

Los riesgos de la inteligencia artificial

Jose Luis Calvo Salanova

1 CONTENIDO

2	Introducción.....	2
3	Perspectiva filosófica y social	2
4	Contextualización de la IA.....	3
5	Coste marginal cero.....	4
6	Problemas actuales.....	5
6.1	Noticias falsas	5
6.2	Polarización social	6
6.3	Privacidad.....	6
6.4	Discriminación.....	7
6.5	Vigilancia masiva	8
6.6	Asimetrías de poder	8
6.7	Decisiones automatizadas	9
7	Posibles desafíos futuros.....	10
7.1	Futuro del empleo	10
7.2	Democracia	11
7.3	Cambio climático.....	12
7.4	Vínculo emocional	12
7.5	Capacidades cognitivas	12
7.6	Armas autónomas.....	13
7.7	Superinteligencia	13
8	Control y regulación.....	14
8.1	Alineamiento.....	14
8.2	Transparencia algorítmica	14
8.3	Regulación y gobernanza global.....	15
8.4	Educación y concienciación social	16
9	Conclusión.....	17
10	Bibliografía.....	18

2 INTRODUCCIÓN

El desarrollo de los ordenadores siempre ha estado acompañado por la posibilidad de que pudiera desarrollarse una forma de inteligencia de manera artificial. Ya a mediados del siglo XIX, Ada Lovelace —considerada la primera programadora— imaginó que la máquina analítica de Charles Babbage podía ir más allá del cálculo numérico y llegar a manipular símbolos de forma general. Con ello anticipaba, de forma embrionaria, la idea de una máquina que pudiera "pensar". Un siglo más tarde, Alan Turing, pionero en la construcción de los primeros ordenadores de propósito general, se preguntaría explícitamente si esas nuevas máquinas podrían llegar a ser inteligentes, e incluso propuso un experimento —el conocido “Test de Turing”— para evaluar dicha capacidad (Turing, 1950).

La mera posibilidad de que las máquinas pudieran estar dotadas de algún tipo de inteligencia ha alimentado desde entonces la imaginación de escritores, cineastas y periodistas. Películas como Terminator o Ex Machina han explorado escenarios en los que sistemas artificiales adquieren autonomía, consciencia o incluso voluntad propia. Esos relatos han calado en la cultura popular, donde no es extraño encontrar titulares que atribuyen propiedades antropomórficas a sistemas de IA: “La IA predice...”, “La IA decide...”, “La IA se rebela...”, como si hubiera una entidad consciente dentro de la máquina. Esta forma de hablar, aunque metafórica, refleja tanto el asombro como la incertidumbre que despierta esta tecnología.

Más allá de la ficción, lo cierto es que la inteligencia artificial posee un enorme potencial transformador. Bien orientada, puede contribuir a mejorar la eficiencia energética, ampliar el acceso a servicios educativos y sanitarios, acelerar la investigación científica y liberar a las personas de tareas tediosas o peligrosas. Sin embargo, como toda tecnología poderosa, la IA también entraña riesgos y dilemas éticos, desde la pérdida de control sobre decisiones automatizadas hasta la concentración del poder tecnológico en pocas manos, pasando por sesgos algorítmicos o impactos ambientales. Como ha ocurrido con otras revoluciones técnicas a lo largo de la historia, los beneficios no están garantizados: dependen de cómo decidamos diseñar, regular y aplicar esta herramienta.

Por eso es importante analizar no solo lo que la IA puede llegar a hacer, sino también cómo puede afectar al futuro de la sociedad. Porque la IA, lejos de ser una herramienta neutral, forma parte de complejas redes sociotécnicas donde actores humanos y no humanos co-construyen realidades sociales, políticas y económicas. En este texto se presentan algunos de los desafíos más relevantes que podrían intensificarse en los próximos años si no se abordan de forma crítica y colectiva.

3 PERSPECTIVA FILOSÓFICA Y SOCIAL

La inteligencia artificial no debe entenderse únicamente como una tecnología neutral o una herramienta eficiente. Su desarrollo y despliegue plantean preguntas fundamentales sobre la naturaleza del conocimiento, el poder, la justicia y la responsabilidad. En este sentido, el análisis filosófico-social es indispensable para comprender no solo qué es la IA, sino qué significa su creciente integración en nuestras sociedades y cómo debería orientarse su uso.

Desde una perspectiva ética, la IA pone en cuestión el concepto de agencia moral. ¿Puede un algoritmo tomar decisiones justas? ¿Cómo se reparte la responsabilidad cuando un sistema automatizado causa daño? Hans Jonas advertía que las tecnologías modernas requieren una ética orientada al futuro, dada su capacidad de transformar de forma irreversible las condiciones de vida humana (Jonas, 1979). Esta ética debe aplicarse especialmente a sistemas que automatizan decisiones sensibles: diagnósticos médicos, evaluaciones crediticias, predicciones policiales o selección de personal.

Uno de los aportes clave de los estudios CTS es el rechazo de la idea de que la tecnología sea un producto puramente técnico y neutral. Autores como Langdon Winner sostienen que los artefactos tienen política: incorporan decisiones sobre quién controla qué, a qué fines sirve una tecnología y quién se beneficia o queda excluido (Winner, 1980). En el caso de los sistemas de IA, al estar entrenados con datos históricos, tienden a reproducir sesgos estructurales como el racismo, el sexismo o incluso la ideología.

Otro eje del análisis social es la asimetría de poder entre quienes desarrollan IA y quienes la utilizan. Empresas como OpenAI, Google, Microsoft o Meta concentran los recursos computacionales, los datos y la capacidad de influencia cultural global. Esta concentración plantea riesgos democráticos: los modelos de lenguaje están mediando la información que consumimos, sin mecanismos claros de rendición de cuentas. Bruno Latour ha mostrado cómo los sistemas técnicos configuran relaciones sociales, y viceversa: no existen tecnologías puramente “objetivas”, sino redes híbridas de actores humanos y no humanos (Latour, 2005).

El pensamiento filosófico-social también puede cumplir una función propositiva al permitirnos imaginar futuros alternativos a los que impone el mercado. ¿Puede diseñarse una IA al servicio del bien común? ¿Es posible desarrollar modelos abiertos, auditables, participativos? Nick Bostrom, desde una perspectiva tecnofilosófica, plantea los riesgos existenciales de una superinteligencia no alineada con los valores humanos (Bostrom, 2014). Frente a esta visión apocalíptica, otras corrientes abogan por el diseño ético y democrático de tecnologías, que atiendan a criterios de justicia, sostenibilidad y equidad.

Incorporar el análisis filosófico y social a la discusión sobre IA no debería ser una necesidad política y ética. Sin él, corremos el riesgo de naturalizar como “progreso” lo que en realidad son decisiones cargadas de valores, muchas veces invisibles. Los estudios CTS ofrecen las herramientas críticas para cuestionar, democratizar y reorientar el rumbo de la inteligencia artificial hacia una sociedad más justa y consciente de sus propios medios.

4 CONTEXTUALIZACIÓN DE LA IA

La inteligencia artificial no es magia, aunque a menudo lo pueda parecer. Como señaló Arthur C. Clarke, “toda tecnología suficientemente avanzada es indistinguible de la magia”, pero en el fondo, la IA no es más que software. Instrucciones ejecutadas por una máquina de Turing. Para comprender sus riesgos es importante distinguir entre el software tradicional y el basado en aprendizaje automático, el motor actual de la inteligencia artificial moderna.

Durante décadas, el desarrollo de software se ha basado en algoritmos deterministas: secuencias de reglas explícitas escritas por programadores. Estos sistemas operan con

precisión y previsibilidad. Sabemos qué hacen porque alguien diseñó cada paso, y por eso no nos sorprende que funcionen. Navegadores web, editores de texto o aplicaciones móviles siguen instrucciones claras creadas de forma manual.

Sin embargo, la revolución reciente en IA se ha producido gracias a los algoritmos probabilísticos y el aprendizaje automático. En lugar de programar explícitamente cada regla, se entrena un modelo con grandes volúmenes de datos. El sistema aprende patrones estadísticos que luego aplica a nuevos datos para hacer predicciones. A este proceso de aprendizaje se le llama entrenamiento, y al uso posterior del modelo para hacer predicciones se le denomina inferencia.

La diferencia es fundamental. En el software tradicional, las reglas están dadas. En la IA, las reglas emergen de los datos. Esto implica que el comportamiento del sistema depende en gran medida de los datos usados durante el entrenamiento. Además, muchos de estos modelos, especialmente las redes neuronales profundas, son altamente complejos y opacos. Los modelos más avanzados tienen cientos de miles de millones de parámetros, lo que los convierte en auténticas cajas negras que dan resultados útiles, pero no sabemos exactamente cómo llegan a ellos.

De manera que la inteligencia artificial no piensa ni decide como los humanos. Aprende de los datos, replica patrones estadísticos y opera bajo lógicas probabilísticas. Comprender esta base es esencial para identificar y mitigar los problemas y desafíos que plantea su adopción masiva, sin confundirnos con la ciencia ficción.

5 COSTE MARGINAL CERO

Una de las principales características del software —y, por extensión, de los sistemas de inteligencia artificial— es que permite crear productos cuyo coste marginal tiende a cero. Es decir, una vez realizada la inversión inicial en desarrollo, entrenamiento o diseño, replicar el producto o servicio digital incurre en un coste prácticamente nulo, independientemente de cuántas veces se reproduzca. Esta propiedad, que no se da en sectores industriales tradicionales donde cada unidad producida requiere nuevos insumos materiales, convierte al software en un bien de rendimiento creciente y escalabilidad extrema.

Históricamente, esta lógica ha transformado la economía global. Desde la comercialización del software para ordenadores personales, liderada por Microsoft en los años 80 y 90, hasta la irrupción de Internet y las plataformas digitales en los 2000 con empresas como Google, Amazon o Facebook, el valor se ha desplazado progresivamente desde los bienes físicos hacia los servicios digitales. Esta tendencia se intensifica en la década de 2020 con la popularización de modelos de inteligencia artificial de propósito general, capaces de automatizar tareas cognitivas en campos tan diversos como la atención al cliente, el diagnóstico médico o la creación de contenidos.

El potencial transformador de esta característica —la posibilidad de ofrecer productos o servicios con un coste marginal tendente a cero— es enorme. Por primera vez en la historia, la tecnología habilita escenarios de abundancia artificial, en los que bienes tradicionalmente escasos pueden distribuirse de forma masiva con un coste próximo a cero (Riffkin, 2015).

En este nuevo paradigma, sectores como la educación, la sanidad o la movilidad podrían experimentar una disrupción radical. Un sistema de tutores inteligentes podría ofrecer educación personalizada, multilingüe y de calidad a cualquier estudiante del mundo, sin depender de la presencia de profesorado local cualificado. Herramientas médicas basadas en IA podrían realizar diagnósticos tempranos y seguimiento remoto de pacientes en zonas con escasez de personal sanitario, reduciendo así desigualdades estructurales. Redes de transporte autónomo y optimizado podrían reducir el coste de desplazamiento a fracciones mínimas, acercando la movilidad universal.

Estas aplicaciones dibujan la posibilidad de un nuevo modelo socioeconómico basado no en la gestión de la escasez —como ha sido la norma desde el surgimiento de la economía política clásica— sino en la gestión de la abundancia digital. Sin embargo, alcanzar ese potencial exige superar obstáculos estructurales como la concentración de poder económico en pocas plataformas, la desigualdad en el acceso a la infraestructura digital y los marcos legales aún anclados en paradigmas industriales. De no abordarse estos retos, el coste marginal cero podría ampliar brechas en lugar de cerrarlas.

La característica del coste marginal nulo no solo redefine las dinámicas de mercado, sino que invita a repensar las bases mismas de la organización económica y social contemporánea.

6 PROBLEMAS ACTUALES

Es importante recordar de que la inteligencia artificial no es más, ni menos, que software. La mayoría de los problemas actuales de la IA a menudo no son nuevos, sino que se ven acentuados por la capacidad del software de generar productos o servicios cuyo coste marginal tiende a cero. Esto significa que una vez desarrollado, el coste de generar copias adicionales o realizar acciones adicionales es prácticamente nulo.

En resumen, los problemas actuales de la IA, aunque a menudo preexistentes, se ven drásticamente magnificados por la capacidad de la tecnología para operar a escala y con costes marginales muy bajos. Estos problemas abarcan desde la manipulación de la información y la cohesión social hasta la erosión de la privacidad, la perpetuación de la discriminación y la creación de nuevos desequilibrios de poder.

6.1 NOTICIAS FALSAS

Uno de los riesgos más visibles y preocupantes de los modelos avanzados de inteligencia artificial es su capacidad para generar noticias falsas y otros contenidos manipulados con una apariencia extremadamente verosímil (Ford, 2021). Aunque la desinformación no es un fenómeno nuevo, la IA ha elevado el problema a una escala sin precedentes al permitir la producción automatizada, masiva y de bajo coste de textos, imágenes, vídeos y audios falsos. Esto no solo facilita la difusión de bulos, sino que reduce drásticamente la barrera técnica y económica para crear contenidos engañosos.

La generación automática de texto falso, anticipada ya en 2019 con el lanzamiento del modelo GPT-2 de OpenAI¹, ha avanzado significativamente con modelos capaces de

¹ Modelos como GPT-2 ya demostraron la capacidad de continuar un texto a partir de un título o unas primeras frases, generando contenido que, aunque no era veraz, sí era verosímil. En 2019,

redactar artículos de apariencia profesional en segundos. Esta capacidad se ha extendido al plano audiovisual con la proliferación de deepfakes, que permiten manipular la voz, la imagen y los gestos de personas reales con un alto grado de realismo (Suleyman, 2023). Así, la IA no solo puede inventar narrativas, sino también atribuir las falsamente a figuras reconocibles, erosionando la confianza en la información y en la autenticidad de las fuentes.

Un caso especialmente preocupante es el de la suplantación de identidad mediante contenido sintético. Aunque las personas públicas —políticos, celebridades, empresarias— son las víctimas más visibles, suelen contar con recursos legales y técnicos para responder. Por el contrario, las personas anónimas carecen de protección efectiva ante la distribución de contenido falso que utiliza su imagen o su voz². El hecho de que estas herramientas estén ya disponibles en móviles y aplicaciones accesibles multiplica el riesgo de daños personales, reputacionales o incluso judiciales (Ford, 2021).

6.2 POLARIZACIÓN SOCIAL

Los sistemas de inteligencia artificial pueden contribuir a la polarización social. Un problema que ya existía, pero que se ve acentuado por la aparición de sistemas de recomendación automáticos. Estos modelos de recomendación, inicialmente desarrollados para gestionar catálogos inmensos de contenido en plataformas como Netflix o Amazon, se han extendido a las redes sociales.

Igual que un modelo de recomendación acerca a un usuario los libros o películas más afines a sus gustos, en las redes sociales prioriza y acerca al usuario las noticias y comentarios que es más propenso a consumir. Esto alimenta su sesgo de confirmación, que hace más probable que uno crea algo que ya piensa. La consecuencia es la creación de cámaras de eco, donde las personas que piensan de manera similar se refuerzan mutuamente en sus opiniones, compartiendo contenidos que confirman sus puntos de vista³.

Esto ha llevado a un auge en la distribución de noticias falsas y de opiniones polarizadas (Coeckelbergh, 2022). Un ejemplo sorprendente es el terraplanismo, un planteamiento que, paradójicamente, hace 50 años habría sido muy difícil que mucha gente creyera, pero que ahora, a través de estos recomendadores en redes sociales, llega masivamente a las personas propensas a creerla⁴.

6.3 PRIVACIDAD

Los sistemas de IA pueden potenciar los problemas de privacidad al permitir la inferencia de información privada a partir de datos que, en apariencia, no son sensibles o no se

OpenAI consideró que este modelo era "demasiado peligroso" para publicarlo abiertamente. <https://www.theguardian.com/technology/2019/feb/14/elon-musk-backed-ai-writes-convincing-news-fiction>

² Un ejemplo de esto fue el desagradable caso de desnudos falsos de menores en Almendralejo. https://www.eldiario.es/extremadura/policia-identifica-extremadura-personas-usaron-fotos-menores-desnudadas-inteligencia-artificial_1_10523765.html

³ El documental "The Social Dilemma" refleja este problema. <https://thesocialdilemma.com/>

⁴ Desde 2016 hay un resurgir de personas que creen en el terraplanismo. <https://www.theguardian.com/science/2016/jan/20/flat-earth-believers-youtube-videos-conspiracy-theorists>

consideran explícitamente privados. Estos algoritmos tienen la capacidad de extraer información de datos de una manera que a veces ni siquiera los humanos podrían.

Un caso sorprendente es el de los modelos de visión que diagnostican retinopatía diabética a partir de un fondo de ojo. Estos modelos no solo diagnostican la enfermedad, sino que también son capaces de determinar si el ojo pertenece a un hombre o a una mujer, algo que los oftalmólogos humanos no saben hacer. Esto muestra cómo los algoritmos de IA pueden extraer información privada (género, en este caso) de un dato médico que aparentemente no contiene esa información.

Otro caso, que nos acerca a una problemática más generalizada, se muestra con las plataformas de streaming como Netflix, que utilizan datos sobre qué películas y series ve el usuario y cómo las ve para recomendar contenido. Es probable que estos datos correlacionen con información privada sensible, como el género, la etnia o la identidad sexual. Aunque la empresa no obtenga intencionadamente esta información sensible, los datos sobre hábitos de consumo que el usuario cede pueden ser utilizados para inferirla como se ha reportado en alguna ocasión en redes sociales⁵.

Esta capacidad de inferir información privada de datos aparentemente inocuos representa un riesgo significativo en cuanto que son mecanismos opacos que socaban la autonomía (Floridi, 2021).

6.4 DISCRIMINACIÓN

La discriminación surge principalmente porque los sistemas de inteligencia artificial, en particular los modelos que se crean a partir de datos existentes, pueden perpetuar y amplificar los sesgos que ya existen en esos datos (Coeckelbergh, 2022). Como los sesgos no son objetivos, sino que son relativos al observador, no es posible crear un sistema de inteligencia artificial completamente libre de ellos. Si los datos utilizados para entrenar un modelo contienen un sesgo que genera discriminación contra las personas, el modelo de IA lo aprenderá, lo aplicará, y lo perpetuará.

Un ejemplo de este tipo de sesgo aparece en Gender Shades⁶, un estudio sobre sistemas comerciales de reconocimiento facial que encontró que estos sistemas fallan más en mujeres negras que en hombres blancos (Buolamwini & Gebru, 2018). Algo que puede tener graves consecuencias si estos sistemas se usan en aplicaciones críticas, como coches autónomos, donde un sesgo podría tener consecuencias graves como atropellos diferenciados por raza o género.

El problema radica en que el sesgo inherente a los datos de entrenamiento se traslada al modelo de IA, haciendo que este funcione peor o tome decisiones injustas para ciertos grupos, perpetuando así la discriminación existente en la sociedad. Este problema se agrava al juntarse con el problema de la privacidad, porque la información que se va a utilizar puede no ser evidente en los datos.

⁵ El usuario Lukas Thoms reportó un orden diferente de la serie “Love, Death + Robots” para heterosexuales y para homosexuales.

<https://www.lavanguardia.com/series/netflix/20190321/461152375897/netflix-desmiente-cambia-orden-episodios-love-death-robots-orientacion-sexual.html>

⁶ Gender Shades. <http://gendershades.org/>

6.5 VIGILANCIA MASIVA

Los sistemas de IA facilitan la vigilancia masiva, ya que, gracias a su coste marginal que tiende a cero, reducen drásticamente el coste de procesar y analizar grandes cantidades de datos de vigilancia, algo que antes era prohibitivamente caro.

Un ejemplo inicial de esto son las tiendas automatizadas sin cajas, como las que probó Amazon en algunas ciudades de EEUU. Utilizan cámaras para rastrear lo que los clientes cogen, permitiendo que se les cargue automáticamente en su tarjeta. Implica un sistema de vigilancia constante dentro de la tienda que se puede considerar legítimo.

Un caso más preocupante es la implementación a nivel estatal, como el sistema de rating social en China⁷. Mediante un despliegue masivo de cámaras en las ciudades, los sistemas de IA pueden procesar la información visual para, por ejemplo, quitar o sumar puntos a un ciudadano si tira un papel al suelo o a la papelera (Coeckelbergh, 2022).

Esta capacidad de vigilancia constante y automatizada plantea preocupaciones sobre la privacidad y el control, dejando de ser escenarios de ciencia-ficción (Coeckelbergh, 2022) y acercándose al concepto distópico del "Gran Hermano" de Orwell.

6.6 ASIMETRÍAS DE PODER

Las asimetrías de poder, tanto entre individuos como entre instituciones, empresas y Estados, se producen cuando una parte dispone de tecnologías automatizadas, escalables y de coste marginal cercano a cero, mientras que la otra carece de medios similares y debe responder de forma manual, costosa y limitada. Ejemplos aparentemente banales, como las llamadas automáticas de telemarketing, ilustran cómo actores con acceso a sistemas de IA pueden saturar a millones de personas con un esfuerzo mínimo, mientras que el coste en tiempo, atención e incomodidad recae enteramente sobre los ciudadanos. Este mismo patrón puede reproducirse a mayor escala en ámbitos como la administración pública, la justicia, la publicidad, la seguridad o la toma de decisiones políticas.

Sin embargo, el problema va más allá de una distribución desigual de esfuerzos: la IA puede reforzar y amplificar estructuras de poder existentes, profundizando desigualdades económicas, sociales y políticas. Las grandes plataformas tecnológicas concentran no solo los beneficios económicos del desarrollo de IA, sino también el acceso al conocimiento, los datos y la capacidad de influencia sobre el comportamiento humano, dando lugar a lo que Shoshana Zuboff denomina capitalismo de vigilancia, un modelo económico basado en la extracción, predicción y monetización de datos personales (Zuboff, 2019). Este fenómeno agrava la brecha entre quienes diseñan y controlan la tecnología y quienes la usan o la padecen. La falta de transparencia en los algoritmos y modelos, la dificultad para atribuir responsabilidades cuando algo falla y la opacidad de los sistemas automatizados generan una "brecha de conocimiento" que deja a buena parte de la población al margen del control efectivo sobre las decisiones que les afectan.

⁷ La vigilancia digital es omnipresente en China.

<https://theconversation.com/digital-surveillance-is-omnipresent-in-china-heres-how-citizens-are-coping-225628>

Además, la IA puede utilizarse como instrumento de vigilancia y control, tanto por parte de Estados como de grandes corporaciones. Desde tecnologías de reconocimiento facial hasta asistentes personales, pasando por sistemas de recomendación y segmentación publicitaria, los entornos digitales están diseñados para guiar decisiones, moldear preferencias y mantener la atención cautiva, a menudo sin un consentimiento informado ni mecanismos efectivos de protección. Esto plantea serios desafíos para la autonomía individual, la deliberación democrática y la libertad de pensamiento.

También se observan formas más sutiles de manipulación algorítmica, como las pequeñas intervenciones (*nudges*) personalizadas que alteran el entorno de elección, reduciendo la agencia de los usuarios (Coeckelbergh, 2022). Estas prácticas, junto con el sesgo algorítmico y la economía de la atención, configuran una arquitectura de poder que opera invisiblemente sobre nuestras decisiones cotidianas.

Si no se gestionan adecuadamente, estas asimetrías pueden derivar en nuevas formas de exclusión y dominación, donde la inteligencia artificial se convierte en una herramienta al servicio de los intereses de unos pocos, en detrimento del bienestar colectivo, la equidad y los principios democráticos. Frente a ello, se hace imprescindible una reflexión ética profunda y una regulación firme que garantice transparencia, equidad, rendición de cuentas y redistribución del poder tecnológico.

6.7 DECISIONES AUTOMATIZADAS

La capacidad de los sistemas de IA para tomar decisiones o realizar acciones de forma automática basándose en datos es un factor subyacente que contribuye a otros problemas, como la discriminación, la privacidad y las asimetrías.

El problema en este caso es que cuando se introduce un sistema automático en un proceso de decisión, con un *human in the loop* porque no se quiere perder el control humano de esa decisión, en realidad, en el medio plazo, se está delegando la decisión al sistema automático (Floridi, 2021).

El motivo es que las personas no estamos preparadas para supervisar un sistema con una alta precisión y detectar una ocasión entre miles en la que comete un error. Si un médico o un juez tienen que supervisar el diagnóstico o el veredicto de un sistema de inteligencia artificial para evitar errores, y el sistema necesita corrección de forma esporádica, es probable que la persona acabe ratificando de manera acrítica las decisiones del sistema.

De esta manera, si no se quiere delegar decisiones a sistemas de inteligencia artificial, es probable que sea mejor no incluirlos en *loop*.

7 POSIBLES DESAFÍOS FUTUROS

Aunque muchas de las implicaciones de la inteligencia artificial ya son visibles en el presente, los escenarios más transformadores aún están por definirse. La velocidad del avance tecnológico, combinada con la creciente disponibilidad de datos y potencia computacional, abre un horizonte de posibilidades que plantea tanto expectativas ambiciosas como riesgos profundos.

A diferencia de otras tecnologías del pasado, la IA no solo automatiza tareas físicas o repetitivas, sino que empieza a intervenir en procesos cognitivos, comunicativos y decisionales tradicionalmente considerados exclusivos de los seres humanos. Esto implica potenciales disrupciones en ámbitos como el mercado laboral, la producción de conocimiento, las dinámicas políticas y la estructura misma del poder social.

En este bloque analizo algunos de los principales desafíos futuros asociados al desarrollo y uso extensivo de la IA. No se trata de hacer predicciones deterministas, sino de identificar tensiones emergentes que podrían intensificarse en los próximos años si no se abordan desde marcos éticos, políticos y regulatorios sólidos. Cada uno de estos desafíos combina componentes técnicos con dimensiones filosófico-sociales que requieren atención desde múltiples disciplinas.

7.1 FUTURO DEL EMPLEO

La inteligencia artificial tiene el potencial de provocar una transformación económica comparable a las grandes revoluciones industriales. Mientras que la primera revolución (siglos XVIII y XIX) sustituyó la fuerza bruta humana y animal mediante la máquina de vapor, la segunda (finales del XIX e inicios del XX) trajo la producción en cadena impulsada por la electricidad, y la tercera (finales del XX) digitalizó procesos mediante los ordenadores, la actual disrupción implica la automatización de tareas para las que las personas necesitamos pensar. Esto la convierte en una tecnología con un potencial impacto sin precedentes en el mercado laboral.

Históricamente, las revoluciones industriales han visto la destrucción de puestos de trabajo, pero también la aparición de otros nuevos. Sin embargo, si estos sistemas de IA pueden automatizar tareas que requieren pensar, es difícil asegurar que los nuevos trabajos que aparezcan estarán protegidos de ser también automatizados.

Estudios recientes refuerzan esta preocupación. La OCDE advierte que una proporción significativa de empleos en economías avanzadas contiene tareas susceptibles de automatización parcial o total, especialmente en sectores administrativos, jurídicos, financieros y técnicos (OECD, 2023).

Además, de manera contraintuitiva, los empleos mejor remunerados y con mayor formación formal están más expuestos a la automatización que otros menos cualificados. Esta paradoja, conocida como la paradoja de Moravec, señala que tareas fáciles para los humanos (como el cuidado o la empatía) son difíciles para la IA, mientras que tareas "difíciles" (como el cálculo o la memorización de datos) resultan relativamente fáciles para las máquinas. A esto se suma la aparición de nuevos modelos de lenguaje que amenazan puestos de trabajo con una necesidad de habilidades con el lenguaje o capacidades de razonamiento limitado (Eloundou, 2023). Esta sustitución de

capacidades humanas por "fuerza laboral digital" no solo desplaza trabajadores, sino que pone en cuestión el modelo mismo de redistribución del valor (Ford, 2021). Si gran parte de la producción de riqueza futura recae en sistemas autónomos operados por una minoría, será necesario repensar mecanismos de reparto como impuestos sobre la automatización o rentas básicas.

Algunos informes estiman que hasta 300 millones de empleos⁸ podrían verse afectados globalmente (Goldman Sachs, 2023). Aunque esa cifra es difícil de precisar, da una idea de la magnitud del cambio. Además, el impacto no será homogéneo y las economías avanzadas serán las más vulnerables (IMF, 2024), precisamente por su alto grado de digitalización y especialización.

En este escenario, la IA no debe entenderse solo como una herramienta productiva, sino como una nueva forma de trabajo autónomo, virtual y escalable, capaz de reemplazar funciones humanas a gran escala. Si no se anticipan sus efectos mediante políticas activas de redistribución, formación y regulación, es posible que su impacto sobre el empleo desemboque en una crisis del modelo socioeconómico actual.

7.2 DEMOCRACIA

La democracia se sostiene, en gran medida, sobre un tejido de confianza pública. Confianza en que la información es verificable, que las instituciones actúan con legitimidad y que los procesos electorales son justos. Sin embargo, el desarrollo y uso creciente de sistemas de inteligencia artificial —especialmente en contextos como la producción de información, la vigilancia y la toma de decisiones automatizadas— podría socavar las bases de esa confianza (Coeckelbergh, 2022).

Uno de los fenómenos más preocupantes es el de la confusión generalizada. La abundancia de noticias falsas, audios manipulados o imágenes generadas puede llevar a que los ciudadanos ya no sean capaces de distinguir lo verdadero de lo falso. Esta saturación de desinformación no solo induce a error, sino que puede sembrar una sospecha permanente. Incluso frente a contenidos reales y verificables, las personas comienzan a dudar de la autenticidad.

Cuando esta dinámica se extiende al terreno institucional, los efectos pueden ser graves. Si los ciudadanos creen que las decisiones políticas están mediadas por algoritmos sesgados o que las elecciones pueden ser manipuladas por campañas de desinformación automatizadas, se deslegitima la autoridad democrática. A esto se suma la dificultad para identificar responsabilidades en entornos dominados por sistemas opacos o "cajas negras", donde ni siquiera las propias instituciones pueden explicar completamente el funcionamiento de los algoritmos que emplean.

En este contexto, la pérdida de confianza ya no se limita a actores individuales o a partidos concretos, sino que afecta al ecosistema institucional en su conjunto: parlamentos, medios de comunicación, sistemas judiciales, fuerzas de seguridad y organismos públicos se ven arrastrados por una ola de escepticismo estructural. Esta desconfianza desmoviliza, polariza y abre la puerta a discursos populistas que se

⁸ En la UE hay unos 200 millones de puestos de trabajo

presentan como alternativa a unas instituciones percibidas como incompetentes o capturadas por intereses ocultos.

El riesgo es que entremos en una fase de cinismo tecnológico, donde la ciudadanía se sienta incapaz de controlar o comprender los sistemas que condicionan su vida pública, lo que debilita no solo la participación democrática, sino también la cohesión social.

7.3 CAMBIO CLIMÁTICO

La IA tiene el potencial de acentuar el problema del cambio climático debido al alto consumo energético requerido para crear y ejecutar los grandes modelos, especialmente durante el entrenamiento (aunque últimamente se ha incrementado significativamente el consumo en la inferencia con los agentes autónomos).

Estudios recientes han indicado las consecuencias amplias del rápido desarrollo y entrenamiento de modelos de IA, mostrando un impacto climático sustancial (CAAD, 2024). Estas huellas de carbono se asocian principalmente con el rápido desarrollo y entrenamiento de algoritmos de IA con poca consideración por el impacto general en el sistema de la Tierra.

Algunos, como Eric Schmidt, antiguo CEO de Google, argumentan que la IA podría ser la solución al cambio climático, por lo que se debe seguir potenciando el desarrollo de la IA, aunque en el corto plazo empeore la situación del cambio climático⁹. El riesgo está en que la solución al cambio climático es un futuro hipotético, mientras que el consumo energético es un hecho actual.

La ética de la IA y la política deben abordar cómo mitigar este impacto ambiental negativo, considerando que el impacto en términos de huella de carbono y consecuencias medioambientales es políticamente problemático tanto para los humanos como para los no humanos.

7.4 VINCULO EMOCIONAL

Existe la preocupación sobre la posibilidad de que las personas desarrollen vínculos emocionales con sistemas de inteligencia artificial que, entre otros problemas, nos pueden dejar expuestos a la manipulación (Coeckelbergh, 2022). La película Her muestra un ejemplo de este escenario.

No es un escenario intrínsecamente peligroso, pero si abre un abanico de potenciales problemas nuevos que habrá que gestionar. Es verdad que también puede ser un camino que mitigue otros problemas actuales como el de la soledad (Coeckelbergh, 2022).

7.5 CAPACIDADES COGNITIVAS

Otro posible desafío es que la dependencia de sistemas de IA para realizar actividades cognitivas pueda afectar negativamente en el desarrollo o mantenimiento de nuestras propias capacidades intelectuales.

⁹ La declaración de Eric Schmidt fue en un evento y la recoge este artículo. “Prioritising AI over climate change would be catastrophic”.

<https://techinformed.com/prioritising-ai-over-climate-change-would-be-catastrophic/>

Un ejemplo de cómo la tecnología puede afectar al desarrollo del cerebro está en el uso del GPS y el desarrollo del hipocampo¹⁰, o cómo los móviles en general están cambiando nuestro cerebro afectando a capacidades como la atención, la concentración, o el recordar números de teléfono¹¹.

De igual manera, por tanto, el uso de herramientas de IA para realizar tareas cognitivas, aunque en una primera instancia potencien nuestras capacidades cognitivas, podría hacernos retroceder en el desarrollo cerebral y el de esas capacidades (Gerlich, 2025).

7.6 ARMAS AUTÓNOMAS

En este caso, el problema principal radica en la asimetría de costes en la guerra. Si un bando utiliza robots autónomos con un coste marginal que tiende a cero por baja humana, mientras el otro lado usa personas, el incentivo del que tiene un coste tan bajo para evitar la guerra será mucho menor que actualmente (Coeckelbergh, 2022).

Es decir, el riesgo de las armas autónomas es considerable no solo por su potencial letal, sino por la asimetría de costes que reduce la barrera para iniciar conflictos, la distancia moral que facilita matar, la asimetría en términos de riesgo y vigilancia, la inherente "brecha de responsabilidad", y el riesgo de escalada incontrolada (Ford, 2021). Estos factores combinados hacen que sea un problema ético y estratégico difícil de evitar o resolver sin una reflexión profunda y, posiblemente, acuerdos internacionales como ocurrió con el armamento nuclear después de la II Guerra Mundial.

7.7 SUPERINTELIGENCIA

Si bien este es un tema de debate y preocupación generalizada, escenarios de un sistema de IA tomando consciencia (como Skynet en el mantra de Terminator) son actualmente ciencia ficción y no parecen posibles con la tecnología actual.

Es recomendable evitar antropomorfizar a los sistemas de IA, es decir, atribuirle características humanas como consciencia, autonomía o agencia que no poseen para evitar un sentimiento colectivo de amenaza de estas herramientas que puede derivar en un miedo infundado, más cercano a la superstición.

Sin embargo, aunque no hay indicios de ningún posible desarrollo de una superinteligencia, la magnitud del riesgo, con potencial de catástrofe existencial (Bostrom, 2014), sí hacen razonable la investigación en la línea del alineamiento y explicabilidad de los sistemas de IA.

Así, el desafío es cómo establecer barreras y reglas para que sistemas cada vez más capaces y autónomos no se desvíen o lleguen a conclusiones dañinas. Esto busca evitar que la IA actúe de formas no deseadas por falta de "sentido común" o alineamiento con los objetivos humanos (Bostrom, 2014).

¹⁰ The Hippocampus and Entorhinal Cortex Encode the Path and Euclidean Distances to Goals during Navigation.

[https://www.cell.com/current-biology/fulltext/S0960-9822\(14\)00526-0](https://www.cell.com/current-biology/fulltext/S0960-9822(14)00526-0)

¹¹ How mobile phones have changed our brains.

<https://www.bbc.com/future/article/20230403-how-cellphones-have-changed-our-brains>

Los principales mecanismos de riesgo identificados son la posibilidad de un crecimiento explosivo de la inteligencia, la obtención de una ventaja estratégica decisiva, la falta de alineación de objetivos con los intereses humanos, la convergencia instrumental que puede llevar a comportamientos engañosos (el giro traicionero), la extrema dificultad de mantener el control sobre una entidad superinteligente y el riesgo de una carrera armamentística (Bostrom, 2014). Estos riesgos se agravan por la diferencia fundamental en la agencia moral y el juicio ético entre humanos y máquinas, y la posibilidad de que la búsqueda de la superinteligencia esté impulsada por valores que pueden ser perjudiciales para la mayoría y para el futuro.

8 CONTROL Y REGULACIÓN

Continuando con esta exploración de los posibles desafíos futuros relacionados con la inteligencia artificial, abordo ahora cómo se puede tomar control y que tendencias regulatorias se están planteando.

La creciente capacidad y potencial impacto de la inteligencia artificial plantean la necesidad de estrategias para controlarla y regular sus usos.

8.1 ALINEAMIENTO

Una línea de trabajo e investigación dentro de las técnicas para controlar sistemas de inteligencia artificial cada vez más capaces y autónomos es el concepto de alineamiento. El desafío es establecer barreras y reglas para asegurar que estos sistemas no se desvíen o lleguen a conclusiones dañinas. Es decir, que “captan nuestras normas y valores, entiendan lo que queremos decir o intentamos, y, sobre todo, hagan lo que queremos” (Christian, 2020).

Ante un caso como el ejemplo hipotético de un sistema de IA optimizando la producción de clips que decide eliminar a la humanidad para obtener recursos (Bostrom, 2014), que ilustra la necesidad de evitar que un sistema de IA actúe de formas no deseadas por falta de "sentido común" o alineamiento con los objetivos humanos (Christian, 2020).

Aún así, no es necesario llegar a un escenario de superinteligencia para tener que desarrollar el alineamiento de los sistemas de IA. Los problemas actuales ya requieren del desarrollo de estas técnicas, que no dejan que los sistemas de IA tengan impactos sin control, a la vez que son menos neutrales porque, en definitiva, el alineamiento implica imponer sesgos explícitos a estos sistemas.

8.2 TRANSPARENCIA ALGORÍTMICA

Un problema relevante de los modelos actuales es su naturaleza de "caja negra", que implica una falta de transparencia (Floridi, 2023). Especialmente en redes neuronales profundas y modelos de lenguaje, no se sabe con exactitud las razones por las que un modelo da una determinada respuesta. Simplemente se obtiene el resultado. Dada la descomunal cantidad de parámetros y pesos, no se entiende completamente el funcionamiento interno o el "racional" detrás de una respuesta concreta. Actualmente, por ejemplo, aunque por supuesto sí se conoce todo el detalle de cómo son entrenados los modelos de lenguaje y cómo funciona su inferencia, no se sabe exactamente cómo pueden manejar la semántica del lenguaje como lo hacen.

Ante esta falta de transparencia inherente, desde una perspectiva técnica, se está trabajando en el área de causalidad y, de forma destacada, en Inteligencia Artificial Explicable (XAI). El objetivo de XAI no es que el modelo no sea una caja negra, sino que, una vez obtenida la respuesta, se pueda explicar el proceso que llevó a ella (Floridi, 2021).

La dificultad en identificar sesgos complejos en los modelos de IA subraya la necesidad de métodos para evaluar y comprender el comportamiento de estos sistemas, lo que se relaciona con la necesidad de transparencia y explicabilidad.

Para este aspecto, la postura del nuevo reglamento de IA de la UE, que apunta a que marcará tendencia global, es el de prohibir en ciertos escenarios, como la denegación de créditos bancarios, el uso de modelos de caja negra, de los que no se extraer razones para una denegación.

8.3 REGULACIÓN Y GOBERNANZA GLOBAL

Problemas ya considerados como la asimetría, donde una empresa con IA tiene mucha información sobre un individuo, pero este no tiene información sobre la empresa, o la capacidad de realizar vigilancia masiva a bajo coste, sugieren áreas donde la política, a través de la regulación, podría ser necesaria para mantener las normas democráticas (Coeckelbergh, 2022).

Así, el debate sobre la regulación abarca varias posturas actualmente en un rango muy amplio, que va desde posiciones en las que se considera que no hacen falta nuevas regulaciones, hasta posturas especialmente estrictas que llegan a considerar las sanciones a países que desarrollen ciertos tipos de IA.

La más ligera de estas posturas defiende que no es necesaria nueva legislación porque los problemas que trae la IA ya existían y la regulación actual es suficiente (Glauner, 2021). Se argumenta que, si algo está prohibido, como suplantar a alguien, sigue estándolo independientemente de si se usa IA o no. Sin embargo, el efecto del coste marginal puede tener un impacto en la escala de las consecuencias de esa acción que haga necesaria una actualización de la regulación existente. Por ejemplo, de forma general los juicios son públicos, pero cuando surge la posibilidad de la retransmisión televisiva, la regulación requiere una adaptación y estos pasan a ser públicos solo en el caso de asistencia física, y requieren del permiso del tribunal para ser retransmitidos.

Otra postura aboga por prohibir la explotación de datos personales porque los convierte en una materia prima para la obtención de beneficios (Zuboff, 2019), de forma similar a como se prohíbe el comercio de órganos humanos. Esta visión sugiere que el Estado debería tener el control de este tipo de datos para ciertos usos, como modelos de salud. No es una postura que afecte directamente al funcionamiento de los sistemas de IA, pero sí afecta a la creación de muchos de ellos.

Existe también la defensa de regular determinados escenarios de forma estricta, comparable al sector farmacéutico. Esto implicaría que los sistemas de IA que se quieran poner en producción o comercializar deban pasar largos procesos administrativos de aprobación antes de su uso público. Este escenario, defendido por el reglamento de IA de la UE (UE, 2024) para los casos de sistemas de alto riesgo, puede frenar la aparición de

nuevas empresas y beneficiar a las grandes que sí pueden asumir el coste de estos procesos.

Una postura más extrema, basada en la posibilidad de problemas existenciales, plantea tratar la IA como se trata el armamento nuclear¹². En este caso se defiende controlar la comercialización de hardware clave como las GPUs y, en un escenario extremo, incluso considerar la posibilidad de iniciar una guerra con un país que desarrolle IA fuera de control o acuerdos internacionales.

8.4 EDUCACIÓN Y CONCIENCIACIÓN SOCIAL

Desde una perspectiva socio-cultural, la discusión de los problemas actuales como las noticias falsas o la polarización implícitamente destaca la necesidad de que la sociedad esté informada y sea crítica ante el contenido generado o filtrado por los sistemas de IA.

Además, la tendencia a presentar la IA como una fuerza inevitable y autónoma, dotada de racionalidad propia, responde a lo que algunos autores han denominado el mito tecnocrático: la creencia de que las soluciones tecnológicas son intrínsecamente objetivas, necesarias y deseables, independientemente del contexto social en que emergen. Este imaginario desactiva el debate político al sugerir que la única opción razonable es adaptarse al “avance” tecnológico. Sin embargo, los artefactos no son neutrales (Winner, 1980). Desde esta perspectiva, la IA no solo resuelve problemas técnicos, sino que organiza relaciones de poder, redefine la agencia y configura quién puede decidir qué es progreso y para quién. El mito tecnocrático, al encubrir estas dimensiones, actúa como una forma de despolitización que impide imaginar futuros alternativos.

Así, surge una suerte de antropomorfización de la IA que incluso la sitúa en un plano superior. Hay un imaginario colectivo, alimentado por ciencia ficción (Coeckelbergh, 2022) y por nuestra tendencia natural a atribuir estados mentales a otras criaturas (Bostrom, 2014). Es decir, a atribuir características humanas como consciencia, autonomía o agencia a los sistemas de IA que no parece que posean.

Un ejemplo que puede ser representativo de este efecto es el titular “Una IA descubre más de 300 nuevas imágenes en Nazca”¹³, que parece que sugiere cierto tipo de agencia o autonomía en el sistema informático, frente a “Descubren más de 300 geoglifos con inteligencia artificial”¹⁴, que refleja mejor el rol de herramienta del sistema de IA en el descubrimiento.

Se debería tomar consciencia de esta tendencia y tener la precaución de no alimentar este imaginario colectivo, especialmente en las guías de estilo de los medios de comunicación.

¹² Eliezer Yudkowsky defendía esta postura en esta carta abierta en Time.
<https://time.com/6266923/ai-eliezer-yudkowsky-open-letter-not-enough/>

¹³ Una IA descubre más de 300 nuevas imágenes en Nazca. Muy Interesante.
<https://www.muyinteresante.com/tecnologia/66251.html>

¹⁴ Descubren más de 300 geoglifos con inteligencia artificial en la Pampa de Nasca, Perú. RTVE.
<https://www.rtve.es/noticias/20240924/mas-300-nuevos-geoglifos-descubiertos-inteligencia-artificial-pampa-nasca-peru/16261052.shtml>

9 CONCLUSIÓN

Aunque la IA entraña riesgos, la historia demuestra que la sociedad ha sido capaz de gestionarlos. Muchas de las métricas que se pueden considerar como reflejo del progreso de la humanidad han mejorado significativamente en los últimos 200 años (Roser, 2016), coincidiendo con una gran concentración de desarrollo tecnológico. Sin embargo, el cambio climático ha sido una contrapartida importante recordándonos que el progreso técnico no siempre implica mejoras equitativas o sostenibles. Por ello, la identificación de los problemas actuales y los desafíos futuros parece imprescindible para abordar la incertidumbre extrema que plantea la IA y así aprovechar su potencial transformador de forma justa y consciente.

Desde una mirada CTS, la IA no puede entenderse como una tecnología aislada, sino como una construcción sociotécnica que encarna valores, intereses y visiones del mundo. Esta perspectiva permite descubrir las asimetrías de poder que se ocultan bajo los argumentos alrededor de la eficiencia o la innovación. En lugar de asumir que los riesgos de la IA son simples externalidades técnicas, el enfoque CTS muestra que muchos de estos problemas —como la discriminación algorítmica, la vigilancia o la pérdida de autonomía— son el resultado de decisiones políticas, económicas y culturales.

Adoptar esta mirada implica también politizar la tecnología, abriendo el debate sobre quién decide qué IA se desarrolla, con qué fines y bajo qué criterios de justicia. Frente a la concentración del poder tecnológico en pocas manos, los estudios CTS proponen democratizar el diseño y la gobernanza de la IA, integrando saberes diversos y recuperando el control social sobre los medios técnicos. Esta orientación contribuye a orientar el desarrollo de esta tecnología hacia una sociedad más equitativa, sostenible y reflexiva, donde la inteligencia, además de ser artificial, sea compartida, deliberada y justa.

10 BIBLIOGRAFÍA

- Acemoglu, D., & Restrepo, P. (2020). Robots and Jobs: Evidence from US Labor Markets. *Journal of Political Economy*. <https://doi.org/https://doi.org/10.1086/705716>
- Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? *ACM Conference on Fairness, Accountability, and Transparency*. <https://doi.org/https://doi.org/10.1145/3442188.3445922>
- Bengio, Y. (2025). *International AI Safety Report*. Obtenido de <https://www.gov.uk/government/publications/international-ai-safety-report-2025>
- Bostrom, N. (2014). *Superintelligence: Paths, Dangers, Strategies*.
- Buolamwini, J., & Gebru, T. (2018). Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. *Proceedings of the 1st Conference on Fairness, Accountability and Transparency*. Proceedings of Machine Learning Research. Obtenido de <https://proceedings.mlr.press/v81/buolamwini18a.html>
- CAAD. (2024). *Artificial intelligence threats to climate change*. Climate Action Against Disinformation. Obtenido de <https://caad.info/analysis/reports/the-ai-threats-to-climate-change/>
- Christian, B. (2020). *The Alignment Problem*.
- Coeckelbergh, M. (2020). *AI Ethics*. MIT Press.
- Coeckelbergh, M. (2022). *Robot Ethics*. MIT Press.
- Coeckelbergh, M. (2022). *The Political Philosophy of AI: An Introduction*. Polity Press.
- Eloundou, T. (2023). GPTs are GPTs: An Early Look at the Labor Market Impact Potential of Large Language Models. Obtenido de <https://arxiv.org/abs/2303.10130>
- Elsner, M., Atkinson, G., & Zahidi, S. (2025). *Global Risks Report 2025*. World Economic Forum.
- Floridi, L. (2021). *Ethics, Governance, and Policies in Artificial Intelligence*. Springer.
- Floridi, L. (2023). *The Ethics of Artificial Intelligence. Principles, Challenges, and Opportunities*.
- Ford, M. (2021). *Rule of the Robots*.
- Gerlich, M. (2025). *AI Tools in Society: Impacts on Cognitive Offloading and the Future of Critical Thinking*. Societies. <https://doi.org/https://doi.org/10.3390/soc15010006>
- Glauner, P. (2021). An Assessment of the AI Regulation Proposed by the European Commission. En *The Future Circle of Healthcare: AI, 3D Printing, Longevity, Ethics, and Uncertainty Mitigation*. <https://doi.org/https://doi.org/10.48550/arXiv.2105.15133>
- Goldman Sachs. (2023). *The Potentially Large Effects of Artificial Intelligence on Economic Growth*. Goldman Sachs Publishing.

- IMF. (2024). *Gen-AI: Artificial Intelligence and the Future of Work*. International Monetary Fund. <https://doi.org/https://doi.org/10.5089/9798400262548.006>
- Innerarity, D. (2025). *Una teoría crítica de la inteligencia artificial*.
- Jonas, H. (1979). *El principio de responsabilidad*.
- Latour, B. (2005). *Reensamblar lo social: una introducción a la teoría del actor-red*.
- Mercier, H. (2020). *Not Born Yesterday*.
- O'Neil, C. (2016). *Weapons of Math Destruction*.
- OECD. (2023). *OECD Employment Outlook 2023: Artificial Intelligence and the Labour Market*. OECD Publishing. <https://doi.org/https://doi.org/10.1787/08785bba-en>
- Riffkin, J. (2015). *Zero Marginal Cost Society*.
- Roser, M. (2016). "The short history of global living conditions and why it matters that we know it." OurWorldinData.org. Obtenido de <https://ourworldindata.org/a-history-of-global-living-conditions>
- Suleyman, M. (2023). *The Coming Wave*.
- Sunstein, C. (2023). *Decisions About Decisions*.
- Turing, A. (1950). Computing Machinery and Intelligence. *Mind*. <https://doi.org/https://doi.org/10.1093/mind/LIX.236.433>
- UE. (2024). Reglamento de Inteligencia Artificial. Obtenido de <https://www.boe.es/buscar/doc.php?id=DOUE-L-2024-81079>
- Winner, L. (1980). Do artifacts have politics? *Daedalus*.
- Zuboff, S. (2019). *The Age of Surveillance Capitalism*.