



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

ASOCIACIÓN EN IA EN BENEFICIO DE LAS PERSONAS Y LA SOCIEDAD, RETOS Y PERSPECTIVAS

Introducción

La PAI es la “Asociación sobre inteligencia artificial en beneficio de las personas y la sociedad” (Partnership on AI to Benefit People and Society) y es una organización sin fines de lucro con sede en San Francisco, California, que reúne a organizaciones académicas, de la sociedad civil, a empresas tecnológicas y de los medios de comunicación para abordar cuestiones sustanciales, básicamente sobre el futuro de la IA, pero también otros importantes retos mundiales como el cambio climático, la alimentación, la desigualdad, la salud y la educación. Tienen cinco programas que contribuyen al desarrollo de recursos, recomendaciones y mejores prácticas para la IA: inteligencia artificial e integridad mediática; IA, trabajo y economía; justicia, transparencia y responsabilidad en aprendizaje automático (ML); investigación y diseño inclusivos; y seguridad para la IA crítica. En este capítulo se destaca que, si bien PAI ha sido criticada por estar dominada en gran medida por grandes empresas tecnológicas, lo que podría limitar su autonomía y autocrítica, es una organización valiosa e influyente en el campo de la ética y la gobernanza de la IA, y desempeña un papel importante en la promoción del desarrollo y el despliegue responsables, no solo de la IA, sino de la tecnología y los medios de comunicación en general.

Los principios éticos que promueve PAI

Partnership on AI [1] es una organización sin fines de lucro fundada en septiembre de 2016 por un grupo de compañías de tecnología líderes, incluyendo Deep Mind, Amazon, Facebook, Google, Microsoft e IBM, con el objetivo de fomentar la colaboración y la investigación en la inteligencia artificial (IA) y explorar cómo la IA puede beneficiar a las personas y a la sociedad en general en un marco ético compartido [2]. En 2023 contaba con 104 socios de 16 países.

Según PAI, su misión es estudiar y formular mejores prácticas en el desarrollo, la implementación y la regulación de la IA, con un enfoque en la ética, la transparencia y la responsabilidad social [2]. La organización reúne a expertos en IA, investigadores, defensores de los derechos civiles y otros interesados en el campo para trabajar juntos en cuestiones críticas relacionadas con el mundo digital. Especialmente se relaciona con las empresas líderes del mercado mundial con quienes trabaja de manera conjunta en sus aplicaciones; entre las principales empresas afiliadas están: Adobe, Amazon, Apple, Deep Mind, Google, IBM, Intel, Intuit, Meta, Microsoft, Samsung y Sony. Medios de comunicación como la BBC de Londres, el New York Times y la Agencia AP; y 60 organizaciones sin fines de lucro, entre las que destaca OpenAI, UNICEF, el Instituto Alan Turing y el instituto Carnegie, entre muchos otros [3].

PAI ha establecido varias iniciativas y grupos de trabajo que se centran en temas como ética y responsabilidad de la IA, inclusión, diversidad, privacidad, seguridad, educación, difusión del conocimiento y colaboración intersectorial. La organización también ha creado un foro para el diálogo público sobre la IA y ha trabajado con gobiernos y organizaciones alrededor del mundo para desarrollar políticas y regulaciones adecuadas para la IA [4].

La asociación publica permanentemente guías como herramientas y recursos para ayudar a los desarrolladores y otros profesionales a implementar prácticas responsables de IA, por lo que ha establecido una serie de grupos de trabajo que se centran en temas específicos tecnológicos.

Los cinco programas que mantienen son: inteligencia artificial e integridad mediática; IA, trabajo y economía; justicia, transparencia y responsabilidad en aprendizaje automático (ML); investigación y diseño inclusivos; y seguridad para la IA crítica. Cada uno de los anteriores ítems con producción específica, artículos y herramientas metodológicas [1].

PAI tiene seis pilares temáticos en los que se funda la asociación y representan conjuntos de temas en los que visualizan los mayores riesgos y oportunidades para la IA:

1. IA de seguridad crítica

Las mejoras en la asistencia sanitaria y el transporte que requieren un alto nivel de seguridad podrían beneficiarse enormemente del uso de la IA. Existe la posibilidad de mejorar la calidad de vida y evitar miles de muertes mediante el uso reflexivo y

estratégico del reconocimiento de patrones, la toma de decisiones automatizada y las tecnologías robóticas [2] [5].

Sin embargo, en los casos en que la IA se utiliza para tomar decisiones junto a los seres humanos o en su lugar, se debe asegurar que sea fiable, segura y respetuosa con los valores y preferencias de las personas a las que afecta.

2. Una IA justa, transparente y responsable

La IA, y especialmente el aprendizaje automático, tiene potencial para añadir valor a la sociedad mediante el reconocimiento de patrones y el análisis de datos, que pueden utilizarse para desarrollar sistemas de diagnóstico y motores de recomendación útiles en campos como la biomedicina, la salud pública, la seguridad, la justicia penal, la educación y la sostenibilidad [2] [6].

Sin embargo, es importante ser conscientes de los posibles sesgos ocultos en los datos utilizados para desarrollar sistemas de IA, así como de otros factores que podrían afectar la calidad de los resultados de los sistemas [7].

3. IA, trabajo y economía

El progreso de la IA tendrá un impacto significativo en el mercado laboral y en la naturaleza del trabajo. Aunque la introducción de nuevas tecnologías promete aportar un valor considerable a la economía, también puede provocar trastornos en el empleo y una reducción de la necesidad de determinados puestos debido a la automatización [2] [8].

Por ello, se está debatiendo cómo minimizar los posibles efectos negativos de la IA en el mercado laboral y garantizar que los beneficios del avance tecnológico se distribuyan de forma equitativa sin dejar de fomentar la competencia y la innovación [9]. Es necesario estudiar y comprender las mejores estrategias para avanzar y participar en los debates en curso [10].

4. Colaboración entre personas y sistemas de IA

La IA puede mejorar la percepción, la cognición y la capacidad de resolución de problemas del ser humano. Algunos ejemplos son las tecnologías de IA que pueden ayudar a los médicos a realizar diagnósticos más precisos y oportunos, y la asistencia de IA proporcionada a los conductores de vehículos para ayudarles a evitar situaciones peligrosas y accidentes [2] [11].

La investigación y el desarrollo en el ámbito de la colaboración entre la IA y el ser humano son necesarios para desarrollar las mejores prácticas. Esto incluye métodos para proporcionar a las personas claridad sobre la comprensión y la confianza que los sistemas de IA tienen sobre aplicaciones específicas, formas de coordinar las contribuciones humanas a la resolución de problemas, y permitir que los sistemas trabajen con las personas de forma armónica.

5. Influencias sociales de la IA

Es un hecho que los avances de la IA afectarán a diversos aspectos de la sociedad y la vida humana, como la privacidad, la democracia, la justicia penal, los medios de comunicación y los derechos humanos. El uso de tecnologías de IA para personalizar la información y ofrecer recomendaciones puede ser beneficioso, pero también existe la posibilidad de que estas tecnologías manipulen involuntaria o intencionadamente a los usuarios e influyan en sus opiniones [2] [12].

Por tanto, es necesario un diálogo abierto y una colaboración reflexiva sobre los posibles efectos de la IA en la sociedad y las personas. El objetivo de PAI es fomentar el debate y la exploración de las formas sutiles y significativas en que la IA podría afectar la cosmovisión de una cultura.

6. IA y bien social

La IA tiene el potencial de promover el bienestar público en áreas como la educación, la vivienda, la salud pública y la sostenibilidad. En Partnership on AI consideran que la colaboración con organizaciones, tanto públicas como privadas, incluidos el mundo académico, las sociedades científicas, las ONGs, los emprendedores sociales y los particulares interesados, puede ayudar a generar debates y catalizar esfuerzos para abordar los retos más acuciantes de la sociedad [2].

Algunos ejemplos de cómo la inteligencia artificial puede ayudar específicamente son: eficiencia energética, detección de fraude eléctrico, vehículos autónomos, gestión del tráfico, salud pública, biología molecular, entre muchos otros campos. De lo anterior, los integrantes de PAI han propuesto seis principios básicos para una ética de la IA:

1. Una IA benéfica

Esta afirmación significa que habrá un esfuerzo para garantizar que las tecnologías de IA se utilicen de forma que tengan un impacto positivo y den poder al mayor

número de personas posible. Sugiere que habrá que centrarse en crear aplicaciones equitativas y accesibles de la IA que beneficien a las personas y a la sociedad en su conjunto, en lugar de beneficiar solo a unos pocos elegidos [2].

2. Escuchar y aprender de todas las voces

PAI hará un esfuerzo por involucrar y comunicarse activamente con el público y las partes interesadas sobre sus iniciativas de IA. La intención de Partnership on AI es educar a la gente sobre el enfoque y los objetivos del trabajo de IA, y buscar opiniones y responder a cualquier pregunta o preocupación que puedan generarse. Esto sugiere un compromiso con la transparencia y la comunicación abierta para garantizar que el desarrollo y la aplicación de las tecnologías de IA se ajuste a las necesidades y los valores de la comunidad en general [2].

3. Implicación de la investigación y diálogo abiertos

Debe existir el compromiso de investigar y facilitar el debate sobre las implicaciones éticas, sociales, económicas y jurídicas de la IA. La intención es garantizar que el desarrollo y el uso de la IA se ajuste a los principios y valores éticos, y abordar de forma proactiva las posibles repercusiones negativas. Esto sugiere un reconocimiento de la necesidad de una toma de decisiones responsable y transparente en relación con el desarrollo y la aplicación de las tecnologías de IA [2].

4. Compromiso activo de todas las partes

Existe la convicción de que la investigación y el desarrollo deben comprometerse activamente y responsabilizarse ante un conjunto diverso de partes interesadas. Estas partes incluyen a individuos, organizaciones y comunidades que pueden verse afectados por la creación y utilización de tecnologías de IA. El objetivo es garantizar que la IA responda a las necesidades e inquietudes de un amplio espectro de personas y grupos, y que las ventajas se repartan equitativamente entre toda la sociedad. [2].

5. Compromiso y atención a todas las partes involucradas

Existe el compromiso de colaborar con la comunidad empresarial en el desarrollo y la aplicación de las tecnologías de IA. La intención es garantizar que se comprendan y aborden las preocupaciones y oportunidades específicas de las distintas industrias y sectores [2].

La inclusión de representantes de la comunidad empresarial ofrece la oportunidad de trabajar en colaboración para identificar posibles casos de uso y áreas de innovación. Esto sugiere un reconocimiento de la importancia de implicar a todas las partes interesadas en el desarrollo y la aplicación de las tecnologías de IA.

6. Maximizar beneficios y enfrentar los retos

Trabajando para preservar la privacidad y la seguridad de las personas, PAI se esfuerza por aprovechar las ventajas de la tecnología de IA y gestionar las posibles dificultades, intentando comprender y respetar los intereses de todas las partes a las que puedan afectar los posibles avances de la IA [2]. PAI trabaja para mantener el compromiso social de los grupos de investigación e ingeniería de IA sobre las posibles repercusiones de la tecnología en general. Garantizar que la investigación y la tecnología de la IA sean sólidas, dignas de confianza y fiables, y que funcionen dentro de límites seguros. Oponerse a la creación y aplicación de tecnología de IA que pueda violar los derechos humanos o los acuerdos internacionales, apoyando al mismo tiempo las medidas de seguridad y las tecnologías beneficiosas. Precisamente del seno de la PAI surgió en marzo de 2023 la propuesta de hacer una tregua y frenar el desarrollo de la IA, más allá de GPT4, mientras se desconozcan sus implicaciones y riesgos a largo plazo [21].

7. Inteligibilidad y explicabilidad

Es fundamental que las personas comprendan e interpreten el funcionamiento de los sistemas de IA para explicar eficazmente su tecnología. Comprender los sistemas de IA es un primer paso hacia la mejora de la accesibilidad y el fomento de la confianza entre las personas que interactúan abiertamente con ellos. Ofrecer explicaciones no solo promueve la aceptación pública, sino que también ayuda a identificar y resolver posibles sesgos o errores [2].

8. Cooperación, confianza y apertura entre científicos

Fomentar una cultura de colaboración y transparencia entre los científicos e ingenieros de IA para alcanzar sus objetivos compartidos [2]. La colaboración y la confianza entre científicos e ingenieros de IA son esenciales para avanzar en el desarrollo de la tecnología de forma ética y sostenible. Una cultura abierta promueve el intercambio de conocimientos, la crítica constructiva y la revisión por pares, lo que puede conducir a mejores soluciones y minimizar los riesgos de consecuencias no deseadas. En última instancia, estos esfuerzos también pueden contribuir a fomentar la confianza pública en la IA y promover su uso responsable.

Haciendo que la IA sea inclusiva

Además de lo anterior, en el documento “Haciendo que la IA sea inclusiva, cuatro principios rectores para el compromiso ético” [13] de Tina M. Park de julio de 2022, PAI ha propuesto:

1. Trabajo participativo

Reconocer el valor de la participación de los usuarios y del público es crucial para fomentar un enfoque integrador del trabajo. Es imperativo que todo el mundo tenga la oportunidad de disfrutar de los beneficios que aporta la IA [13].

Para construir sistemas de IA se necesitan grandes cantidades de datos, por lo que la tecnología depende mucho de la participación del público. Los datos que los usuarios proporcionan son necesarios para que la tecnología pueda funcionar correctamente y generar beneficios [14].

La participación se refiere a cualquier contribución directa o indirecta a la creación, desarrollo, implementación y sostenibilidad de un sistema de IA. Esto significa que cualquier acción que ayude a crear o mejorar los sistemas se considera una forma de participación. También lo son la provisión de datos, la retroalimentación sobre el uso de los sistemas o la capacitación para su desarrollo [15]. La producción de textos, fotografías, videos, audios, y etiquetado entran en la misma categoría.

La importancia de la participación pública en la creación de estos sistemas radica en que están diseñados para satisfacer las necesidades y expectativas de los usuarios. Sin la participación de una amplia gama de personas que representen diversas perspectivas y necesidades, no se puede lograr esto [16].

2. Terminar con asimetrías

Para Park [1] la participación de las partes interesadas debe abordar las asimetrías de poder inherentes. Muchos ciudadanos dudan a la hora de apoyar iniciativas participativas dirigidas por empresas u otras entidades, especialmente porque hay escepticismo en las comunidades históricamente oprimidas y explotadas [13].

En el pasado, los grupos como la comunidad afroamericana de Estados Unidos han sido tratados injustamente, lo que ha llevado a que tengan desconfianza hacia los sistemas tecnológicos que se desarrollan. Por esta razón, pedirles que participen

en la creación de sistemas de IA puede generar suspicacia, ya que sienten que se les está pidiendo que trabajen sin remuneración para beneficiar a otros, en lugar de ser incluidos como verdaderos socios en el desarrollo de la tecnología [1]. Cosa que, dicho sea, es verdad.

Los usuarios y el público en general carecen a menudo de la capacidad de obtener información privilegiada o de influir significativamente en las decisiones cruciales que toman las organizaciones de IA y aprendizaje automático. Los desarrolladores, que actúan como intermediarios, tienen el poder de controlar el acceso a la información y tomar decisiones importantes. Sin embargo, incluso los propios desarrolladores pueden no tener la autoridad última para conceder dicho acceso. Además, los desequilibrios de poder derivados de factores históricos como la segregación y la discriminación contribuyen a estas relaciones inequitativas [1].

Reconocer y abordar esta dinámica es esencial para establecer relaciones respetuosas y mutuamente beneficiosas entre quienes promueven las iniciativas de IA y los miembros de la comunidad implicados. Es crucial reconocer el impacto de factores sociales como el racismo y la misoginia, ya que estas dinámicas pueden tener un efecto continuo en la relación, aunque las interacciones interpersonales parezcan más equitativas [1].

3. Inclusión y participación

La inclusión y la participación pueden integrarse en todas las fases del ciclo de vida del sistema. Sigue siendo frecuente encontrar prácticas participativas implementadas en el desarrollo de IA hacia el final del ciclo de vida, en lugar de estar integradas a lo largo de todo el proceso. En otras palabras, estas prácticas participativas son, a menudo, una ocurrencia tardía en lugar de un componente central del desarrollo de la IA [13].

Debido al enfoque anterior es esencial garantizar que las prácticas participativas se integren plenamente en todo el ciclo de vida del desarrollo de la IA para promover un proceso de desarrollo más inclusivo y colaborativo que tenga en cuenta las necesidades y perspectivas de todas las partes interesadas, no solo cuando el producto ya está hecho [16].

Se afirma que las personas que abogan por la equidad y la inclusión han destacado la importancia de implicar a los usuarios y a los miembros de las comunidades afectadas en todas las fases del proceso de desarrollo de los proyectos de IA. Esto significa que las aportaciones y comentarios de todas las partes interesadas deben

buscarse e incorporarse desde las primeras fases del proyecto hasta el despliegue final.

La manera más eficaz de lograrlo es aplicando prácticas participativas integradoras que creen relaciones significativas entre todas las partes. Al darles el espacio para dirigir el propósito y la intención del proyecto de IA los miembros de la comunidad pueden estar facultados para garantizar que el proyecto se alinea con sus valores y prioridades, lo que puede dar lugar a resultados más éticos y equitativos [18].

4. Transversalidad inclusiva

Según Park [1], incorporar la inclusión y la participación es de vital importancia a la hora de aplicar otros principios de la IA responsable. A menudo, los debates sobre el uso ético de la IA se centran en normas individuales por separado, pasando por alto la interconexión de estos principios [13]. Por ejemplo, una norma puede hacer hincapié en el acceso universal a la tecnología de la IA, mientras que otra destaca la importancia de un uso justo. Si bien estas normas tienen importancia por sí mismas, es igualmente vital que estén interconectadas y se apoyen mutuamente, trabajando en armonía para lograr prácticas de IA responsables.

Conclusiones parciales

De acuerdo con lo planteado por Partnership on AI para maximizar los beneficios y afrontar los retos de la IA, es crucial garantizar que los sistemas sean favorables para toda la sociedad. Esto requiere el compromiso activo de todas las partes involucradas en su desarrollo y despliegue, incluidos científicos, ingenieros, distribuidores, políticos y usuarios. Alcanzar estos objetivos también requiere una investigación que tenga en cuenta las implicaciones de la tecnología de IA para la sociedad y el medio ambiente.

Para lograr una comprensión holística de las consecuencias de la IA, es esencial escuchar y aprender de todas las voces del sector, incluidas las de procedencias y perspectivas que no estén a favor del desarrollo de la tecnología. Esto significa crear una cultura de apertura y cooperación entre los científicos e ingenieros de IA, así como colaborar con comunidades más amplias y expertos de otras disciplinas, por ejemplo, ética, filosofía, derecho, sociología, antropología, economía y ciencias cognitivas.

PAI ha sido criticada porque, aunque la organización incluye a una amplia gama de sectores representados, sigue estando dominada en gran medida por las grandes empresas tecnológicas, lo que puede limitar su capacidad para ser plenamente autónoma, autocrítica y representar los intereses globales [19].

También se ha señalado de ser una panacea para protegerse las grandes empresas de un verdadero marco regulatorio internacional [22]. Otra crítica es que las iniciativas y proyectos de PAI tienden a centrarse en la autorregulación voluntaria y las normas impulsadas por la industria, en lugar de reglamentos jurídicamente vinculantes y supervisión gubernamental. Sin una normativa y supervisión más estrictas, las empresas pueden no tener que rendir cuentas de sus actos y dar prioridad a los beneficios económicos frente a las consideraciones éticas [20].

A pesar de estas críticas, PAI sigue siendo una organización influyente y necesaria en el campo de la ética y la gobernanza de la IA, y desempeña un papel decisivo en la promoción del desarrollo y el despliegue responsables de la IA.

En conclusión, PAI representa una iniciativa prometedora que puede ayudar a afrontar los retos y liberar el vasto potencial de la IA para la mejora de la humanidad. Al fomentar la colaboración y la cooperación entre las partes interesadas de diversos sectores y regiones, la asociación puede contribuir al desarrollo de una IA responsable, ética e integradora. Gracias a la dedicación persistente y al compromiso colectivo de todos los implicados, el futuro de la IA, desde PAI, es prometedor para servir al bien común y mejorar el bienestar humano. Esta organización sin ánimo de lucro, impulsada por su misión, sigue dedicada a fomentar la colaboración entre diversas voces de distintos sectores, disciplinas y grupos demográficos para cumplir con su objetivo primordial que es que los avances en IA produzcan resultados positivos, tanto para las personas como para la sociedad en su conjunto [2].

Referencias

- [1] Partnership on AI, “Partnership on AI is bringing together diverse voice from across the AI community,” Partnershiponai.org. [En línea]. Acceso mar. 2023. [En línea]. Disponible: <https://partnershiponai.org/>
- [2] Partnership on AI, “About us,” Partnershiponai.org. Acceso mar. 2023. [En línea]. Disponible: <https://partnershiponai.org/about/>
- [3] Partnership on AI, “Our Funding. PAI relies on a multitude of funding sources to accomplish our goals,” Partnershiponai.org. Acceso mar. 2023. [En línea]. Disponible: <https://partnershiponai.org/funding/>

- [4] Partnership on AI, "Our Resource Library. Our collected papers, resources, and other outputs," Partnershiponai.org. Acceso mar. 2023. [En línea]. Disponible: <https://partnershiponai.org/resources/>
- [5] Partnership on AI, "Safety Critical AI," Partnershiponai.org. Acceso mar. 2023. [En línea]. Disponible: <https://partnershiponai.org/program/safety-critical-ai/>
- [6] Partnership on AI, "Fairness, Transparency, and Accountability & About ML," Partnershiponai.org. Acceso mar. 2023. [En línea]. Disponible: <https://partnershiponai.org/program/fairness-transparency-and-accountability-about-ml/>
- [7] J. Fjeld, N. Achten, H. Hilligoss, A. C. Nagy, y M. Srikumar, "Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-based Approaches to Principles for AI," Berkman Klein Center for Internet & Society, Research Publication No. 2020-1, 2020, doi:10.2139/ssrn.3518482
- [8] Partnership on AI, "AI, Labor, and the Economy," Partnershiponai.org. Acceso mar. 2023. [En línea]. Disponible: <https://partnershiponai.org/program/ai-labor-and-the-economy/>
- [9] D. Acemoglu, D. Autor, J. Hazell y P. Restrepo, "Artificial Intelligence and Jobs: Evidence from Online Vacancies," *J. Labor Econ.*, vol. 40, 2022. Disponible: <https://bsu.buap.mx/ciR>
- [10] P. Moradi y K. Levy, "The Future of Work in the Age of AI: Displacement or Risk-Shifting?," en *The Oxford Handbook of Ethics of AI*, M. Dubber, F. Pasquale, y S. Das, Eds., Oxford University Press, 2020, pp. 270-288.
- [11] Partnership on AI, "PAI Research Promotes Responsible Collaborations between People and AI Systems," Partnershiponai.org. Acceso mar. 2023. [En línea]. Disponible: <https://partnershiponai.org/cpais-research/>
- [12] Partnership on AI, "Fairer Algorithmic Decision-Making & Its Consequences," Partnershiponai.org. Acceso mar. 2023. [En línea]. Disponible: <https://partnershiponai.org/paper/fairer-algorithmic-decision-making-and-its-consequences/>.
- [13] T. Park, "Making AI Inclusive: 4 Guiding Principles for Ethical Engagement," Partnershiponai.org. Acceso mar. 2023. [En línea]. Disponible: <https://partnershiponai.org/paper/making-ai-inclusive-4-guiding-principles-for-ethical-engagement/>.
- [14] L. Floridi, "Establishing the Rules for Building Trustworthy AI," en *Ethics, Governance, and Policies in Artificial Intelligence*, L. Floridi, Ed., Springer Intern. Publish., 2021, pp. 41-45.
- [15] L. Floridi et al., "An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations," en *Ethics, Governance, and Policies in Artificial Intelligence*, L. Floridi, Ed., Springer Intern. Publish., 2021, pp. 19-39, doi: 10.1007/s11023-018-9482-5.
- [16] A. Tsamados, N. Aggarwal, J. Cows, 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.

- [17] M. Taddeo, T. McCutcheon, y L. Floridi, "Trusting artificial intelligence in cybersecurity is a double-edged sword," *Nat. Mach. Intell.*, vol. 1, pp. 557-560, 2019.
- [18] 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.
- [19] R. Ochigame, "The invention of 'ethical AI,'" Acceso mar. 2023. [En línea]. Disponible: <https://theintercept.com/2019/12/20/mit-ethical-ai-artificial-intelligence/>
- [20] Rességuier y R. Rodrigues, "AI ethics should not remain toothless! A call to bring back the teeth of ethics," *Big Data & Soc.*, vol. 7, no. 2, p. 2053951720942541, 2020.
- [21] Future of Life Institute, "Pause Giant AI Experiments: An Open Letter," [Futureoflife.org](https://futureoflife.org). Acceso jun. 2023. [En línea]. Disponible: <https://bsu.buap.mx/b8d>
- [22] T. Hagendorff, "The Ethics of AI Ethics: An Evaluation of Guidelines," *Mind and Mach.*, vol. 30, no. 1, pp. 99-120, 2020. DOI: 10.1007/s11023-020-09517-8