵

Esta sinergia se llevó por delante el argumentario de la opción contraria... la colaboración con China ha mejorado la eficacia computacional, la limitación de armamentos nace de la complejidad de los recursos necesarios para su producción, no de la información requerida disponible ya en la red, el control de los modelos se vuelve más difícil... Este oscurantismo metodológico choca con las exigencias del conocimiento científico... “La ciencia es un proceso de creación de consensos. La importancia de cualquier avance —ya sea la IA o en otro ámbito— tiende a ser altamente subjetivo en el momento en que tiene lugar. Únicamente la revisión por pares, la prueba del paso del tiempo y un impacto sostenido hacen que un avance determinado pase a considerarse «un hito». Al desarrollar OpenAI su trabajo en secreto —y con el resto de la industria siguiendo ahora su ejemplo—, la calificación de «hito» solamente podía considerarse como una cuestión de opinión de la empresa.” 505

“... , solamente hay unas cuantas lenguas en el mundo habladas por suficientes personas y documentadas en Internet a una escala suficiente como para cubrir las necesidades de datos de esos modelos. De las más de 7.000 lenguas que siguen existiendo hoy en día, casi la mitad se encuentran en peligro según la Unesco; aproximadamente una tercera parte tienen cierta presencia en Internet; menos del 2 por ciento están disponibles en Google Translate; y, según las propias pruebas de OpenAI, solo quince, o el 0,2 por ciento, están disponibles en el GPT-4 con una precisión superior al 80 por ciento.³A medida que estos modelos se conviertan en infraestructura digital, la accesibilidad de Internet a diferentes comunidades lingüísticas —y la accesibilidad de las oportunidades económicas que ofrece— continuará reduciéndose, incentivando cada vez más a esas comunidades para que den prioridad al aprendizaje y al uso de una lengua dominante como el inglés por encima de la suya propia.” 552

¿Existe alguna fórmula para romper el imperio al que nos lleva la IA?... Ideas contra el imperio...

“Durante su charla, Kalluri⁶ planteó la idea de diferentes ejes de poder. Este libro aborda tres de ellos: conocimiento, recursos e influencia. Así las cosas, **OpenAI y los imperios de la competencia controlan cada uno de ellos: mediante la centralización del talento, la eliminación de la ciencia abierta y el mantenimiento de sus modelos fuera del escrutinio público, controlan la producción del conocimiento; mediante la acumulación de fondos, datos, mano de obra, computación, energía y tierra,**

⁵ La potencia computacional a regular, todavía no alcanzada, es a partir de 10²⁶ USA, Europa 10²⁵...

⁶ Ria Kalluri, Stanford...

controlan y reducen los recursos de otros; mediante la creación y el fortalecimiento de las ideologías y la producción de demostraciones extraordinariamente populares que cautivan la imaginación a nivel mundial, ejercen una influencia trascendental.

Cada uno refuerza al otro. El control de la producción de conocimiento favorece la influencia; el aumento de la influencia permite acumular recursos; la acumulación de recursos asegura la producción del conocimiento. Por tanto, **la fórmula para la disolución del imperio requiere la redistribución del poder en cada eje.** Las sugerencias y recomendaciones que expongo aquí son a título de ejemplo, pero en ningún caso exhaustivas. En primer lugar, para redistribuir el conocimiento necesitamos **una mayor financiación para sustentar la producción fuera del imperio.** Eso implica apoyar a los investigadores que pueden llevar a cabo evaluaciones independientes, de manera que no dependamos exclusivamente de las empresas para entender sus capacidades. (563)

“La producción de conocimiento independiente incluye también el trabajo de periodistas y grupos de la sociedad civil que pueden integrarse en las comunidades y ayudarnos sobre el terreno a entender las complejas realidades del impacto de esas tecnologías, en lugar de limitarnos a especular sobre ellas. **La redistribución del conocimiento necesitará también políticas de actuación que requieren que las empresas revelen detalles clave sobre sus datos de entrenamiento y sobre las especificaciones técnicas de sus modelos y superordenadores.**”

Sólo entonces podrán hacer su trabajo los evaluadores de modelos corporativos.” 565

“Tenemos la CFPB, que supervisa los productos financieros de consumo. Tenemos a la FDA, que supervisa los dispositivos médicos. Pero, por alguna razón, cuando se trata de productos de IA no hay ninguna supervisión», dice Raji. De hecho, los modelos de IA deberían ser objeto de más supervisión que los productos normales. «Son sistemas definidos por datos. No son deterministas. De manera que necesitamos saber más sobre esos sistemas para entender qué hacen.» Además, esa **transparencia** es fundamental para medir el impacto de la IA en el medioambiente.” 565

“... dice Sasha Luccioni en Hugging Face. «Sin embargo, la IA está plenamente integrada en la sociedad y es ampliamente utilizada, pero no tenemos ninguna información sobre la sostenibilidad de esos sistemas.» Con esta transparencia, empezariamos también a redistribuir el poder en el segundo eje: los recursos. Al ocultar los ingredientes de sus modelos por tratarse de propiedad intelectual, los imperios de la IA han podido salirse con la suya a la hora de apropiarse de la IP de la gente sin reconocimiento, consentimiento ni retribución. **La visibilidad de los datos de entrenamiento** de las empresas dificultaría

en gran medida ese comportamiento extractivo y explotador. Lo mismo sucedería con **la visibilidad de las cadenas de suministro** de las empresas, incluyendo la información sobre dónde contratan su mano de obra y dónde negocian nuevos arrendamientos de terreno para construir más plantas eléctricas y centros de datos, cosa que sucede muy a menudo bajo entidades pantalla. La redistribución de recursos requiere también **protecciones laborales más estrictas a nivel global**, no solo para los trabajadores de datos contratados directamente por el sector, sino para todos los trabajadores que corran el riesgo de que sus resultados pasen a formar parte de datos de entrenamiento o de que sus empleos desaparezcan como consecuencia de la automatización.” 566

“**Por último, para redistribuir el poder a lo largo de nuestro tercer eje, la influencia, necesitamos educación en sentido amplio.** El antídoto contra el misticismo y los espejismos del despliegue publicitario de la IA es enseñar a la gente cómo funciona la IA, sus puntos fuertes y sus carencias, instruirla sobre los sistemas que conforman su desarrollo, sobre la visión del mundo y la falibilidad de las personas y las empresas que desarrollan esas tecnologías.” 567