

## Riesgos sociotécnicos y códigos deontológicos en inteligencia artificial: ¿Hasta dónde llega la ética de la tecnología?

Moreno Inglés, Aarón

**Publication date**

2025

**Document Version**

Final published version

**Published in**

Niaia. Revista Interdisciplinar sobre Formación y Resolución de Problemas Morales

**Citation (APA)**

Moreno Inglés, A. (2025). Riesgos sociotécnicos y códigos deontológicos en inteligencia artificial: ¿Hasta dónde llega la ética de la tecnología? *Niaia. Revista Interdisciplinar sobre Formación y Resolución de Problemas Morales*, (3), 169-180.

**Important note**

To cite this publication, please use the final published version (if applicable).  
Please check the document version above.

**Copyright**

Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

**Takedown policy**

Please contact us and provide details if you believe this document breaches copyrights.  
We will remove access to the work immediately and investigate your claim.

# Riesgos sociotécnicos y códigos deontológicos en inteligencia artificial: ¿Hasta dónde llega la ética de la tecnología?

Aarón Moreno Inglés.  
Universidad Técnica de Delft

A estas alturas poca gente pasa por alto que el uso de los sistemas de inteligencia artificial (IA) supone una gran cantidad de riesgos a escala social. El problema de los *deepfakes* y otras repercusiones de la IA generativa (García Haendler 2024), los sesgos racistas y machistas en la clasificación de los datos (Noble 2018; Gebru 2020; Akter *et al.* 2021), las reiteradas violaciones de privacidad y recopilación de datos de forma indebida (Bartneck *et al.* 2021), o la vigilancia masiva con motivos políticos y comerciales (Feldstein 2019) son tan solo una muestra del impacto dañino que ha ocasionado su adopción estos últimos años.

Es importante recordar también las lógicas extractivistas y neocoloniales en las que se enmarca el diseño y la producción de la IA. No nos olvidamos del uso de materiales en escasez, como el litio, con comunidades enteras trabajando en condiciones paupérrimas para extraerlos, dejando auténticos eriales tras de sí (Crawford 2021). Tampoco olvidamos el impacto medioambiental del consumo desorbitado de agua y electricidad de los procesadores de IA, superando cifras de consumo de países enteros. Igualmente recordamos los escándalos en relación al entrenamiento

de sistemas de IA o en su moderación de contenido como, por ejemplo, la gente que fue empleada en Kenia por apenas 1 dólar al día viendo contenido explícito (realmente perturbador, con las secuelas psicológicas que eso conllevó para gran parte de las personas contratadas) con intención mejorar los algoritmos de ChatGPT (Perrigo 2023).

Sin embargo, en el actual estadio político-económico capitalista neoliberal, el uso de la IA de cara a la «optimización» de distintos procesos conllevaría un potencial aumento de la tasa de productividad del trabajo, que se vería traducido en el crecimiento de la tasa de ganancia de instituciones alrededor del mundo. Corporaciones y estados ven en la IA una aliada ideal. Esto quiere decir que, a pesar de todos los daños que el uso de la IA está causando, tenemos que asumir que esta tecnología está aquí para quedarse, al menos a medio plazo.

Una forma recurrente de intentar mitigar parte del impacto dañino producido por la IA es la creación de estándares éticos que puedan, apoyándose en la legislación y en la presión popular, guiar en su diseño e implementación. En muchos casos, estos estándares ofrecen valores ético-políticos que seguir de forma deontológica, *à la* Kant. Los valores pueden partir de conceptos muy amplios (justicia, seguridad, igualdad) o más concretos, relacionados con los sistemas y artefactos en particular que se pretenden estudiar (explicabilidad, disputabilidad, transparencia). Escoger e interpretar estos valores supone necesariamente integrar una visión del mundo específica en el diseño y la producción tecnológica. Tal y como Langdon Winner exponía en 1980, los artefactos tienen política, independientemente de si somos conscientes o no de ello. Los distintos códigos deontológicos que se construyan deberán tener esto en cuenta y, en la medida de lo posible, explicitar cómo se concretan dichos valores en torno a cada sistema de IA en particular.

Este capítulo pretende evaluar la posibilidad de diseñar códigos deontológicos para el desarrollo de sistemas de IA y, a su vez, problematizar lo insuficiente de algunas perspectivas éticas para enfrentar de forma eficaz los riesgos y daños descritos anteriormente. Como punto de partida, se presenta la teoría de los sistemas sociotécnicos como herramienta para analizar el impacto de la IA. Además, se presenta una aproximación interdisciplinar a los códigos morales y protocolos de actuación, que trascienda la ética, beba de la economía política y tenga en cuenta la materialidad de la IA, cuestionando su diseño y proceso de producción al completo.

## LA INTELIGENCIA ARTIFICIAL Y LOS SISTEMAS SOCIOTÉCNICOS

A día de hoy, los debates sobre la IA giran en torno a las tecnologías algorítmicas en concreto, sin enfocarse apenas en los sistemas más amplios donde se encuadran. Como indican Olya Kudina e Ibo van de Poel (2024), esto es cierto incluso cuando se trata de estudiar los problemas éticos que plantea la IA. Es relativamente sencillo encontrar artículos y propuestas varias sobre el diseño de una «IA justa» o una «IA transparente», sin tener en cuenta que los conceptos de justicia o transparencia no dependen únicamente de la tecnología, sino también de los *sistemas sociotécnicos* en los que se integra.

La teoría de los sistemas sociotécnicos propone una perspectiva holística para entender el desarrollo y funcionamiento de las tecnologías. En este sentido, las tecnologías no operarían como entes aislados, sino a través de la compleja interacción de múltiples elementos y procesos, que son parcialmente autónomos (teniendo en cuenta las normativas y regulaciones tecnológicas, el flujo de los mercados, el significado cultural de cada tecnología y sus redes de distribución, entre otras). Esto no quiere decir que las especificaciones técnicas de cada sistema algorítmico pasen a un segundo plano. La clave es que estas no deben entenderse en abstracto, sino concretadas en un lugar y momento histórico, social y político determinado. Entre otros elementos, es especialmente importante el factor humano: las relaciones entre personas (agentes) que moldean las instituciones, las que a su vez mediarán las interacciones entre humanos y tecnología.

Aunque se trata de una teoría interdisciplinar, en el ámbito de la filosofía de la tecnología destacan las contribuciones de Kudina (2023), Franssen (2015), Van de Poel (2020), Vermaas *et al.* (2011), Ottens *et al.* (2006), y Kroes *et al.* (2006). En castellano, destaca el trabajo de Hevia Martínez (2019) en la conceptualización de los sistemas sociotécnicos desde una aproximación bungiana. El artículo «*A sociotechnical system perspective on AI*», de Kudina y Van de Poel, publicado en junio de 2024, constituye probablemente la contribución más completa a esta discusión, y es la principal fuente en la que se basa esta sección.

Hay muchas razones por las que aplicar un enfoque sociotécnico al estudio de la IA. Kudina y Van de Poel explican que, ante todo, esta perspectiva ayudaría a comprender mejor los sistemas de IA (cómo funcionan, qué efectos sociales tienen, qué problemas éticos plantean y cómo abordarlos), destacando que la IA está integrada en sistemas sociotécnicos más amplios que

determinan su funcionamiento y sus efectos. A su vez, la IA transformaría los sistemas sociotécnicos existentes, como el sistema de telecomunicaciones, de salud, o de transportes. Una de las particularidades de la IA es que tiene un potencial disruptor mayor que el de otras tecnologías, por lo que debemos prestar especial atención a la hora de analizar su impacto social. Es importante asumir también que enfrentarse a los riesgos que supone la IA no es tan sólo un reto técnico, y que requerirá necesariamente del desarrollo de nuevas competencias para la gente de a pie, nuevas formas de organizarse, nuevas instituciones y quizás nuevos esfuerzos legislativos. Kudina y van de Poel comentan que se podrían llegar a necesitar nuevos conceptos morales, jurídicos y filosóficos para afrontar este reto.

Otra razón para aplicar este enfoque es que nos ayuda a ampliar el prisma a través del cual vemos los daños que causa la IA, y a «visibilizar lo invisible». Al entender los riesgos de la IA como *riesgos sociotécnicos*, seremos capaces de extender nuestro análisis a las distintas instituciones y relaciones humanas que influyen en el diseño y la producción de la IA, y no sólo quedarnos en las que nos afectan más de cerca. Por ejemplo, si ignoramos las cadenas de producción y logística que permiten la existencia de la IA y no las entendemos como parte del sistema sociotécnico, es más probable que pasemos por alto el problema de la manufacturación de los semiconductores, o el de la geopolítica de los cable submarinos, entre muchos otros.

Además, entender cómo la IA se enmarca en un sistema sociotécnico nos puede ayudar a escapar del plano metafórico y de las falsas expectativas en torno a ella. El propio término «inteligencia artificial» es en sí mismo una metáfora que utilizamos para describir, entre otras cosas, los sistemas algorítmicos y sus aplicaciones. A su vez, sobre este término, se superponen infinidad de metáforas: la de la caja negra para explicar, precisamente, la falta de «explicabilidad» de los algoritmos que usamos; la de la «nube» para referirnos al almacenamiento masivo de datos con acceso vía internet; la de la «inteligencia artificial general» (o *AGI* por sus siglas en inglés) para las narrativas (ahora mismo, cercanas a la ciencia ficción) en torno a una posible adquisición de conciencia por parte de la IA; o incluso «el nuevo petróleo», por la importancia de la IA en el mercado financiero global. La IA es difícil de describir por muchos motivos, como su novedad o su complejidad, y las metáforas «se pegan», moldeando nuestras percepciones incluso cuando han superado su propósito. Describir con más precisión lo que realmente puede hacer la IA en lugar de adoptar metáforas aporta mayor claridad y podría permitirnos actuar con más responsabilidad.

De entre los argumentos que se proponen en favor del uso de la perspectiva de los sistemas sociotécnicos para el estudio de la IA, destaca el argumento de Kudina y Van de Poel sobre la gobernanza de la IA. La atención que reciben actualmente las investigaciones en «ética de la tecnología» son crecientes, y con razón. Sin embargo, si entendemos a la IA como parte de un sistema sociotécnico complejo, tenemos que asumir que un abordaje adecuado de los problemas éticos que plantea la IA no puede ser tan solo una cuestión de directrices éticas. Requiere necesariamente prestar atención a cuestiones políticas y de gobernanza más amplias. Para ello, hay que trabajar de forma coordinada e interdisciplinar, enfocándonos en el sistema económico-político, y apoyándonos en distintas alternativas, técnicas y políticas. En resumidas cuentas, estos autores entienden que no sólo precisamos de la ética, sino también de una *política* de la tecnología.

## RIESGOS SOCIOTÉCNICOS DE LA IA: UN BREVE REPASO

Antes de seguir ahondando en este asunto de la política de la tecnología, conviene dar un repaso general a los riesgos sociotécnicos que ocasiona la IA, para tener un mapa más completo del asunto. Cabe tener en cuenta que los mencionados a continuación son tan solo *algunos* de los principales riesgos sociotécnicos. Se trata de una lista no exhaustiva, que bien puede ser ampliada en distintas direcciones.

### Sesgos y discriminación algorítmica

Uno de los riesgos más discutidos en la literatura sobre IA es la posibilidad de que los algoritmos reproduzcan o incluso amplifiquen sesgos y desigualdades sociales preexistentes. La IA, especialmente en áreas como la contratación, la justicia penal, el crédito y la atención médica, se alimenta de grandes cantidades de datos históricos que evidentemente tienen sesgos implícitos relacionados con la clase social, el color de piel, la procedencia o el género. Estos sesgos se incorporan a los modelos algorítmicos, lo que produce decisiones automatizadas que discriminan a ciertos grupos, perpetuando desigualdades estructurales. El riesgo sociotécnico radica aquí en que las tecnologías basadas en IA operan con una aparente objetividad y *autoridad epistémica*, lo que puede dificultar la identificación de estos sesgos y, por ende, su mitigación. Esto puede llevar a situaciones en las que las decisiones tecnológicas, tomadas bajo el supuesto de ser «neutrales», afecten de forma desproporcionada a grupos vulnerables.

## Responsabilidad

Uno de los mayores riesgos socio-técnicos asociados a la inteligencia artificial es la dificultad para establecer una clara línea de responsabilidad en los casos de daño o perjuicio derivados de decisiones algorítmicas. A medida que los sistemas de IA se automatizan y toman decisiones que afectan la vida de las personas, surge la cuestión de quién es responsable cuando un algoritmo comete un error o causa un daño. Este riesgo plantea importantes desafíos en términos de justicia y rendición de cuentas. Si una IA comete un error en un contexto legal, médico o financiero, por ejemplo, ¿quién debe ser considerado responsable? ¿El desarrollador del algoritmo, el operador del sistema, o incluso el propio sistema? La falta de un marco normativo claro y global que regule la responsabilidad de las decisiones automatizadas crea un vacío legal que puede dejar desprotegidos a los individuos afectados por errores algorítmicos, además de diluir la rendición de cuentas en los sistemas sociotécnicos más amplios.

## Privacidad

El buen funcionamiento de la IA depende de grandes volúmenes de datos, que a menudo incluyen información personal, sensible o incluso privada. Esta información se extrae de internet, en muchas ocasiones, de forma ilícita o ignorando las distintas leyes de propiedad intelectual y *copyright*. Los sistemas de IA, al analizar estos datos, podrían violar sistemáticamente la privacidad de los individuos. Además, los algoritmos de IA pueden ser vulnerables a ataques cibernéticos que comprometan tanto los datos como el funcionamiento de los sistemas en los que se encuadran. Los riesgos asociados con la recopilación, almacenamiento y uso de grandes bases de datos pueden ser amplificadas por la falta de regulación y la dificultad de implementar controles de seguridad efectivos a medida que los sistemas de IA se hagan más complejos.

## Falta de transparencia y explicabilidad

La opacidad de muchos algoritmos de IA, especialmente los basados en aprendizaje profundo (*deep learning*), presenta un riesgo significativo en términos de responsabilidad y gobernanza. La toma de decisiones automatizada se está convirtiendo en una parte central de infinidad de prácticas organizacionales, gubernamentales y empresariales, pero los procesos internos de muchos algoritmos son difíciles de entender incluso para sus propios creadores: a esto es precisamente a lo que nos referimos con el término *Black Box*.

Desde una perspectiva sociotécnica, la falta de transparencia y de explicabilidad de las decisiones algorítmicas crean una desconexión entre la tecnología y las personas afectadas por ella. La falta de comprensión del funcionamiento de la IA puede estudiarse como un problema de *injusticia episódica*: esta tecnología se implementa en multitud de ámbitos de la vida de la gente, sin que la propia gente sepa cómo funciona en realidad, lo que dificulta darle sentido a muchas de sus experiencias diarias.

Trabajo (el que vemos, y el que no vemos)

Otro riesgo sociotécnico importante es el impacto de la automatización y la inteligencia artificial sobre el mercado laboral. La IA tiene el potencial de transformar industrias al reemplazar ciertas tareas humanas, lo que podría resultar en una disminución significativa de la demanda de trabajo en ciertos sectores. En particular, las profesiones creativas, como la escritura, el diseño gráfico, la música y las artes visuales, se están viendo afectadas por los generadores de imágenes, como Dall-E, o de texto, como Chat GPT. También es importante resaltar la influencia de la IA en la precarización de ciertos empleos, como los relacionados con la llamada «economía de plataformas», de entrega de comida o de taxi, por ejemplo.

Además, un aspecto fundamental del funcionamiento de los sistemas de IA es el trabajo invisible que se esconde detrás de su uso. Aunque muchas aplicaciones de IA parecen funcionar de manera autónoma, su desarrollo y mantenimiento dependen de un conjunto de tareas realizadas por seres humanos, a menudo en condiciones precarias y con escasa visibilidad. Este trabajo invisible incluye desde el etiquetado de datos, la moderación de contenido y la supervisión de los algoritmos, hasta las tareas de entrenamiento de modelos y el ajuste de las redes neuronales.

Las personas encargadas de estas tareas a menudo enfrentan largas jornadas laborales, compensación inadecuada y condiciones laborales precarias. Además, muchas de estas tareas están subcontratadas a empresas en países con bajos estándares laborales, lo que perpetúa la explotación de la mano de obra en el paradigma capitalista. El riesgo sociotécnico aquí reside en que este trabajo invisible se oculta detrás de la apariencia de autonomía de la IA, lo que dificulta el reconocimiento de las desigualdades laborales y la explotación que sustenta la operación de muchos sistemas basados en inteligencia artificial.



## Salud mental

La integración de la IA en la vida cotidiana, especialmente a través de redes sociales, aplicaciones de salud y asistentes virtuales, puede tener efectos negativos sobre la salud mental a escala social. Por ejemplo, los algoritmos que alimentan las redes sociales pueden contribuir a la creación de cámaras de eco y la propagación de desinformación, afectando la percepción de la realidad de los usuarios y exacerbando problemas de ansiedad, depresión o estrés. Existen ya múltiples investigaciones que conectan el uso de Instagram con el desarrollo de trastornos de la conducta alimentaria en adolescentes (O'Brien 2015; Teo y Collinson 2019), entre otros problemas.

## Medio ambiente

El desarrollo y la implementación de tecnologías de IA a gran escala conllevan riesgos medioambientales significativos, especialmente relacionados con el consumo masivo de energía y la utilización de materiales raros en la fabricación de los dispositivos necesarios para operar estos sistemas. Podemos encontrar ejemplos de investigaciones en esta dirección en el trabajo de Crawford (2021) o George *et al.* (2023).

Los modelos de IA más avanzados, como los basados en redes neuronales profundas, requieren enormes cantidades de energía para ser entrenados y mantenidos, y el *hardware* requiere de la extracción de metales raros, como el litio. A medida que el ritmo de producción y consumo de IA aumenta, el costo energético asociado con su funcionamiento lo hace también, lo que contribuye significativamente al agotamiento de recursos naturales y al aumento de las emisiones de gases de efecto invernadero. Además, el reciclaje y la gestión de residuos tecnológicos asociados con la IA representan otro desafío medioambiental, dado el rápido ritmo de obsolescencia de los dispositivos y el creciente volumen de desechos electrónicos generados.

No son pocos los riesgos sociotécnicos que se han traducido ya en daños reales alrededor del globo. La pregunta que se nos presenta a continuación, en relación con este volumen y con el Seminario Permanente, sería: ¿hasta qué punto pueden los códigos deontológicos ayudarnos en la prevención de dichos riesgos?

## **CÓDIGOS DEONTOLÓGICOS EN IA Y LAS FRONTERAS DE LA ÉTICA: HACIA UNA POLÍTICA DE LA TECNOLOGÍA**

Durante el pasado ciclo del Seminario Permanente de Niaiá (2023-24) se debatió la importancia y adecuación de los códigos deontológicos en distintos

ámbitos, como el periodístico, el médico, o el científico, entre otros. Cuando hablamos de códigos deontológicos, nos referimos a los conjuntos de principios, normas y directrices éticas que guían el comportamiento de los miembros de una determinada disciplina o profesión, aunque la deontología puede extenderse a otros espacios más allá del profesional. Al hablar de deontología nos acordamos inevitablemente de Immanuel Kant y su imperativo categórico, con su formulación más clara en: «obra sólo según aquella máxima por la cual puedas querer que al mismo tiempo se convierta en ley universal». Esta frase sirve de síntesis de la deontología, y de evidente inspiración para distintos códigos éticos, como lo es el juramento hipocrático, por ejemplo.

En lo que respecta a la IA (y a la tecnología en general), el uso de protocolos deontológicos no está muy extendido. Se suele trabajar con distintos paradigmas de diseño socio-técnico que también aporten una visión ética. Uno de los ejemplos más prominentes es el del «*Design for Values*» (DfV) o «Diseño en Valores», que encuentra en el trabajo de Ibo van de Poel su mayor exponente, y que promueve la incorporación activa de valores éticos, sociales y humanos en el diseño tecnológico. Este enfoque reconoce que la tecnología y el diseño tienen un impacto directo en la vida de las personas y, por ende, deberían reflejar ciertos valores clave para la sociedad en la que se enmarcan. Esta teoría escapa de la neutralidad del diseño y asume, por tanto, que cualquier sistema siempre integrará ciertos valores. El objetivo será que estos valores se escojan explícitamente y guíen las decisiones en cada etapa del diseño. Dicho diseño deberá promover la inclusión de múltiples partes interesadas (usuarios, expertos, comunidades) en el proceso. Cuando surjan conflictos entre los valores a implementar, se propone un proceso de análisis y deliberación para entender bien la raíz de estos conflictos, y poder priorizar ciertos valores o aportar otras soluciones, creativas y dialogadas (van de Poel 2013). También es interesante explorar a fondo cada valor y aportar guías de diseño particulares, sobre todo en valores con significados más amplios, como el de la justicia (van de Poel *et al.*, 2024).

Aun así, también podría ser interesante desarrollar códigos deontológicos que puedan servir de brújula ética en el diseño y producción de la IA. Podrían incluso servir para priorizar ciertos valores utilizando el mismo enfoque de *Diseño en Valores*. Podemos encontrar una aproximación inicial a los códigos deontológicos en la IA en un artículo publicado en el blog de Niaiá, titulado «*Estándares Éticos en la Inteligencia Artificial Generativa*», escrito por Borja García Haendler (2024). Dicho texto pretende dar una serie de pautas en torno a las que construir estos códigos éticos, más concretamente, en

torno a los generadores de imagen y sonido, como puede ser *Dall-E*. García Haendler presenta once categorías de reflexión en torno a estas tecnologías. Estas son: seguridad y fiabilidad; transparencia y explicabilidad, privacidad y protección de datos; equidad y no discriminación; el problema del deep fake; derechos de propiedad intelectual; responsabilidad; IA generativa con valores centrados en el ser humano; sostenibilidad ambiental; cooperación global; y monitoreo y evaluación continuos.

Aunque algunas de estas pautas están muy centradas en la IA generativa (como sería llevar a cabo un estudio cauteloso del problema del *deepfake*), otras se podrían replicar en el desarrollo de más sistemas de IA. Reflexionar sobre transparencia, responsabilidad, o el impacto climático de la IA es urgente y se podría integrar en cualquier código deontológico. Sin embargo, surgen dudas sobre la operacionalización de estos códigos. ¿Qué quiere decir realmente «respetar la privacidad» o los «derechos de propiedad intelectual», si la mayoría de sistemas algorítmicos se basan en la extracción masiva de datos? ¿Cómo se concreta la ambigüedad de la tan manida expresión «valores centrados en el ser humano»? ¿Cómo cooperamos globalmente, si siempre hay una parte del globo más afectada por el impacto de la IA (que suele ser también la que menos poder de decisión tiene)? ¿Estarían los estados, los mercados, y la sociedad civil dispuestos a cumplir remotamente alguno de estos códigos, si esto supone repensar radicalmente lo que es la IA?

Estas cuestiones no son exclusivas a los protocolos deontológicos, sino que se pueden extender a la reflexión ética en general. Por un lado, la ética de la tecnología suele atomizar el foco de la deliberación moral, centrándose en la persona concreta, la empresa concreta, o el gobierno concreto y evaluando lo que estos pueden hacer en relación a una tecnología en singular. En general, faltan estructuras de análisis más complejas que nos permitan entender cómo operan los sistemas y lo que es necesario para intervenir en ellos. Pensar en la «transparencia» como valor abstracto suele ser poco productivo. Es más interesante reflexionar sobre por qué los algoritmos no son transparentes y a qué actores del sistema sociotécnico beneficia la falta de transparencia. Teniendo en cuenta esto, la clave es pensar en cómo *garantizar* la transparencia, y en muchos casos, disponer de una guía ética para el diseño no parece suficiente.

De hecho, uno de los problemas más habituales en la aplicación de la ética al diseño tecnológico es el del «*ethics washing*», que podríamos traducir como *lavado de cara ético*. Esta es la práctica que realizan infinidad de organizaciones adoptando y promoviendo ciertos estándares éticos para mejorar su imagen pública, pero sin realizar cambios sustanciales en sus

políticas, productos o procesos. Es una forma de lavado de imagen similar al «*greenwashing*», que ocurre cuando una empresa exagera su compromiso con la sostenibilidad. Lo mismo sucede con el «*participation-washing*», donde las instituciones invitan a distintas comunidades a participar en el proceso del diseño tecnológico de forma no vinculante. Al final, las empresas o estados que desarrollan la tecnología seguirán manteniendo el poder sobre las decisiones entorno a esta. De nuevo, la clave no está en qué valores (como la sostenibilidad o la participación) se aceptan, sino en cómo podemos asegurarnos de que estos realmente se lleven a cabo.

Es aquí donde la ética encuentra sus fronteras, y donde debemos empezar a buscar una *política de la tecnología*, asumiendo un enfoque interdisciplinar que combine la ética con la filosofía política, la economía política, la geopolítica, y otras áreas. Sólo de esta forma podremos tratar la cuestión sobre las estructuras de poder que promueven ciertos formatos de diseño, producción, logística y comercialización tecnológica. Sin esto, cualquier propuesta respecto a la cooperación global en el monitoreo ético de la tecnología será insuficiente. Los códigos éticos, sin una voluntad política clara, corren el riesgo de ser cooptados por el propio sistema que causa los daños que buscamos prevenir.

En este sentido, es necesario seguir investigando en dos direcciones. En primer lugar, como estrategia preventiva dentro del marco legal, construyendo regulaciones que garanticen que las decisiones sociales respecto a la IA se lleven a cabo. A medio plazo, sin embargo, necesitamos trascender la mera regulación y buscar *modelos alternativos de propiedad* de la IA, para evitar, como es el caso actual, que su control esté en manos de unas pocas personas (Maas y Moreno Inglés 2024). Tenemos que transformar radicalmente las estructuras de control de la tecnología, lo que pasa inevitablemente por reformular las relaciones de producción en torno a las mismas. En definitiva, no es cuestión de renegar de la ética, sino de ver dónde están sus límites, y complementarla con una *política* que esté a la altura.

## Fuentes

- Akter, S.; McCarthy, G.; Sajib, S.; Michael, K.; Dwivedi, Y. K.; D'Ambra, J.; and Shen, K. N. 2021. Algorithmic bias in data-driven innovation in the age of AI. *International Journal of Information Management*, 60.
- Crawford, K. (2021). *The Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence*. Yale University Press.
- Feldstein, S. 2019. The global expansion of AI surveillance. Technical Report. Washington DC: Carnegie Endowment for International Peace.

- Franssen, M. (2015). Design for Values and Operator Roles in Sociotechnical Systems. In J. van den Hoven, P. E. Vermaas, & I. van de Poel (Eds.), *Handbook of Ethics, Values, and Technological Design: Sources, Theory, Values and Application Domains* (pp. 117–149). Springer Netherlands.
- Gebru, T. 2020. Race and gender. In *The Oxford handbook of ethics of AI*, edited by M. D. Dubber; F. Pasquale; S. Das, 251-69. Oxford University Press.
- George, A. S., George, A. H., & Martin, A. G. (2023). The environmental impact of AI: a case study of water consumption by chat GPT. *Partners Universal International Innovation Journal*, 1(2), 97-104.
- Hevia Martínez, G. (2019). La sociedad como artefacto: Sistemas sociotécnicos, sociotecnologías y sociotécnicas. *CTS: Revista iberoamericana de ciencia, tecnología y sociedad*, 14(40), 267-295.
- Kroes, P., Franssen, M., Van de Poel, I., & Ottens, M. (2006). Treating socio-technical systems as engineering systems: Some conceptual problems. *Systems Research and Behavioral Science*, 23(6), 803–814.
- Kudina, O. (2023). *Moral Hermeneutics and Technology: Making Moral Sense Through Human-Technology-World Relations*. Lexington Books.
- Kudina, O., & van de Poel, I. (2024). A sociotechnical system perspective on AI. *Minds and Machines*, 34(3), 21.
- Maas, J., y Moreno Inglés, A. (2024). Beyond Participatory AI. In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society* (Vol. 7, pp. 932-942).
- Noble, S. U. 2018. Algorithms of oppression: How search engines reinforce racism. In *Algorithms of oppression*. New York: New York University Press.
- O'Brien, K. (2015). The Cultivation of Eating Disorders Through Instagram. Master's thesis, University of South Florida, Tampa FL.
- Ottens, M., Franssen, M., Kroes, P., & Van de Poel, I. (2006). Modeling engineering systems as socio-technical systems. *International Journal of Critical Infrastructures*, 2, 133–145.
- Perrigo, B. (2023). Exclusive: OpenAI used Kenyan workers on less than \$2 per hour to make ChatGPT less toxic. *Time Magazine*, 18, 2023.
- Teo, N. S. Y., y Collinson, S. L. (2019). Instagram and risk of rumination and eating disorders: An Asian perspective. *Psychology of Popular Media Culture*, 8(4), 491.
- van de Poel, I. (2013). Translating values into design requirements. *Philosophy and engineering: Reflections on practice, principles and process*, 253-266.
- van de Poel, I., Secomandi, F., Abraham, E., van Uffelen, N., Feenstra, M., Liem, C., Rocco, R., y Moreno Inglés, A. (2024). White Paper: Design for Justice. Delft Design for Values Institute.
- Vermaas, P., Kroes, P., van de Poel, I., Franssen, M., & Houkes, W. (2011). *A Philosophy of Technology: From Technical Artefacts to Sociotechnical Systems* (Vol. 6).
- Winner, L. (2017). Do artifacts have politics?. In *Computer ethics* (pp. 177-192). Routledge.