

AI ETHICS RESEARCH THEMATIC AREA NARRATIVE IN ENGLISH ARABIC FRENCH PORTUGUESE AND SPANISH

Rachel Adams , Kelly Stone

Rachel Adams , Kelly Stone

©2025, RACHEL ADAMS , KELLY STONE



This work is licensed under the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/legalcode>), which permits unrestricted use, distribution, and reproduction, provided the original work is properly credited. Cette œuvre est mise à disposition selon les termes de la licence Creative Commons Attribution (<https://creativecommons.org/licenses/by/4.0/legalcode>), qui permet l'utilisation, la distribution et la reproduction sans restriction, pourvu que le mérite de la création originale soit adéquatement reconnu.

IDRC GRANT / SUBVENTION DU CRDI : - GLOBAL INDEX ON RESPONSIBLE ARTIFICIAL INTELLIGENCE

Índice Global sobre IA Responsable

Dimensión: Capacidades nacionales para una IA responsable

Subdimensión: Inversiones

Área temática: [Investigación sobre la ética de la IA](#)

Definiciones

Se puede definir la [ética](#) como "un conjunto de **principios morales** que rigen el comportamiento de una persona o la realización de una actividad determinada".¹ En este sentido, la ética puede considerarse como un conjunto de principios [basados en nociones de lo correcto y lo incorrecto](#) que ayudan a determinar lo que constituye una conducta y/o un comportamiento aceptable e inaceptable.

De este modo, se puede definir la [ética de la IA](#) como "un sistema de principios y técnicas morales",² destinado a orientar el diseño, desarrollo, uso y despliegue de las tecnologías de IA. Aunque las normas de la ética de la IA están en constante evolución, la Recomendación de la UNESCO sobre la Ética de la Inteligencia Artificial ([Recomendación](#)) articuló cuatro valores fundamentales subyacentes a la ética de la IA: (1) derechos humanos y dignidad humana; (2) vivir en sociedades pacíficas, justas e interconectadas; (3) garantizar la diversidad y la inclusión; y (4) prosperidad del medio ambiente y los ecosistemas. Además, la Organización para la Cooperación y el Desarrollo Económicos (OCDE) identificó cinco [principios basados en valores](#) para la IA, que incluyen: (1) el crecimiento inclusivo, el desarrollo sostenible y el bienestar; (2) los valores centrados en el ser humano y en la equidad; (3) la transparencia y explicabilidad; (4) la solidez, la seguridad y la protección; y (5) la rendición de cuentas.

La [investigación sobre la ética de la IA](#) es una [área de estudio](#) emergente, en el ámbito de la ética aplicada y de la filosofía de la tecnología, que se ocupa de cómo se aplican, actualmente, los principios éticos a la IA y de si constituyen [salvaguardias adecuadas](#) para hacer frente a algunos de los riesgos que plantea el desarrollo y despliegue de sistemas y herramientas de IA.

Justificaciones

Desde hace mucho tiempo que el estudio de la Inteligencia artificial suscita preocupaciones éticas. Las últimas aplicaciones de la IA han tenido consecuencias como la discriminación, la vigilancia generalizada y la falta de responsabilidad.

¹ Traducción no oficial

² Traducción no oficial

También han suscitado inquietud por el controvertido uso de la IA en la [guerra](#). Para hacer frente a estas consecuencias negativas, expertos de todo el mundo han elaborado [principios éticos](#), al objeto de diseñar una IA con más responsabilidad, equidad y privacidad. Entre estos principios se encuentran los Principios de la IA de la OCDE y la Recomendación de la UNESCO sobre la Ética de la IA, cuyo objetivo es crear una "base dinámica para la [evaluación y la orientación normativas](#) de las tecnologías de la IA, tomando como referencia la dignidad humana, el bienestar y la prevención de daños".

Los [marcos normativos](#) para la ética de la IA pretenden promover "[faros](#)" para el uso y despliegue de las tecnologías de IA con el fin de mitigar diversos riesgos, como los prejuicios, la discriminación y las amenazas a las libertades y los derechos humanos. Sin embargo, es esencial tener en cuenta la [complejidad](#) de [aplicar](#) estos principios en la práctica y prever los nuevos retos éticos que pueden surgir con la introducción de nuevas tecnologías basadas en la IA, como la IA generativa. Además, la IA puede dar lugar a diversas cuestiones éticas con un énfasis diferente en los principios específicos de la IA en los distintos países y regiones. Por consiguiente, los países deben reforzar sus capacidades en materia de investigación sobre la ética de la IA, sobre todo invirtiendo en este ámbito. El objetivo no es sólo adherirse a un conjunto de principios éticos que se están elaborando y orientan el uso responsable de las herramientas y sistemas de IA, sino también comprender las condiciones que se necesitan a nivel local para aplicar estos principios. Así será posible determinar su pertinencia y utilidad para las poblaciones locales. La inversión en investigación sobre la ética de la IA debe tratar de (1) promover una [comprensión interdisciplinaria](#) de las cuestiones éticas; (2) identificar y anticipar las implicaciones sociales, económicas y políticas de las tecnologías de IA; y (3) plantear [cuestiones críticas](#) sobre la conveniencia de utilizar la IA en diversas facetas de la vida.

Esto pone de relieve la importancia de la investigación sobre la ética de la IA para documentar el debate con datos sobre las directrices actuales, elaborar nuevos principios, analizar los enfoques y destacar las áreas importantes en las que tiene que centrarse la ética de la IA. De este modo, esta área temática pretende evaluar y explorar cómo los gobiernos y otras partes interesadas allanan el camino para esta investigación, mediante la publicación de artículos sobre la ética de la IA, el desarrollo de investigaciones y la creación de condiciones institucionales.

Identificaciones

Esta área temática examina las medidas adoptadas por los diferentes países para apoyar la investigación en el ámbito de la ética de la IA. En concreto, el análisis debe tener en cuenta: 1) los **marcos jurídicos** relativos a políticas, programas o directrices que definen la financiación y las prácticas de la investigación sobre la ética de la IA; 2) las **acciones gubernamentales** de apoyo a la investigación sobre la ética de la IA, que invierten en programas centrados en la integración de la ética de la IA en los departamentos, agencias y organismos gubernamentales, en la creación de redes de investigación, de fondos o concursos centrados en la investigación sobre la ética de la IA; y 3) los **agentes no estatales** que trabajan para hacer avanzar la ética de la IA, mediante la creación de redes o institutos de investigación independientes para formar unidades dedicadas a la ética de la IA, publicaciones en este ámbito o financiando la investigación sobre la ética de la IA.

Los *marcos jurídicos* del país pueden asumir la forma de políticas adoptadas, libros blancos o directrices. Las *acciones gubernamentales* pueden incluir proyectos de políticas o directrices, la creación de entidades gubernamentales como organismos de supervisión con la misión de formular recomendaciones políticas sobre el tema y/o hacer cumplir la normativa. Estas acciones también pueden implicar la aplicación de políticas destinadas a abordar la cuestión y/o promover la sensibilización o recopilar más datos sobre la investigación ética de la IA. Los *Agentes No Estatales* (ANE) pueden ser organizaciones no gubernamentales (ONG), pero también empresas multinacionales, organizaciones militares privadas, medios de comunicación, grupos étnicos organizados, instituciones académicas, grupos de presión, sindicatos o movimientos sociales que promueven la pertinencia e importancia de la investigación sobre la ética de la IA.

Ejemplos:

Marcos jurídicos

En una [estrategia para la IA](#), el gobierno federal alemán ha subrayado que para promover la investigación el enfoque debería ser europeo "cuando se trata de establecer directrices éticas claras, en la investigación de base y aplicada a medio y largo plazo".³ En esta estrategia, el gobierno también aboga por un enfoque "ético desde el propio diseño"⁴ [centrado en soluciones técnicas] y, en general, se ha comprometido a destinar [5 mil millones de euros](#) hasta el final de 2025, para promover la IA. Sin embargo, hay que tener en cuenta la distinción entre la investigación general sobre aplicaciones de IA y la investigación específicamente orientada a la ética de la IA.

Acciones gubernamentales

Una iniciativa gubernamental visible es, por ejemplo, la financiación de centros de competencia para la IA, financiados por el Ministerio de Educación e Investigación alemán, por los estados federales participantes y por el [DFKI](#) (uno de esos centros), que también cuenta con un [equipo de Ética para la IA](#). Además, a través de la Fundación Alemana de Investigación se financian proyectos individuales, como el [artículo](#) de Thilo Hagendorff (2022) "Blind spots in AI ethics" por el Cluster of Excellence Machine Learning-New Perspectives for Science.

Agentes no estatales

Existen organizaciones de la sociedad civil, como [Algorithm Watch](#), específicamente dedicadas a "observar, desentrañar y analizar los sistemas automatizados de toma de decisiones (ADM) y su impacto en la sociedad",⁵ que, por ejemplo, ha elaborado una base de datos mundial de [167 directrices éticas](#). Además, la Fundación Bertelsmann Stiftung está desarrollando un proyecto sobre la [Ética de los Algoritmos](#) con el que pretende contribuir "diseñando sistemas algorítmicos que conduzcan a una mayor participación de todos".⁶ En 2019, Meta concedió una [subvención de financiación inicial](#) para la apertura del Instituto de Ética en Inteligencia Artificial en la Universidad de Múnich.

³ Traducción no oficial

⁴ Traducción no oficial

⁵ Traducción no oficial

⁶ Traducción no oficial

