



# PRINCIPIOS NORMATIVOS PARA UNA ÉTICA DE LA INTELIGENCIA ARTIFICIAL

**FABIO MORANDÍN-AHUERMA**

# PRINCIPIOS NORMATIVOS PARA UNA ÉTICA DE LA INTELIGENCIA ARTIFICIAL

Fabio Morandín-Ahuerma

ISBN: 978-607-8901-78-4  
Primera edición, México, 2023

# RECOMENDACIÓN DEL CONSEJO SOBRE INTELIGENCIA ARTIFICIAL DE LA OCDE: DESIGUALDAD E INCLUSIÓN

## Introducción

La “Recomendación del Consejo de la OCDE sobre inteligencia artificial” es un conjunto de directrices y principios para el desarrollo y despliegue de sistemas de IA alineados con los valores y derechos humanos, la transparencia y la rendición de cuentas. Fue adoptada por el Consejo de la Organización para la Cooperación y el Desarrollo Económicos (OCDE) en mayo de 2019 e incluye cinco principios clave para la gobernanza de la IA. Ésta debe beneficiar a las personas y al planeta impulsando el crecimiento inclusivo, el desarrollo sostenible y el bienestar; los sistemas de IA deben respetar el estado de derecho, los derechos humanos, los valores democráticos y la diversidad; también deben ser transparentes para garantizar que las personas comprendan los resultados y puedan cuestionarlos; los sistemas deben ser sólidos, seguros y protegidos durante todo su ciclo de vida; y las partes interesadas deben colaborar para maximizar los beneficios de la IA y minimizar sus riesgos. Se concluye aquí que los Principios de IA de la OCDE han sido criticados por no ser vinculantes y carecer de un mecanismo formal de aplicación. Tampoco abordan preocupaciones sociales amplias, como el desplazamiento de puestos de trabajo y la concentración de poder en unas pocas empresas. Además, es posible que los principios no tengan en cuenta los retos a los que se enfrentan los países en desarrollo y estén demasiado centrados en los intereses de las economías avanzadas. Sin embargo, la Recomendación de la OCDE establece una serie de principios ampliamente aceptados y respaldados tanto por los miembros, como por los que no lo son. Entre los valores figuran la transparencia, la solidez, la seguridad, la responsabilidad, la justicia, la equidad y la determinación humana. Este conjunto compartido de principios proporciona una base sólida para el desarrollo responsable de la IA, más allá de las fronteras de los “países ricos”.

## Recomendación del consejo de la OCDE sobre inteligencia artificial

El Consejo de la Organización para la Cooperación y el Desarrollo Económicos (OCDE), a nivel ministerial, aceptó la Directriz sobre inteligencia artificial en mayo de 2019 a propuesta del Comité de Política de Economía Digital (CDEP por sus siglas en inglés) [1]. La “Recomendación del consejo de la OCDE sobre inteligencia artificial” (Recommendation of the Council on OECD Legal Instruments Artificial Intelligence - OECD-LEGAL-0449) [2] es un documento que promueve la administración responsable de la IA, al tiempo que trata de garantizar el respeto de los derechos humanos y los ideales democráticos, con el fin de promover la innovación y confianza en la IA.

La Recomendación de OCDE, se centran en cuestiones específicas de la IA y establece una norma aplicable y flexible, para que no quede pronto obsoleta, complementando las normas existentes de la organización en áreas como la privacidad, la gestión de riesgos de seguridad digital y la conducta empresarial responsable.

La OCDE, que tiene su sede en París y está formada por 38 Estados con el objetivo de coordinar sus políticas económicas y sociales, también se conoce como el “club de los países ricos”, debido a que entre sus agremiados se concentra más del 60% del PIB nominal mundial [3]. Sin embargo, hay que señalar que países miembros como Turquía, Colombia, Chile y México aún enfrentan retos económicos importantes como para ser considerados “países ricos” [4].

La “Recomendación del consejo sobre inteligencia artificial” es un documento no vinculante que fue aprobado y firmado el 21 de mayo de 2019 por los 38 países miembros, a los que se adhirieron ese mismo día, Argentina, Brasil, Perú y Rumania (no miembros); posteriormente, lo hicieron otros no miembros, como Ucrania (29 de octubre de 2019), Malta (19 de diciembre de 2019), Egipto y Singapur (30 de marzo de 2021). Estas directrices están en concordancia con otras de la propia organización en ámbitos como la protección de datos, la gestión de riesgos para la seguridad digital y las prácticas empresariales éticas [5].

Así, el Comité de expertos en inteligencia artificial de la OCDE publicó en mayo de 2019 este conjunto de principios éticos que fue adoptado por los países miembros y adherentes.

## **1. Crecimiento inclusivo, desarrollo sostenible y bienestar**

---

El principio destaca la capacidad de la IA fiable para promover los objetivos de desarrollo mundial, el progreso y la prosperidad generales para todos los involucrados alrededor del orbe.

La OCDE ha hecho un llamamiento a las naciones más avanzadas tecnológicamente, instándolas a asumir un deber internacional de solidaridad con las naciones menos avanzadas tecnológicamente. Esto sirve para garantizar que los beneficios de la IA se distribuyan de manera que se fomente el acceso y la participación, contribuyendo a un orden mundial más equitativo con respecto a la información, la educación, la cultura, la investigación y la estabilidad de las naciones [2].

Para lograr este objetivo, es esencial establecer un marco de cooperación y colaboración internacional. Dicho marco implicaría el intercambio de conocimientos y experiencia, la provisión de recursos financieros y tecnológicos y el compromiso activo de todas las partes interesadas. Si se construye una comunidad mundial integradora y cooperativa, se puede crear un entorno en el que la mayoría disfrute de los beneficios de la IA y en el que el progreso tecnológico se aproveche para consolidar el desarrollo social [6].

## **2. Valores y equidad centrados en el ser humano**

---

Para garantizar una sociedad justa y equitativa, la OCDE considera que los sistemas de IA deben crearse de manera que respeten el estado de derecho, los principios democráticos y la diversidad.

A lo largo de la vida útil de los sistemas de inteligencia artificial, los agentes deben respetar la ley, los derechos humanos y los principios democráticos. Por ejemplo, la diversidad, la equidad, la justicia social, la no discriminación, la autonomía, privacidad, protección de datos y los derechos laborales reconocidos. Para lo anterior, los actores de IA deben implementar garantías y métodos que sean apropiados y en línea con las mejores prácticas, como la prioridad del juicio humano y la autodeterminación [2].

### **3. Transparencia y explicabilidad**

---

Este principio sirve para que los usuarios sepan cuándo están interactuando con sistemas de IA y puedan cuestionar los resultados; este principio enfatiza la divulgación responsable para que cualquiera entienda los mecanismos de decisión automática [2].

La OCDE destaca la importancia de que los agentes de la IA se comprometan a proporcionar datos pertinentes que se ajusten al estado actual de la tecnología y sean adecuados para la situación específica. Por ejemplo, esto incluye aumentar la responsabilidad entre las partes interesadas sobre sus interacciones con los sistemas de IA, tanto en el lugar de trabajo como fuera de él. También implica fomentar una comprensión general de los sistemas de IA para que los afectados puedan comprender los resultados, sobre todo si son desfavorables. Esta comprensión permite a las personas cuestionar las decisiones basándose en información transparente y comprensible sobre las variables implicadas y el razonamiento que subyace a las predicciones, recomendaciones y decisiones [7].

### **4. Robustez, seguridad y protección**

---

Este principio se refiere a que, a lo largo de su vida útil, los sistemas de IA deben funcionar de manera fiable, segura y protegida de los posibles riesgos, que siempre deben ser evaluados y abordados oportunamente [2].

La OCDE considera que los sistemas de IA deben ser fiables, seguros y protegidos durante todo su ciclo de vida para que puedan funcionar correctamente y no supongan un peligro excesivo para la seguridad en situaciones de uso normal y previsible [2].

Para permitir el análisis de los resultados de los sistemas de IA y las respuestas a las consultas es crucial que las partes interesadas garanticen la trazabilidad a lo largo del ciclo de vida del sistema. Esto implica hacer un seguimiento de los conjuntos de datos, los procesos y las decisiones, y garantizar que las respuestas sean adecuadas al contexto específico y coherentes con la tecnología y las prácticas más recientes.

La trazabilidad es un concepto clave en este contexto, ya que permite el seguimiento de todos los procesos desde los datos de entrada iniciales hasta los resultados de salida. Al garantizar el seguimiento y registro de toda la información

pertinente, las partes interesadas pueden identificar y abordar cualquier problema o error que pueda surgir [8].

Además, las partes deben tener en cuenta la importancia de cada contexto, las circunstancias particulares en las que opera el sistema de IA a la hora de determinar la respuesta adecuada, e incluir factores como la diversidad cultural y lingüística, así como los marcos jurídicos y normativos.

Garantizar la trazabilidad es esencial para promover la responsabilidad y la transparencia en los sistemas de IA. Mediante el seguimiento y el análisis de todos los conjuntos de datos, procesos y decisiones pertinentes, los involucrados pueden contribuir a garantizar que los sistemas de IA funcionen de forma ética y responsable, en consonancia con las mejores prácticas y tecnologías [9]. Para gestionar los riesgos vinculados a los sistemas de IA, como la privacidad, la seguridad digital, y la parcialidad, la OCDE afirma que se debe aplicar continuamente una estrategia sistemática de gestión de riesgos en cada fase del ciclo de vida del sistema, en función del contexto y la capacidad de actuación [2].

## **5. Responsabilidad**

---

El correcto funcionamiento de la IA, de acuerdo con las directrices de la OCDE, debe ser responsabilidad de las organizaciones y personas que los gestionan, desarrollan y despliegan, en función de sus compromisos, del contexto, y de acuerdo con la situación imperante. Los actores de la IA deben ser responsables del correcto funcionamiento de los sistemas y no soslayar el hecho de que, detrás de todo sistema, finalmente son personas quienes los administran [2].

En octubre de 2021, una nueva Convención denominada “Poniendo en práctica los principios de IA de la OCDE: avances y perspectivas futuras” [10] revisó los principios formulados dos años antes para validar su actualidad, y consideraron cinco recomendaciones más para los gobiernos en materia de IA:

- Invertir en investigación y desarrollo.
- Fomentar un ecosistema digital propicio.
- Dar forma a un entorno político adecuado sobre IA.
- Desarrollo de capacidades para un nuevo mercado laboral y,
- Cooperación internacional para una IA confiable.

Las prioridades, por tanto, además de dar financiamiento a las instituciones y proyectos de investigación y desarrollo sobre IA, abordan los problemas sociales y fomentan la adopción de la IA por parte de las empresas [11].

Asimismo, buscan impulsar el diálogo social inclusivo, así como dotar a la población de las competencias necesarias para utilizar la IA y promover una transición justa de los trabajadores al nuevo mercado laboral, con una alta tendencia a la automatización de los procesos [12].

Estas fueron algunas de las prioridades estratégicas y políticas establecidas por la OCDE en 2021. La Recomendación también impulsó a la OCDE a crear un “Observatorio de políticas de IA” [16] para ayudar a poner en práctica las directrices y vigilarlas.

## Conclusiones parciales

Para la OCDE la IA debe promover el crecimiento equitativo, el desarrollo sostenible y el bienestar general en beneficio de las personas y el medio ambiente; para garantizar una sociedad justa y equitativa, los sistemas de IA deben construirse de forma que respeten el estado de derecho, los derechos humanos, los principios democráticos, la diversidad e incorporen las protecciones adecuadas para permitir la intervención humana cuando sea necesario [13].

También garantizar que las personas sean conscientes de cuándo están interactuando con los sistemas de IA y tengan la capacidad de impugnar los resultados de esas interacciones. Así, los sistemas de IA deben guiarse por la transparencia y la ejecución responsable.

El principal objetivo de la “Recomendación del consejo de la OCDE” es fomentar la innovación y la confianza en la IA, al tiempo que se defienden los derechos humanos y los principios democráticos. La Recomendación trata de alcanzar un consenso sobre la importancia y los conceptos rectores de una tecnología fiable y apoya una IA propositiva pero segura [14].

Al igual que otras propuestas, las críticas de este instrumento han sido que sus principios no son vinculantes, lo que significa que no existe un mecanismo formal para hacerlos cumplir. Los investigadores sostienen que esta falta de aplicación

hace que los principios no tengan fuerza legal y quede a consideración de los desarrolladores en las empresas y los gobiernos acatarlas, aun siendo signatarios de la OCDE [7].

Las críticas se han centrado en que, dado que los principios son amplios, están abiertos a la interpretación, lo que puede dificultar su aplicación en la práctica. Los principios se concentran en la gobernanza de la IA, pero no abordan preocupaciones más amplias relacionadas con el impacto que la inteligencia artificial pueda tener en la sociedad; por ejemplo, se argumenta que no abordan específicamente cuestiones como el desplazamiento de puestos de trabajo, el sesgo algorítmico o la concentración de poder en manos de unas pocas empresas y la falta de inclusividad [15]. Esto es que no tienen en cuenta los retos y contextos específicos de los países en desarrollo y pueden estar demasiado cerca de los intereses de las economías avanzadas [15].

En conclusión, la “Recomendación del consejo de la OCDE sobre inteligencia artificial” [2] subraya la urgente necesidad de abordar los dilemas del poder y la inclusión para garantizar que los beneficios de la IA sean compartidos por todos, al tiempo que se minimizan los riesgos potenciales y los impactos negativos. Corresponde a los responsables políticos, a las partes interesadas y a la sociedad civil trabajar juntos hacia un enfoque de la IA centrado en el ser humano que promueva un desarrollo y un despliegue responsables, transparentes e inclusivos de esta poderosa tecnología.

Si bien la OCDE puede representar los intereses de sus miembros, lo cierto es que este breve documento, sin ser legalmente normativo, también es una propuesta seria, directa y comprometida por sentar las bases para una ley de aspiraciones universalistas para el bien común.

## Referencias

---

- [1] OCDE, “Cuarenta y dos países adoptan los Principios de la OCDE sobre Inteligencia Artificial,” OECD.org, 2019. Acceso mar. 2023. [En línea]. Disponible: <https://bsu.buap.mx/b2r>
- [2] OCDE, “Recommendation of the Council on Artificial Intelligence,” OECD.org 2019. OECD.org, 2019. Acceso mar. 2023. [En línea]. Disponible <https://legalinstruments.oecd.org/en/instruments/oecd-legal-0449>
- [3] Bank, W., “GDP growth OECD members,” 2021. Acceso mar. 2023. [En línea]. Disponible: <https://data.worldbank.org/indicator/NY.GDP.MKTP.KD.ZG?locations=OE>

- [4] J. Clifton y D. Díaz-Fuentes, "From 'Club of the Rich' to 'Globalisation à la carte'? Evaluating Reform at the OECD," *Global Policy*, vol. 2, no. 3, pp. 300-311, 2011, doi: 10.1111/j.1758-5899.2011.00103.x.
- [5] OCDE, "Papers," OECD.org. Acceso feb. 2023. [En línea]. Disponible: <https://www.oecd-ilibrary.org/papers>
- [6] A. Korinek, "Integrating Ethical Values and Economic Value to Steer Progress in Artificial Intelligence," in *The Oxford Handbook of Ethics of AI*, M. Dubber, F. Pasquale, y S. Das, Eds. Oxford, UK: Oxford University Press, 2020, pp. 475-491.
- [7] AlgorithmWatch, "AlgorithmWatch is a non-profit research and advocacy organization that is committed to watch, unpack and analyze automated decision-making (ADM) systems and their impact on society," 2022. Acceso feb. 2023. [En línea]. Disponible: <https://algorithmwatch.org/>
- [8] S. Russell y P. Norvig, "Philosophy, ethics, and safety of AI," en *Artificial Intelligence: A Modern Approach*, Londres: Pearson, 2022, pp. 1032-1062.
- [9] R. Blackman, "Ethical Machines: Your Concise Guide to Totally Unbiased, Transparent, and Respectful AI." Harvard, MA, USA: Harvard Business Review Press, 2022.
- [10] OCDE, "Putting the OECD AI Principles into practice: progress and future perspectives," OECD.ai, 2021. Acceso feb. 2023. [En línea]. Disponible: <https://oecd.ai/en/mcm>
- [11] K. Yeung, "Recommendation of the Council on Artificial Intelligence (OECD)," *Intern. Legal Materials*, vol. 59, no. 1, pp. 27-34, 2020.
- [12] OCDE, "Employment database - Unemployment indicators," OECD.org. Acceso feb. 2023. [En línea]. Disponible: <https://t.ly/CEO76>
- [13] C. Burr, M. Taddeo, y L. Floridi, "The Ethics of Digital Well-Being: A Thematic Review," *Sci. Eng. Ethics*, vol. 26, no. 4, pp. 2313-2343, 2020, doi: 10.1007/s11948-020-00175-8
- [14] S. Milano, M. Taddeo, y L. Floridi, "Recommender systems and their ethical challenges," *AI & Soc.*, vol. 35, no. 4, pp. 957-967, 2020, doi: 10.1007/s00146-020-00950-y
- [15] Nature, "International AI ethics panel must be independent," Nature.com. ago. 2019. Acceso feb. 2023. [En línea]. Disponible: <https://www.nature.com/articles/d41586-019-02491-x>
- [16] OECD, "OECD.AI Policy Observatory". Acceso feb. 2023. [En línea]. Disponible: <https://oecd.ai/en/>