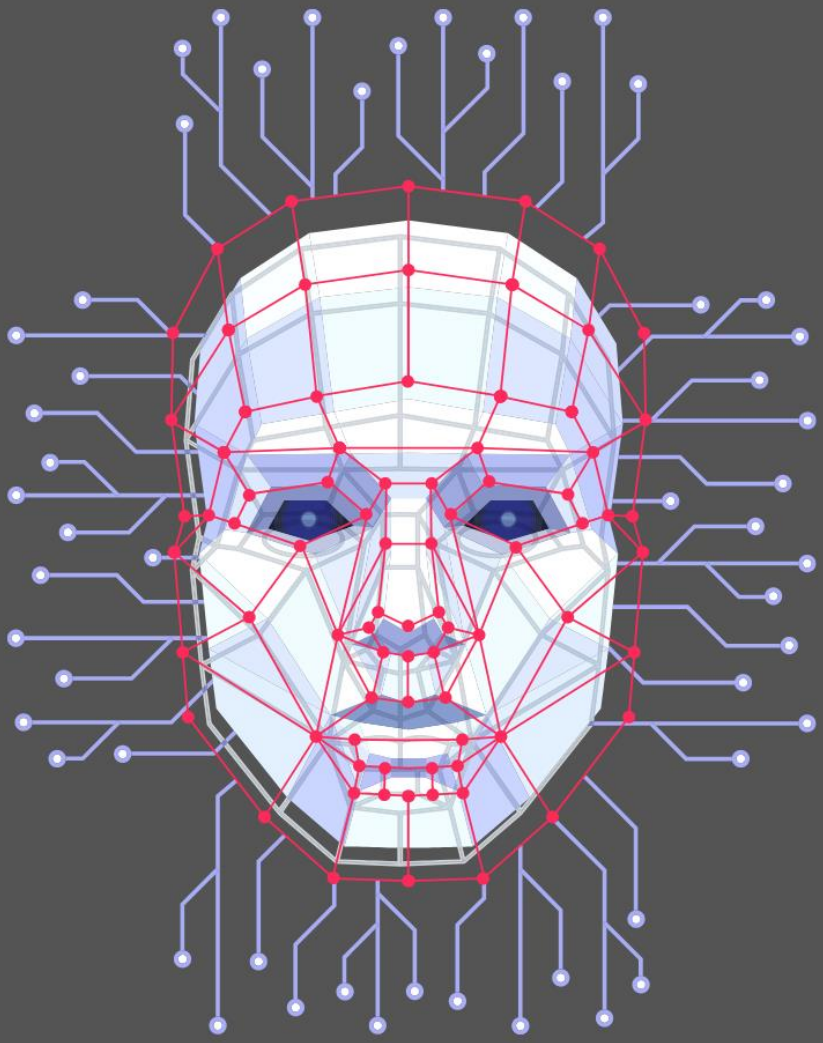
CONSEJO DE CIENCIA Y TECNOLOGÍA DEL ESTADO DE PUEBLA



PRINCIPIOS NORMATIVOS PARA UNA ÉTICA DE LA INTELIGENCIA ARTIFICIAL

FABIO MORANDÍN-AHUERMA

PUEBLA Un gobierno presente
Secretaría de Educación
CONCYTEP Consejo de Ciencia y Tecnología del Estado de Puebla
Benemérita Universidad Autónoma de Puebla 1578
Complejo Regional Nororiental

PRINCIPIOS NORMATIVOS PARA UNA ÉTICA DE LA INTELIGENCIA ARTIFICIAL

Fabio Morandín-Ahuerma

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

DIEZ RECOMENDACIONES DE LA UNESCO SOBRE ÉTICA DE LA INTELIGENCIA ARTIFICIAL

Introducción

La “Recomendación sobre la ética de la inteligencia artificial” es un documento elaborado por el Grupo especial de expertos (GEE) y adoptado por la UNESCO en noviembre de 2021. El documento presta especial atención a las implicaciones éticas de los sistemas de IA en relación con la cultura, la educación, la ciencia, la información y la comunicación. El documento pretende orientar a los responsables políticos y a las partes interesadas sobre cómo garantizar que la IA se desarrolle y utilice de forma ética. Se subraya aquí su visión de futuro y el desafortunado hecho de que no todos los países son parte del acuerdo internacional y, por tanto, no se sienten obligados a seguir estas directrices, por no ser incluidos o autoexcluirse. Se concluye que sienta las bases para futuros instrumentos normativos que puedan contribuir a su aplicación y que se den los pasos necesarios para garantizar que la ética se lleve a la práctica. Se espera que el instrumento ayude a las naciones y a las empresas a mejorar sus marcos reglamentarios éticos basándose en una visión universalista.

Recomendación sobre la Ética de la Inteligencia Artificial

La “Recomendación de la Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura sobre la Ética de la Inteligencia Artificial” (Recommendation on the Ethics of Artificial Intelligence - SHS/BIO/REC-AIETHICS/2021) fue adoptada por la Conferencia General de la UNESCO, que se reunió en la ciudad de París, a partir del día 9 hasta el 24 de noviembre de 2021, en su 41° período de sesiones, misma que esboza diez principios para la IA [1].

La UNESCO tiene 193 países miembros. Estos principios incluyen la transparencia y la explicabilidad, la no discriminación y la equidad, el respeto de la autonomía humana, la prevención del daño, la responsabilidad, la privacidad y la gobernanza de datos, el beneficio social, la sostenibilidad, la rendición de cuentas y la inclusión. Las recomendaciones son un marco normativo mundial que orienta a los países en la creación de sus propios marcos jurídicos para garantizar que la IA se despliegue de forma ética.

En su Reunión número 41° la UNESCO generó los siguientes valores y principios para el desarrollo de una IA responsable:

1. Proporcionalidad e inocuidad

Según la UNESCO, el principio de proporcionalidad subraya la importancia de garantizar que la IA se desarrolle y utilice de forma que se ajuste a su finalidad prevista, evitando al mismo tiempo cualquier exceso o peligro innecesarios. En esencia, esto implica que la aplicación de la IA debe mantenerse dentro de límites para lograr el objetivo predeterminado, sin aventurarse en una utilización excesiva o innecesaria que supere el objetivo establecido. Por ejemplo, si se utiliza la IA para detectar y prevenir delitos, se debe asegurar que las medidas tomadas sean proporcionales al riesgo de cometerlo y no utilizar demasiada fuerza o recursos en un objetivo más allá de lo establecido. Por ello, se debe aplicar la evaluación de riesgos y resultados colaterales [1].

La inocuidad, por su parte, se refiere a la necesidad de evitar el uso de la IA para fines que puedan causar daño o perjuicio a las personas o al medio ambiente. La IA no debe ser utilizada para discriminar a ciertos grupos, para la vigilancia masiva o para la manipulación psicológica. Para garantizar una mejor comprensión, el método elegido de IA debe ser adecuado y equilibrado para lograr un objetivo legítimo específico. No debe vulnerar los principios fundamentales de los derechos humanos y debe adaptarse bien a las circunstancias concretas que se presenten, basándose en principios científicos fiables [2].

2. Seguridad y protección

Para garantizar la seguridad de los seres humanos, el medio ambiente y los ecosistemas, es crucial abordar y mitigar los daños no deseados, riesgos de seguridad y las vulnerabilidades que pueden dar lugar a ataques a lo largo de todo el ciclo de vida de los sistemas de IA. Esto implica tomar medidas proactivas para prevenir y eliminar tales riesgos. Para promover eficazmente la seguridad de la IA, es esencial

establecer marcos sostenibles para acceder a los datos. Estos marcos deben dar prioridad a la privacidad al tiempo que permiten mejorar el entrenamiento y la validación de los modelos de IA utilizando datos de alta calidad. De este modo, se puede sentar una base sólida para sistemas de IA que den prioridad al bienestar de las personas y a la preservación del entorno natural.

3. Equidad y no discriminación

Las entidades de IA están obligadas a salvaguardar los principios de equidad, justicia social y no discriminación. Este mandato plantea una perspectiva integradora para garantizar la disponibilidad y accesibilidad equitativas de los beneficios de la tecnología de la IA para todos, teniendo en cuenta las distintas necesidades de grupos demográficos dispares, como los delimitados por la edad, la cultura, el idioma, la capacidad física o cognitiva, el género y la situación socioeconómica [1].

Corresponde a los Estados miembros de UNESCO cultivar un entorno que fomente la entrada, sin trabas, a los sistemas de IA que proporcionan contenidos y servicios pertinentes a nivel local, manteniendo, al mismo tiempo, el respeto por el multilingüismo y la pluralidad cultural, que se extiende tanto a los actores locales como foráneos. Este esfuerzo tiene como objetivo reducir el abismo digital y fomentar el acceso equitativo y la participación en la evolución de la IA. A nivel nacional, los Estados miembros deben luchar por la equidad en el acceso y la participación a lo largo del ciclo de vida de los sistemas de IA, entre las zonas rurales y urbanas, y las personas de todos los géneros, edades, religiones, razas, afiliaciones políticas, etnias, posiciones socioeconómicas, discapacidades o cualquier otro motivo [3].

La UNESCO ha hecho un llamamiento a las naciones tecnológicamente avanzadas para que mantengan una obligación global de solidaridad con las naciones menos desarrolladas con el fin de garantizar una distribución equitativa de los beneficios de la IA, de modo que se facilite la participación y el acceso a los sistemas de información. Este esfuerzo sirve para promover un orden mundial más equitativo en lo que respecta a la información, la educación, la cultura, la investigación, las condiciones socioeconómicas e incluso la estabilidad política [4].

4. Sustentabilidad

Según la UNESCO, para el desarrollo de sociedades sostenibles es necesario el avance no solo de las dimensiones medioambientales, sino también sociales, culturales y económicas. Dependiendo de cómo se apliquen las tecnologías de la IA en

naciones con distintos grados de desarrollo, pueden ayudar o dificultar la consecución de los objetivos de sustentabilidad [1].

La eficiencia energética se refiere a que los sistemas de IA deben diseñarse minimizando el consumo y, por ende, reducir su huella de carbono; el ciclo de vida es considerar los impactos ambientales y sociales de los sistemas de IA durante su vida útil y posterior a ella, desde la producción hasta el desecho; y, la interoperabilidad significa que los sistemas de IA deben ser compatibles con otros sistemas y tecnologías existentes, lo que permite reducir la necesidad de crear nuevos sistemas y reducir la duplicación de esfuerzos [4].

5. Derecho a la intimidad y protección de datos

Para la UNESCO la privacidad debe ser respetada, salvaguardada y promovida porque es un derecho humano fundamental y, por tanto, defenderse como parte de la dignidad humana. La recopilación, uso, intercambio, archivo y eliminación de datos deben realizarse de conformidad con los marcos jurídicos aplicables [1].

Con el propósito de establecer normas globales de protección de datos y procedimientos de gobernanza, sustentados por sistemas jurídicos, se recomienda un enfoque multilateral. Los principios y reglamentos que rigen la cosecha, el uso y la difusión de datos personales, junto con el ejercicio de los derechos por parte de los interesados, deben asimilarse dentro de las leyes locales de protección de datos y cualquier mecanismo que los acompañe. Estos marcos también deben garantizar la presencia de una justificación objetiva y jurídica válida para el tratamiento de datos personales, lo que incluye la obtención de un consentimiento informado. El consentimiento informado implica obtener la autorización de una persona para el tratamiento lícito de su información; de lo contrario, debe considerarse un abuso [6].

La intimidad es el hecho de que los datos personales no sean divulgados. Los datos personales son cualquier información relacionada con una persona física identificada o identificable. Por ejemplo, información directa, como ya se dijo antes en otro capítulo: nombre y apellido, dirección física y dirección de correo electrónico, teléfono, fecha de nacimiento, género, nacionalidad, pasaporte o número de identificación oficial, información financiera, tarjeta de crédito o detalles de cuenta bancaria, número de seguro social, RFC o número de contribuyente, CURP o DNI, información médica, empleo, educación, calificaciones, dirección IP y geolocalización.

6. Supervisión y decisión humanas

Este principio se refiere a la atribución de responsabilidad, esto es, la capacidad de identificar quién es responsable de las decisiones de los sistemas de IA. Las decisiones son tomadas automáticamente por algoritmos y modelos matemáticos basados en datos. Por lo tanto, puede ser difícil identificar quién es responsable de las decisiones de la IA. La supervisión humana se refiere, tanto al control individual como colectivo, si fuera necesario [1].

La decisión de ceder el control a la IA en determinadas situaciones puede ser tomada por las personas por razones de eficiencia, pero solo será así en circunstancias limitadas. Aunque los humanos puedan utilizar sistemas de IA para ayudarles a tomar decisiones y realizar tareas, el sistema no podrá sustituir la responsabilidad última del operador o de quien dé la orden, y su obligación de rendir cuentas de sus actos, por lo tanto, hay decisiones que no deben dejarse en manos de las máquinas; por ejemplo, decisiones que pongan en riesgo la vida o la salud de otra persona [7].

7. Transparencia y explicabilidad

Para la UNESCO los requisitos previos fundamentales para garantizar el respeto, la protección y la promoción de los derechos humanos, las libertades fundamentales y los principios éticos son la transparencia y la explicabilidad de los sistemas de IA [1].

La transparencia es necesaria para el funcionamiento eficaz de las normativas nacionales e internacionales, cuando existan, en materia de responsabilidad, y facilitar el examen de las decisiones adoptadas por la IA y fortalecer así los ámbitos jurídicos en los que pueden emplearse estos sistemas.

El grado de transparencia y de explicabilidad debe ser proporcional al contexto y al efecto en cualquier régimen democrático.

El público debe ser debidamente informado cuando una decisión provenga de algoritmos de IA, especialmente si afecta la seguridad o los derechos humanos. Por ello, el público debe tener la oportunidad de pedir aclaraciones e información al agente responsable de la IA o a las entidades gubernamentales pertinentes en estas circunstancias [8].

Los sujetos de la IA deben poder comprender los fundamentos de cualquier decisión que les atañe, además de tener la opción de reclamar ante un miembro calificado del personal de una empresa del sector privado o de una institución pública,

para que se revise y, en su caso, se modifique cualquier resultado si le es injustamente adverso [9].

El objetivo de la transparencia es dar la información adecuada a cada destinatario para que pueda comprender y desarrollar confianza en el sistema y, en última instancia, en su propio gobierno o empresa que le da un servicio. La transparencia puede ayudar a las personas a comprender cómo se lleva a cabo cada etapa y revela información si se han establecido las garantías adecuadas, tales como medidas de seguridad o imparcialidad.

Por su parte, la explicabilidad es la capacidad de hacer comprensibles los resultados de los sistemas de IA. Esto también incluye comprender la entrada, la salida, el funcionamiento y la contribución de cada elemento de construcción algorítmica al producto final. Los resultados y los subprocesos también deben ser comprensibles y rastreables [9].

8. Responsabilidad y rendición de cuentas

La UNESCO enfatiza que la responsabilidad debe estar en conformidad con la legislación internacional y nacional de cada país, en particular en materia de derechos humanos. Los actores de la IA deben asumir siempre su deber ético y responsabilidad de las decisiones y acciones que se basen de algún modo en un sistema de IA. Para garantizar la responsabilidad deben diseñarse procesos adecuados de supervisión, de impacto y de evaluación a través de auditorías de cada proceso [1]. Este enfoque multifacético busca reforzar el marco ético en materia de rendición de cuentas y promueve una cultura de transparencia, confianza e implantación responsable de la IA.

9. Sensibilización y educación

La UNESCO aconseja que, para garantizar la eficacia de las políticas públicas, es imperativo que los gobiernos, las organizaciones no gubernamentales, el mundo académico, los medios de comunicación, los líderes comunitarios, las asociaciones civiles y el sector privado colaboren en la promoción de la alfabetización mediática, el compromiso cívico y la formación en competencias digitales. Estos esfuerzos deben tener en cuenta la diversidad cultural y lingüística existente. Se calcula que en la actualidad hay unas doscientas lenguas oficiales reconocidas en el mundo [1].

Además, debe hacerse un esfuerzo concertado para cultivar una cultura de IA ética en todos los sectores, incluidos el académico, la investigación, el gobierno y la

industria. Para ello, los sistemas de IA deben diseñarse teniendo en cuenta consideraciones prácticas responsables, como la transparencia, la rendición de cuentas y el respeto de los derechos humanos. Establecer mecanismos de supervisión y evaluación continuas de los sistemas de IA para garantizar su conformidad con las normas y estándares éticos [10].

10. Gobernanza y colaboración adaptativas

La UNESCO recomienda que las grandes empresas multinacionales de uso masivo de información deben respetar tanto la soberanía nacional como el derecho internacional. Los países son libres de regular los datos creados en su territorio o que pasan por él, y de adoptar medidas para una vigilancia eficaz del uso de la información de sus ciudadanos, incluida la protección y el respeto del derecho a la intimidad, y otras normas de derecho fundamental [1].

Para garantizar la distribución equitativa de los beneficios y la promoción del desarrollo sostenible en IA, es crucial contar con la participación de todas las entidades pertinentes, incluidos los gobiernos, las organizaciones no gubernamentales, la sociedad civil, el mundo académico, los medios de comunicación, los educadores y los responsables de la toma de decisiones de los sectores público y privado. La colaboración entre las partes involucradas es necesaria para facilitar la adopción de normas abiertas e interoperables [11].

El organismo multilateral también recomienda tomar medidas para tener en cuenta los cambios tecnológicos, la creación de nuevos grupos y la participación significativa de comunidades e individuos marginados, así como, el respeto a la autogestión de sus datos. La alfabetización digital puede ser posible en todos los ámbitos, sin traicionar costumbres o cultura [12].

En diciembre de 2022 se realizó el “Primer foro global sobre la ética de la inteligencia artificial”, con el tema “Asegurando la inclusión en el mundo de la IA” y fue patrocinado por UNESCO y organizado por la República Checa en el marco de la presidencia del Consejo de la Unión Europea. Este foro marcó nuevas pautas en la construcción de una coalición internacional para garantizar el desarrollo y el uso éticos de la inteligencia artificial en todo el mundo.

Conclusiones parciales

Hasta aquí los diez principios y valores que la UNESCO recomienda para la creación, desarrollo, uso y aplicación de los sistemas de IA firmados por 193 países. Debe considerarse su carácter propositivo y que, desafortunadamente, algunos países no son signatarios del acuerdo internacional y, por tanto, no se sienten obligados a acatar estas recomendaciones.

Sin embargo, podría decirse, que es lo más cercano a un Acuerdo Internacional avalado por la Organización de las Naciones Unidas en materia de uso responsable de sistemas de IA. La Recomendación es exhaustiva y abarcan una amplia gama de temas. Las recomendaciones específicas piden que el desarrollo y el uso de las tecnologías de IA respeten la dignidad, los derechos y el bienestar humanos, así como la protección del medio ambiente y la justicia social. Esto puede servir de guía a todos los países y empresas del sector.

Pide que la IA aumente y potencie las capacidades humanas, no que las disminuya. Una IA que sirva al ser humano y que mejore y amplifique sus capacidades, en lugar de sustituirlas.

La recomendación trata de reconocer los beneficios que la IA ofrece a la sociedad y reducir los peligros que ésta puede significar. Al abordar cuestiones de transparencia, rendición de cuentas, privacidad y ofrecer aspectos políticos orientados a la acción sobre gobernanza de datos, educación, cultura, trabajo, salud y economía, garantiza que las transformaciones digitales promuevan los derechos humanos y ayuden a alcanzar los Objetivos de Desarrollo Sostenible de la ONU.

Finalmente, el punto central es que sienta las bases de los instrumentos normativos futuros que podrían coadyuvar en su ejecución y que las medidas adecuadas para garantizar que la ética se aplique en los hechos. Se espera que la Recomendación ayude a los países y empresas a mejorar sus marcos ético-normativos desde una visión cosmopolita.

Referencias

[1] UNESCO, “Recommendation on the ethics of artificial intelligence,” 2021, acceso feb. 2023. [En línea]. Disponible: https://bsu.buap.mx/b2m

[2] B. Shneiderman, “Human-centered artificial intelligence: Reliable, safe & trustworthy,” Int. J. Hum. Comput. Interact., vol. 36, no. 6, pp. 495-504, 2020. arXiv:2002.04087v2.

[3] A. Tsamados, N. Aggarwal, J. Cowls, J. Morley, H. Roberts, M. Taddeo y L. Floridi, "The ethics of algorithms: key problems and solutions," *AI & Soc.*, vol. 37, no. 1, pp. 215-230, Feb. 2022. doi: 10.1007/s00146-021-01154-8.

[4] P. Boddington, "Normative Modes: Codes and Standards," en *The Oxford Handbook of Ethics of AI*, M. Dubber, F. Pasquale, y S. Das, Eds. Oxford: Oxford University Press, 2020, pp. 124-140.

[5] J. Cowls, A. Tsamados, M. Taddeo y L. Floridi, "The AI gambit: leveraging artificial intelligence to combat climate change—opportunities, challenges, and recommendations," *AI & Soc.*, vol. 38, pp. 283-307, 2023. doi:10.1007/s00146-021-01228-6.

[6] C. Véliz, "Privacy Is Power: Why and How You Should Take Back Control of Your." NY: Penguin Random House, 2022.

[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," Algorithmwatch.org. Acceso feb. 2023. [En línea]. Disponible: https://algorithmwatch.org/

[8] A. Smith, "Using Artificial Intelligence and Algorithms," FTC.gov. Acceso feb. 2023. [En línea]. Disponible: https://www.ftc.gov/business-guidance/blog/2020/04/using-artificial-intelligence-and-algorithms.

[9] N. Diakopoulos, "Transparency. Accountability, Transparency, and Algorithms," in *The Oxford Handbook of Ethics of AI*, M. Dubber, F. Pasquale, y S. Das, Eds. Oxford, UK: Oxford University Press, 2020, pp. 197-213.

[10] D. Peters, P. H. Kuss, T. Calders y F. Schram, "Responsible AI—two frameworks for ethical design practice," *IEEE Trans. on Tech. and Soc.*, vol. 1, no. 1, pp. 34-47, 2020. Disponible: https://bsu.buap.mx/ciQ

[11] M. Taddeo y L. Floridi, "How AI Can Be a Force for Good – An Ethical Framework to Harness the Potential of AI While Keeping Humans in Control," en *Ethics, Governance, and Policies in Artificial Intelligence*, L. Floridi, Ed. Cham: Springer International Publishing, 2021, pp. 91-96.

[12] T. Powers y J. Ganascia, "The Ethics of the Ethics of AI," en *The Oxford Handbook of Ethics of AI*, M. Dubber, F. Pasquale, y S. Das, Editors, Oxford University Press, 2020, pp. 26-51.