¿Qué es la ética de la IA?



[Artificial Intelligence](https://www.ibm.com/mx-es/think/artificial-intelligence)

**17 de septiembre de 2024**

* [¿Qué es la ética de la IA?](https://www.ibm.com/mx-es/think/topics/ai-ethics#%C2%BFQu%C3%A9+es+la+%C3%A9tica+de+la+IA%3F)
* [Establecimiento de principios para la ética de la IA](https://www.ibm.com/mx-es/think/topics/ai-ethics#Establecimiento+de+principios+para+la+%C3%A9tica+de+la+IA)
* [Principales preocupaciones de la IA en la actualidad](https://www.ibm.com/mx-es/think/topics/ai-ethics#Principales+preocupaciones+de+la+IA+en+la+actualidad)
* [Cómo establecer la ética de la IA](https://www.ibm.com/mx-es/think/topics/ai-ethics#C%C3%B3mo+establecer+la+%C3%A9tica+de+la+IA)
* [Organizaciones que promueven la ética de la IA](https://www.ibm.com/mx-es/think/topics/ai-ethics#Organizaciones+que+promueven+la+%C3%A9tica+de+la+IA)
* [El punto de vista de IBM sobre la ética de la IA](https://www.ibm.com/mx-es/think/topics/ai-ethics#El+punto+de+vista+de+IBM+sobre+la+%C3%A9tica+de+la+IA)

¿Qué es la ética de la IA?

La ética es un conjunto de principios morales que nos ayudan a discernir entre el bien y el mal. La ética de la IA es un campo multidisciplinario que estudia cómo optimizar el impacto beneficioso de la [inteligencia artificial (IA)](https://www.ibm.com/mx-es/think/topics/artificial-intelligence) mientras se reducen los riesgos y los resultados adversos.



What is AI Ethics? (6:10 min)

Algunos ejemplos de cuestiones éticas de la IA son la responsabilidad y la privacidad de los datos, la imparcialidad, la [explicabilidad](https://www.ibm.com/mx-es/think/topics/explainable-ai" \t "_self), la robustez, la [transparencia](https://www.ibm.com/mx-es/think/topics/ai-transparency), la sustentabilidad medioambiental, la inclusión, la agencia moral, la alineación de valores, la responsabilidad, la confianza y el mal uso de la tecnología. Este artículo tiene como objetivo proporcionar una visión integral del mercado de la ética de la IA en la industria actual. Para aprender más sobre el punto de vista de IBM, consulte nuestra página de ética de IA aquí.

Con la aparición de [big data](https://www.ibm.com/mx-es/think/topics/big-data" \t "_self), las compañías aumentaron su enfoque para impulsar la automatización y la toma de decisiones basada en datos en todas sus organizaciones. Si bien la intención suele ser, si no siempre, mejorar los resultados comerciales, las compañías están experimentando consecuencias imprevistas en algunas de sus aplicaciones de IA, particularmente debido a un diseño de investigación inicial deficiente y conjuntos de datos sesgados.

A medida que salieron a la luz casos de resultados injustos, surgieron nuevas directrices, principalmente de las comunidades de investigación y ciencia de datos, para abordar las preocupaciones en torno a la ética de la IA. Las compañías líderes en el campo de la IA también mostraron un interés personal en dar forma a estas directrices, ya que ellas mismas comenzaron a experimentar algunas de las consecuencias de no mantener los estándares éticos en sus productos. La falta de diligencia en esta área puede generar exposición reputacional, regulatoria y legal, lo que se deriva en sanciones costosas. Como ocurre con todos los avances tecnológicos, la innovación tiende a superar la regulación gubernamental en campos nuevos y emergentes. A medida que se desarrolle la experiencia adecuada dentro de la industria gubernamental, podemos esperar que las compañías sigan más protocolos de IA, lo que les permitirá evitar cualquier violación de los derechos humanos y las libertades civiles.

Establecimiento de principios para la ética de la IA

Mientras se desarrollan normas y protocolos para gestionar el uso de la IA, la comunidad académica ha aprovechado el [Informe Belmont](https://www.hhs.gov/ohrp/sites/default/files/the-belmont-report-508c_FINAL.pdf) como medio para orientar la ética dentro de la investigación experimental y el desarrollo algorítmico. Hay tres principios fundamentales que se desprenden del Informe Belmont y que sirven de guía para el diseño de experimentos y algoritmos, que son:

1. **Respeto a las personas:** este principio reconoce la autonomía de los individuos y defiende la expectativa de que los investigadores protejan a las personas con autonomía disminuida, lo que podría deberse a una variedad de circunstancias como enfermedad, discapacidad mental o restricciones de edad. Este principio se concentra principalmente en la idea del consentimiento. Las personas deben ser conscientes de los posibles riesgos y beneficios de cualquier experimento en el que participen, y deben poder elegir participar o retirarse en cualquier momento antes y durante el experimento.
2. **Beneficencia:** este principio proviene de la ética de la atención médica, donde los médicos hacen el juramento de “no hacer daño”. Esta idea se puede aplicar fácilmente a la inteligencia artificial, donde los algoritmos pueden amplificar los sesgos en torno a la raza, el género, las inclinaciones políticas, etc., a pesar de la intención de hacer el bien y mejorar un sistema determinado.
3. **Justicia:** este principio trata temas como la equidad y la igualdad. ¿Quién debería cosechar los beneficios de la experimentación y el machine learning? El Informe Belmont ofrece cinco formas de distribuir cargas y beneficios, que son:
   * Igualdad
   * Necesidad individual
   * Esfuerzo individual
   * Contribución social
   * Mérito



Las últimas novedades e insights sobre IA

Descubra insights y noticias de expertos sobre IA, la nube y mucho más en el boletín semanal Think.

[Suscríbase hoy](https://www.ibm.com/account/reg/signup?formid=news-urx-52954)

Principales preocupaciones de la IA en la actualidad

Hay una serie de cuestiones que están al frente de las conversaciones éticas en torno a las tecnologías de IA en el mundo real. Por ejemplo:

Modelos fundacionales e IA generativa

El lanzamiento de ChatGPT en 2022 marcó un verdadero punto de inflexión para la inteligencia artificial. Las capacidades del chatbot de OpenAI (desde redactar reportes legales hasta depurar códigos) abrieron una nueva constelación de posibilidades sobre lo que la IA puede hacer y cómo se puede aplicar en casi todas las industrias.

ChatGPT y herramientas similares se basan en modelos fundacionales, modelos de IA que pueden adaptarse a una amplia gama de tareas posteriores. Los modelos fundacionales suelen ser modelos generativos a gran escala, compuestos por miles de millones de parámetros, que se entrenan con datos no etiquetados mediante autosupervisión. Esto permite a los modelos fundacionales aplicar rápidamente lo que han aprendido en un contexto a otro, lo que los hace altamente adaptables y capaces de realizar una amplia variedad de tareas diferentes. Sin embargo, hay muchos posibles problemas y preocupaciones éticas en torno a los modelos fundacionales que son comúnmente reconocidos en la industria de la tecnología, como la parcialidad, la generación de contenido falso, la falta de explicabilidad, el mal uso y el impacto social. Muchos de estos problemas son relevantes para la IA en general, pero adquieren una nueva urgencia a la luz del poder y la disponibilidad de los modelos fundacionales.

Singularidad tecnológica

[La singularidad tecnológica](https://www.ibm.com/mx-es/think/topics/technological-singularity) es un escenario teórico en el que el crecimiento tecnológico se vuelve incontrolable e irreversible, culminando en cambios profundos e impredecibles en la civilización humana. Si bien este tema atrae mucha atención pública, a muchos investigadores no les preocupa la idea de que la IA supere a la inteligencia humana en un futuro cercano o inmediato.

[La IA sólida](https://www.ibm.com/mx-es/think/topics/strong-ai) (IA que poseería inteligencia y conciencia iguales a las de los humanos) y la superinteligencia siguen siendo hipotéticas, las ideas plantean algunas preguntas interesantes a la hora de considerar el uso de sistemas autónomos, como los coches autónomos. Es poco realista pensar que un coche sin conductor nunca tendrá un accidente de coche, pero ¿quién es el responsable en esas circunstancias? ¿Deberíamos seguir apostando por los vehículos autónomos, o limitamos la integración de esta tecnología para crear solo vehículos semiautónomos que promuevan la seguridad de los conductores? El debate aún no ha concluido, pero estos son los tipos de debates éticos que se producen a medida que se desarrolla la nueva e innovadora tecnología de IA.

Impacto de la IA en los empleos

Si bien gran parte de la percepción pública de la inteligencia artificial se centra en la pérdida de empleos, esta preocupación probablemente debería replantearse. Con cada nueva tecnología disruptiva, vemos que cambia la demanda de roles laborales específicos.

Por ejemplo, al considerar la industria automotriz, muchos fabricantes, como GM, están cambiando su enfoque hacia la producción de vehículos eléctricos para ajustarse a las iniciativas ecológicas. La industria energética no va a desaparecer, pero la fuente de energía está cambiando de una economía de combustible a una eléctrica.

La inteligencia artificial debe verse de manera similar, donde la inteligencia artificial desplazará la demanda de puestos de trabajo a otras áreas. Será necesario que haya personas que ayuden a administrar estos sistemas a medida que los datos crezcan y cambien todos los días. Seguirá siendo necesario contar con recursos para atender problemas más complejos en las industrias con más probabilidades de verse afectados por los cambios en la demanda de empleo, como el servicio de atención al cliente. Lo importante de la inteligencia artificial y su efecto en el mercado laboral será ayudar a las personas en la transición a estas nuevas áreas de demanda del mercado.

Privacidad

La privacidad tiende a discutirse en el contexto de la privacidad, protección y seguridad de los datos, y estas preocupaciones permitieron a los responsables políticos avanzar más en este sentido en los últimos años. Por ejemplo, en 2016, se creó la legislación GDPR para proteger los datos personales de las personas en la Unión Europea y el Espacio Económico Europeo, dando a los individuos más control sobre sus datos. En Estados Unidos, los estados individuales están desarrollando políticas, como la Ley de Privacidad del Consumidor de California (CCPA), que exige a las empresas que informen a los consumidores sobre la recopilación de sus datos.

Esta y otras leyes recientes han obligado a las empresas a repensar cómo almacenan y utilizan los datos de identificación personal (PII). Como resultado, las inversiones en seguridad se han convertido en una prioridad cada vez mayor para las empresas, que tratan de eliminar cualquier vulnerabilidad y oportunidad de vigilancia, piratería informática y ataques cibernéticos.

Prejuicios y discriminación

Los casos de sesgo y discriminación en una serie de sistemas de machine learning han planteado muchas cuestiones éticas sobre el uso de la inteligencia artificial. ¿Cómo podemos protegernos contra el sesgo y la discriminación cuando los conjuntos de datos de formación pueden prestarse al sesgo? Aunque las empresas suelen tener intenciones bien intencionadas en torno a sus esfuerzos de automatización, la incorporación de la IA en las prácticas de contratación puede tener consecuencias imprevistas. En su esfuerzo por automatizar y simplificar un proceso, Amazon [sesgó involuntariamente a posibles candidatos](https://incidentdatabase.ai/cite/37/) por género para puestos técnicos abiertos y, en última instancia, tuvo que desechar el proyecto. A medida que surgen eventos como estos, [Harvard Business Review](https://hbr.org/2019/04/the-legal-and-ethical-implications-of-using-ai-in-hiring) ha planteado otras preguntas puntuales sobre el uso de la IA en las prácticas de contratación, como qué datos *debería* poder utilizar al evaluar a un candidato para un puesto.

El sesgo y la discriminación tampoco se limitan a la función de recursos humanos; se pueden encontrar en una serie de aplicaciones, desde software de reconocimiento facial hasta algoritmos de redes sociales.

A medida que las empresas se vuelven más conscientes de los riesgos de la IA, también se vuelven más activas en la discusión sobre la ética y los valores de la IA. Por ejemplo, el año pasado, el CEO de IBM, Arvind Krishna, compartió que IBM ha puesto fin a sus productos de reconocimiento y análisis facial de IBM de propósito general, enfatizando que *“IBM se opone firmemente y no condona los usos de ninguna tecnología, incluida la de reconocimiento facial que ofrecen otros proveedores para vigilancia masiva, perfiles raciales, violaciones de derechos humanos y libertades básicas, o cualquier propósito que no sea consecuente con nuestros valores y principios de confianza y*transparencia”.

Responsabilidad

No existe una legislación universal y general que regule las prácticas de IA, pero muchos países y estados trabajan para desarrollarlas y aplicarlas a nivel local. En la actualidad existen algunas normas sobre IA, y muchas otras están a punto de entrar en vigor. Para llenar el vacío, han surgido marcos éticos como parte de una colaboración entre especialistas en ética e investigadores para regir la creación y distribución de modelos de IA dentro de la sociedad. Sin embargo, por el momento, estos solo sirven como guía, y la [investigación](https://arxiv.org/ftp/arxiv/papers/1903/1903.03425.pdf) muestra que la combinación de responsabilidad distribuida y la falta de previsión de las posibles consecuencias no conduce necesariamente a la prevención de daños a la sociedad.

**AI Academy**



Confianza, transparencia y gobernanza en IA

Es probable que la confianza en la IA sea el tema más importante en el campo de la IA. También es, comprensiblemente, un tema abrumador. Desentrañaremos cuestiones, como las alucinaciones, el sesgo y el riesgo, y compartiremos medidas para adoptar la IA de manera ética, responsable y justa.

[Ir al episodio](https://www.ibm.com/mx-es/think/videos/ai-academy/generative-ai-governance)

Cómo establecer la ética de la IA

La inteligencia artificial se desempeña de acuerdo con la forma en que se diseña, desarrolla, entrena, sintoniza y utiliza, y la ética de la IA trata de establecer un ecosistema de estándares éticos y límites alrededor de todas las fases del ciclo de vida de un sistema de IA.

Las organizaciones, los gobiernos y los investigadores, en conjunto, han comenzado a crear marcos para abordar las preocupaciones éticas actuales sobre la IA y dar forma al futuro del trabajo en este campo. Si bien cada día se incorporan más estructuras a estas pautas, existe cierto consenso en cuanto a incorporar lo siguiente:

Gobernanza

La gobernanza es el acto de una organización de supervisar el ciclo de vida de la IA a través de políticas y procesos internos, personal y sistemas. La gobernanza ayuda a garantizar que los sistemas de IA funcionen según lo previsto en los principios y valores de una organización, como esperan los stakeholders y según lo exija la normativa pertinente. Un programa de gobernanza exitoso:

* define las funciones y responsabilidades de las personas que trabajan con IA.
* educa a todas las personas involucradas en el ciclo de vida de la IA sobre cómo crear IA de manera responsable.
* establecerá procesos para construir, gestionar, monitorizar y comunicar sobre la IA y los riesgos de la IA.
* aprovecha las herramientas para mejorar el rendimiento y la confiabilidad de la IA a lo largo de su ciclo de vida.

Un comité de ética de IA es un mecanismo de gobernanza particularmente eficaz. En IBM, el Consejo de Ética de IA está compuesto por diversos líderes de toda la compañía. Proporciona un proceso centralizado de gobernanza, revisión y toma de decisiones para las políticas y prácticas éticas de IBM. [Obtenga más información sobre el Comité de Ética de IA de IBM](https://www.ibm.com/mx-es/think/insights/how-our-commitment-to-ethics-trust-and-transparency-is-differentiating-ibm).

Principios y áreas de interés

La estrategia de una organización en ética de AI puede guiarse por principios que se pueden aplicar a productos, políticas, procesos y prácticas en toda la organización para ayudar a habilitar una IA confiable. Estos principios deben estructurar y apoyar áreas de enfoque, como la explicabilidad o la equidad, en torno a las cuales se pueden desarrollar estándares y alinear prácticas.

Cuando la IA se construye con la ética en el centro, es capaz de tener un gran potencial para impactar a la sociedad positivamente. Empezamos a ver esto en su integración en áreas de la salud, como la radiología. La conversación en torno a la ética de la IA también es importante para evaluar y mitigar adecuadamente los posibles riesgos relacionados con los usos de la IA, comenzando la fase de diseño.

Organizaciones que promueven la ética de la IA

Dado que los estándares éticos no son *la* principal preocupación de los ingenieros y científicos de datos en el sector privado, surgieron varias organizaciones para promover la conducta ética en el campo de la inteligencia artificial. Para aquellos que buscan más información, las siguientes organizaciones y proyectos proporcionan recursos para promulgar la ética de la IA:

* **AlgorithmWatch**: esta organización sin fines de lucro se enfoca en un algoritmo explicable y rastreable, así como en un proceso de decisión en programas de IA. Haga clic [aquí](https://algorithmwatch.org/en/) para aprender más.
* **AI Now Institute:**esta organización sin ánimo de lucro de la Universidad de Nueva York investiga las implicaciones sociales de la inteligencia artificial. Haga clic [aquí](https://ainowinstitute.org/) para aprender más.
* **DARPA:** la [Defense Advanced Research Projects Agency](https://www.darpa.mil/research/programs/ai-next) del Departamento de Defensa de EE. UU. se centra en promover la IA explicable y la investigación sobre la IA.
* **CHAI:** el [Center for Human-Compatible Artificial Intelligence](https://humancompatible.ai/) es una cooperación de varios institutos y universidades para promover una IA confiable y sistemas beneficiosos demostrables.
* **NASCAI:** la [Comisión de Seguridad Nacional sobre Inteligencia Artificial](https://cybercemetery.unt.edu/nscai/20211005220330/https:/www.nscai.gov/) es una comisión independiente "que estudia los métodos y medios necesarios para avanzar en el desarrollo de la inteligencia artificial, el aprendizaje automático y las tecnologías asociadas para abordar de forma integral las necesidades de seguridad nacional y defensa de Estados Unidos".

El punto de vista de IBM sobre la ética de la IA

IBM también estableció su propio punto de vista sobre la ética de la IA, creando [Principios de confianza y transparencia](https://www.ibm.com/mx-es/policy/trust-transparency) para ayudar a los clientes a entender dónde se encuentran sus valores dentro de la conversación sobre la IA. IBM tiene tres principios básicos que dictan su enfoque hacia el desarrollo de datos e IA, que son:

1. **El objetivo de la IA es potenciar la inteligencia humana.** Esto significa que no buscamos reemplazar la inteligencia humana con IA, sino apoyarla. Dado que cada nueva innovación tecnológica implica cambios en la oferta y la demanda de roles laborales particulares, IBM se compromete a apoyar a los trabajadores en esta transición invirtiendo en iniciativas globales para promover la formación de habilidades en torno a esta tecnología.
2. **Los datos e insights pertenecen a su creador.** Los clientes de IBM pueden estar seguros de que ellos, y solo ellos, son propietarios de sus datos. IBM no proporcionó ni proporcionará acceso gubernamental a los datos de los clientes para ningún programa de vigilancia, y sigue comprometida con la protección de la privacidad de sus clientes.
3. **Los sistemas de IA deben ser transparentes y explicables** IBM cree que las empresas de tecnología deben tener claro quién entrena sus sistemas de IA, qué datos se utilizaron en el entrenamiento y, lo más importante, qué se incluyó en las recomendaciones de sus algoritmos.

IBM también ha desarrollado [cinco pilares](https://www.ibm.com/mx-es/policy/ibm-artificial-intelligence-pillars/) para guiar la adopción responsable de las tecnologías de IA. Estas incluyen:

* **Explicabilidad:** un sistema de IA debe ser transparente, particularmente sobre lo que se incluyó en las recomendaciones de su algoritmo, según sea relevante para los stakeholders con una variedad de objetivos.
* **Equidad:** se refiere al trato equitativo de las personas o grupos de personas por parte de un sistema de IA. Cuando se calibra adecuadamente, la IA puede ayudar a los humanos a tomar decisiones más justas, contrarrestar los sesgos humanos y promover la inclusión.
* **Fortaleza:** los sistemas impulsados por IA deben defenderse activamente de los ataques de adversarios, minimizando los riesgos de seguridad y permitiendo la confianza en los resultados del sistema.
* **Transparencia:** para reforzar la confianza, los usuarios deben poder ver cómo funciona el servicio, evaluar su funcionalidad y comprender sus fortalezas y limitaciones.
* **Privacidad:**Los sistemas de IA deben priorizar y salvaguardar la privacidad y los derechos de datos de los consumidores y proporcionar garantías explícitas a los usuarios sobre cómo se utilizarán y protegerán sus datos personales.

Estos principios y áreas de enfoque forman la base de nuestro enfoque de la ética de la IA. Para obtener más información sobre los puntos de vista de IBM en torno a la ética y la inteligencia artificial, lea más [aquí](https://www.ibm.com/mx-es/impact/ai-ethics).