LITORALPRESS

Fecha: 04-11-2024
Medio: Le Monde Diplomatique
Supl. : Le Monde Diplomatique
Tipo: Noticia general
Título: **Políticas de la Inteligencia Artificial**

Pág. : 25
Cm2: 631,4
VPE: $ 1.187.584

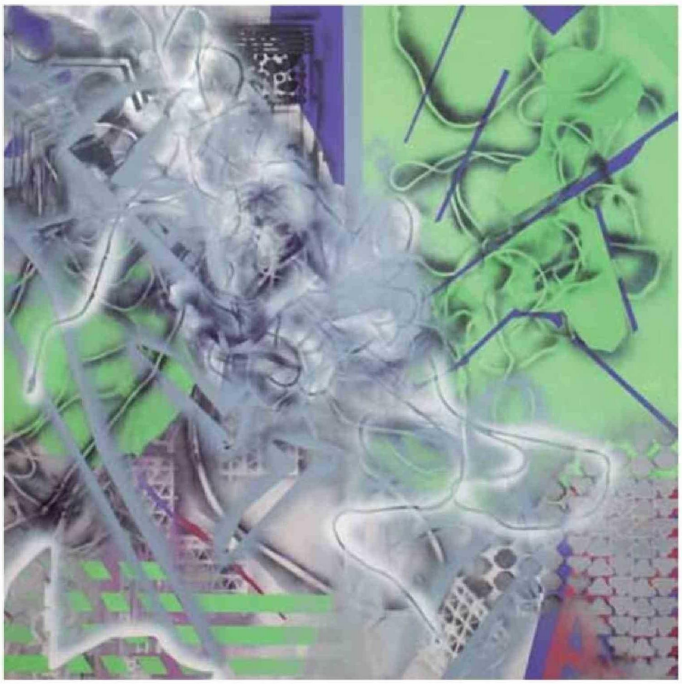
Tiraje: 6.200
Lectoría: 18.600
Favorabilidad: No Definida

## La batalla por el control del futuro tecnológico

# Políticas de la Inteligencia Artificial

por Victor Chaix, Auguste Lehuger, Zako Sapey-Triomphe*

El conflicto en OpenAI revela una grieta inesperada: ¿es posible conciliar el altruismo eficaz, que busca salvar a la humanidad con responsabilidad, y el aceleracionismo, que empuja a un progreso sin límites? Guattari soñaba con una nueva alianza entre humanos y máquinas. ¿Estamos preparados para construirla?



Raimundo Edwards, *Atajo*, (Pintura acrílica y spray sobre tela) 2019 (www.raimundoedwards.com)

En noviembre de 2023, el proyecto OpenAI, famoso por su producto estrella ChatGPT, fue el escenario de un curioso conflicto de gobernanza. El comité de dirección, encabezado por Ilya Sutskever, informático y cofundador de la empresa, destituyó al director general, Sam Altman, también informático y cofundador. Altman terminaría recuperando su puesto, pero el episodio reveló una fisura interna entre dos ideologías en apariencia opuestas, pero no tan disímiles: el altruismo eficaz (*Effective Altruism*) y el aceleracionismo eficaz (*Effective Accelerationism*). Los partidarios del primero intentaron –sin éxito– apartar a los gurúes del segundo, por temor a que condujeran a la humanidad a su pérdida.

Desarrollado en Estados Unidos en los años 2000, el altruismo eficaz pretende responder a la cuestión de la utilización óptima de los recursos para el bien común. Los defensores de esta corriente de pensamiento se sienten bien diseñados por sus capacidades intelectuales, financieras y técnicas superiores para jerarquizar y resolver los principales problemas humanos, entre los que se destacan el riesgo de pandemias, de una guerra nuclear y la aparición de una "inteligencia artificial general", a veces llamada "singularidad". Con una definición lo suficientemente imprecisa como para que unos consideren que ya ocurrió, mientras que otros imaginan que emergerá de acá a medio siglo, ese sistema de inteligencia artificial consciente engendrado en nuestro océano de datos podría conducir a la humanidad a una era de prosperidad universal o hacerla desaparecer.

Más radical que el altruismo eficaz, el aceleracionismo eficaz propone el desarrollo tecnológico desenfrenado para llegar lo más rápido posible a esa entidad suprahumana y hacer que la humanidad pase a un estadio de evolución superior, escapando así de los peligros que enfrenta. Mientras tanto, sería conveniente levantar todos los frenos reglamentarios y éticos, ignorar las cuestiones de propiedad intelectual o de respeto de los datos personales y, sin perder un instante, acelerar. Este tecno-liberalismo desinhibido justifica la comercialización de sistemas cuyo funcionamiento e implicancias aún cuesta comprender –como el ChatGPT, que Altman hizo público sin esperar–. Percibimos allí cómo aparece el modelo de sociedad presentado por la industria digital y sus aliados en el poder, el del imperativo funcional, que el filósofo Marcello Vitali-Rosati describe como "la declinación capitalista del imperativo racional, una racionalidad sometida a la necesidad de producir riqueza y de acumular mercancías" (1). Suficiente para instalar en el imaginario colectivo el reemplazo del humano por la máquina como horizonte de las transformaciones socioeconómicas y tecnológicas actuales.

Horizonte probable, pero no inevitable. En los últimos años, la planificación industrial se ha vuelto a imponer a ambos lados del Atlántico. Las élites occidentales la consideran indispensable para competir con el desarrollo asiático. La planificación ecológica también abre su camino. Desde los demócratas estadounidenses partidarios de un "Green New Deal" hasta la presidenta de la Comisión Europea, Ursula von der Leyen, todos pretenden mover los recursos del poder público y de las nuevas tecnologías para organizar la transición hacia una economía más verde, pero liberal. La izquierda propone alinear la producción con las necesidades sociales y las obligaciones medioambientales (2). En su seno, algunas voces apoyan la posibilidad de una coordinación industrial basada en sistemas de toma de decisión colectiva, que sacan partido de las recientes tecnologías de la información (3). "¿Podemos imaginar tecnologías de la información y de la comunicación que no nos exploten, no nos engañen y no nos suplanten? –preguntaba el escritor británico James Bridle–. Sí, podemos, una vez que salgamos de las redes de poder comercial que han definido a la ola actual de la IA" (4).

### Máquinas sesgadas

Así, ambos lados del espectro partidario depositan sus esperanzas en progresos técnicos que bastaría con adaptar a sus preferencias ideológicas. Ahora bien, desde su concepción hasta su realización, la inteligencia artificial (IA) no es neutra. Para desenmarañar el enredo entre técnica y política que habita en el seno de la construcción de una IA, hay que abrir la caja negra: comprender aquello de lo que se trata y cómo funcionan sus mecanismos de aprendizaje. A menudo el debate público deja a un lado esta etapa esencial, que sin embargo permitiría disipar las fantasías gemelas de la magia solucionista y de la ansiedad antropomórfica.

En la intersección de las ciencias matemática e informática, la inteligencia artificial funciona concretamente como un sistema de entrada/salida: una función matemática que aprende, a partir de una base de datos, a proporcionar las respuestas correctas a preguntas precisas, con la intención de maximizar un objetivo definido. Si se trata por ejemplo de identificar animales, debe predecir correctamente la etiqueta asociada a cada imagen ("perro" o "pelícano"). Para ello, los ingenieros entrenan al programa a partir de un banco de imágenes etiquetadas, con el objetivo de minimizar el error de predicción. Un protocolo reajusta los parámetros después de cada intento y, cuando la probabilidad de fracaso se torna aceptable, la empresa distribuye el sistema.

Hasta aquí la teoría. En la práctica, la fría neutralidad del proceso oculta elecciones eminentemente políticas, como la de los componentes en que se basa el aprendizaje. Sin ser necesariamente conscientes de ello, los ingenieros trasladan los sesgos discriminatorios inherentes a las condiciones en las cuales son producidos los datos que nutren a la máquina. La IA Pulse, desarrollada en 2020 por estudiantes de la universidad de Duke en Estados Unidos, que sirve para despixelar imágenes, tendía a blanquear a las personas de color, al punto de engendrar un "Obama blanco" (5). Nada de intencional, por supuesto: para construir su algoritmo, el equipo de Pulse utilizó otro sistema de inteligencia artificial (StyleGAN, desarrollado por la empresa Nvidia). Concebido para generar imágenes de caras humanas "verosímiles", este último sobrerrepresenta espontáneamente a los hombres blancos, debido a su propio aprendizaje. Si bien el algoritmo de Pulse no contenía ningún sesgo intrínseco, indirectamente incorpora los de StyleGAN: cuando despixela la cara real de Obama, el programa lo convierte en un hombre blanco. Así, las presuposiciones y los estereotipos se incorporan subrepticiamente en la técnica que los naturaliza: ¿no son las máquinas consideradas objetivas y desprovistas de ideología? Algunos, víctimas de algoritmos de la policía predictiva que incorporan variables discriminatorias, deberán responder ante la justicia y aprenderán, a su costa, que eso no es así.

Si bien a veces a los datos les falta representatividad, la formalización del objetivo deja también mucho que desear. A través de una fórmula matemática, se trata de sintetizar la finalidad de la tarea intelectual a la que se apunta. Así, mientras el objetivo de los algoritmos de recomendación consiste, en teoría, en seleccionar contenidos pertinentes, cuando examinamos la traducción matemática de esta tarea aparece un objetivo completamente diferente: maximizar el tiempo que se

LITORALPRESS

Fecha: 04-11-2024
Medio: Le Monde Diplomatique
Supl. : Le Monde Diplomatique
Tipo: Noticia general
Título: **Políticas de la Inteligencia Artificial**

Pág. : 26
Cm2: 350,3
VPE: $ 658.934

Tiraje: 6.200
Lectoría: 18.600
Favorabilidad: No Definida

pasa frente a una pantalla, proponiendo el algoritmo contenidos adictivos y sensacionales con el fin de capturar la atención del usuario.

**No se trata de elegir entre el medio artificial y el entendimiento humano, sino de construir la "nueva alianza con la máquina".**

Más ampliamente, un mundo en el cual los autómatas orquestan nuestra vida digital plantea una cuestión rara vez abordada: ¿incumbe a las empresas privadas decidir ellas solas sobre los objetivos perseguidos por esas IA? Esas decisiones técnico-políticas fundamentales, pasadas por alto por los dirigentes nacionales e internacionales, ante todo preocupados por regular los desbordes demasiado dramáticos o censurar los contenidos, justificarían sin embargo una deliberación colectiva y un control público más estricto, como lo sugiere un número creciente de actores del sector (6). La acumulación de una gran cantidad de datos no puede reemplazar a la reflexión democrática y al diálogo crítico. Ahora bien, todo parece dispuesto para impedirlo, desde la organización de la investigación hasta la denominación misma de "inteligencia artificial". Esta expresión, inmediatamente comprensible por el gran público, tiene como particularidad que alude a lo contrario de lo que pretende designar. En rigor, habría que hablar de "autómatas computacionales" (7), expresión claramente menos favorecedora pero más justa, dado que esas máquinas alcanzan sus objetivos calculando el mejor medio para reiterar resultados pasados. Por el contrario, la noción de inteligencia sugiere una forma de desautomatización esencial en cualquier dinámica creativa: un esfuerzo de superación de las ideas preconcebidas y estereotipadas.

## Inteligencia colectiva vs. algoritmos

Poner a las tecnologías digitales al servicio de las decisiones colectivas, es decir, también de nuestras capacidades de invención, de imaginación y de interpretación, supone una visión de "la inteligencia" diferente de la sostenida por los industriales de Silicon Valley y los transhumanistas. Según la asociación Ars Industrialis, "lo que es tonto o inteligente, no es tanto tal individuo o tal medio, sino la relación que los vincula uno con otro" (8). Tal enfoque nutría las obras de los informáticos utopistas de los años 1960 y 1970 (9). No se trata de elegir entre el medio artificial y el entendimiento humano, sino más bien de construir la "nueva alianza con la máquina", tan anhelada en 1992 por el filósofo Félix Guattari (10).

Por lo pronto, el asunto parece mal encarado, porque incluso a los investigadores más agudos les resulta difícil comprender qué pasa en la caja negra de los algoritmos. "Explicar" el funcionamiento de los modelos de IA, es decir, traducir la respuesta del sistema en una "serie de etapas vinculadas entre sí por lo que un ser humano puede sensatamente interpretar como causas o razones" (11): este principio elemental de higiene intelectual hoy por hoy ya no está entre las condiciones previas para la puesta en servicio de un modelo, pero es la frutilla del postre. La ingeniería domina la investigación hasta el punto de que los investigadores solo entienden qué es lo que las IA hacen varios años después de su comercialización o su puesta en línea. Por consiguiente, ¿cómo puede el legislador establecer normas de evaluación de sistemas que nadie sabe cómo funcionan, particularmente en los sectores sensibles de la salud o de la educación? Como ejemplo del malestar general, el MIT Media Lab forjó y popularizó la expresión "AI Alchemy" (12) como metáfora de nuestra interacción con esas cajas negras y como concepto metodológico para interpretar su incomprensibilidad.

Mientras tanto, la inteligencia artificial sigue siendo esa extraña mezcolanza entre ámbito de investigación científica, conjunto de tecnologías y mercado en pleno auge, los tres dominados por un puñado de actores cuyas capacidades financieras y pericia en política industrial siguen de cerca las de algunos países del G20. El acortamiento del proceso de innovación, desde la investigación fundamental hasta la puesta en el mercado en algunos años, incluso en algunos meses, tiene abiertamente como fuente el aceleracionismo. Las exigencias de los mercados de rentabilidad a corto plazo y la debilidad de los frenos reglamentarios refuerzan ese movimiento. Secciones enteras de la producción científica se alinean con esos imperativos, como demuestra la influencia que ejercen sobre los principales coloquios del ámbito (Sistemas Neuronales de Procesamiento de Información [NeurIPS], o la Sociedad Internacional de Aprendizaje de Máquinas [ICML]). Los laboratorios privados, dotados de recursos colosales, pueden marcar más fácilmente el ritmo en esos ámbitos en los que la infraestructura de cálculo cuesta caro y donde los mejores postores reclutan a las mentes más formadas.

En este ámbito, como en muchos otros, la "nueva alianza" de Guattari pasa por separar el Estado del mercado. ■

**1.** Marcello Vitali-Rosati, *Éloge du bug. Être libre à l'époque du numérique*, Zones, París, 2024.
**2.** "Propuesta macroeconómica. Programa del nuevo Frente Popular", junio de 2024.
**3.** Cédric Durand y Razmig Keucheyan, *Comment bifurquer. Les principes de la planification écologique*, Zones, 2024.
**4.** James Bridle, "So, Amazon's 'AI-powered' cashier-free shops use a lot of... humans. Here's why that shouldn't surprise you", *The Guardian*, Londres, 10 de abril de 2024.
**5.** Kevin Truong, "This image of a White Barack Obama is AI's racial bias problem in a nutshell", *Vice*, 23 de junio de 2020, www.vice.com
**6.** Joana Varon (dir.), "Fostering a Federated AI Commons ecosystem", T20 policy Briefing, junio de 2024, https://codingrights.org
**7.** Anne Alombert y Giuseppe Longo, "Il n'y a pas d'intelligence artificielle: parlons d'automates numériques pour rompre avec les idéologies publicitaires!", *L'Humanité*, París, 11 de julio de 2023.
**8.** Victor Petit, "Vocabulaire d'Ars Industrialis", en Bernard Stiegler, *Pharmacologie du Front national*, Flammarion, París, 2013.
**9.** Evgeny Morozov, "Otra inteligencia artificial es posible", *Le Monde diplomatique*, agosto de 2024.
**10.** Félix Guattari, "Pour une refondation des pratiques sociales", *Le Monde diplomatique*, octubre de 1992.
**11.** Christophe Denis. "Esquisses philosophiques autour de la compréhension de phénomènes complexes avec des outils de prédiction basés sur de l'apprentissage machine", Conférence francophone sur l'Extraction et la Gestion des Connaissances – Atelier Explain'AI, Blois, enero de 2022.
**12.** http://aialchemy.media.mit.edu/

*Doctorando en Humanidades Digitales, ingeniero de Investigación en IA e ingeniero, respectivamente. Este texto es la síntesis de una nota publicada en octubre en el sitio web del grupo de reflexión X-Alternative.

Traducción: Micaela Houston