

[Inicio](#) / [topics](#) / [Ética de la IA](#)

# ¿Qué es la ética de la IA?

La ética de la IA guía a los científicos e investigadores de datos para crear sistemas de IA de manera ética que beneficien a la sociedad en su conjunto.

**Suscríbase al boletín de IBM** →

## ¿Qué es la ética de la IA?

Establecimiento de principios para la ética de la IA

Principales inquietudes actuales de la IA

Cómo establecer la ética de la IA

Organizaciones de IA ética

Punto de vista de IBM en cuanto a la ética de la IA

Ética de la IA e IBM

Soluciones relacionadas

Dé el siguiente paso

# ¿Qué es la ética de la IA?

Este artículo tiene como objetivo proporcionar una visión integral del mercado de la ética de la IA en la industria actual. Para conocer más acerca del punto de vista de IBM, vea nuestra página de ética de la IA [aquí](#).

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 conjunto de pautas que asesoran sobre el diseño y los resultados de la inteligencia artificial. Los seres humanos tienen todo tipo de sesgos cognitivos, como los sesgos de actualidad y de confirmación, y esos sesgos inherentes se exhiben en nuestros comportamientos y, posteriormente, en nuestros datos. Dado que los datos son la base de todos los algoritmos de machine learning, es importante estructurar experimentos y algoritmos con esto en mente, ya que la inteligencia artificial tiene el potencial de amplificar y escalar estos sesgos humanos a un ritmo sin precedentes.

Con la aparición del big data, las empresas ahora se centran más en impulsar la automatización y la toma de decisiones basadas en datos en todas sus

A medida que se han descubierto casos de resultados injustos, han surgido nuevas pautas, principalmente de las comunidades de investigación y ciencia de datos, para abordar las preocupaciones en torno a la ética de la IA. Las empresas líderes en IA también se han interesado en estructurar estas pautas, ya que ellas mismas han comenzado a experimentar algunas de las consecuencias de no respetar los estándares éticos en sus productos. La falta de diligencia en esta área puede tener repercusiones legales y normativas, y dar lugar a costosas sanciones, además de afectar a la reputación. Como ocurre con todos los avances tecnológicos, la innovación tiende a superar la regulación gubernamental en campos nuevos y emergentes. A medida que la experiencia adecuada se vaya desarrollando dentro del sector gubernamental, podemos esperar que las empresas sigan más protocolos de inteligencia artificial, lo que les permitirá evitar cualquier violación a los derechos humanos y las libertades civiles.



## Conozca a watsonx

Aplique la potencia de la IA en su empresa con la plataforma de datos e IA de última generación de IBM

[Vea lo que sigue →](#)

## Cómo avanzar en la ética de la IA hoy

[Descubra cómo poner en práctica los principios éticos de la IA. \(326 KB\)](#)



# Establecimiento de principios para la ética de la IA

A medida que se desarrollan reglas y protocolos para gestionar el uso de IA, la comunidad académica se ha basado en el [Informe Belmont](#) (enlace externo a ibm.com) (PDF, 121 KB) para guiar la ética en la investigación experimental y el desarrollo algorítmico. Hay tres principios fundamentales que surgieron del Informe Belmont que sirven como guía para el diseño de experimentos y algoritmos:

1. **Respeto a las personas:** este principio reconoce la autonomía de las personas 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 una enfermedad, una discapacidad mental, restricciones de edad, etc. Este principio se enfoca principalmente en la idea del consentimiento. Las personas



deben ser conscientes de los posibles riesgos y beneficios de cualquier experimento del que participen y deben poder elegir participar o retirarse en cualquier momento antes y durante el experimento.

2. **Beneficencia:** este principio toma el ejemplo de la ética del cuidado de la salud, donde los médicos hacen un 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 aborda cuestiones como la imparcialidad y la igualdad. ¿Quién debería aprovechar los beneficios de la experimentación y el machine learning? El Informe Belmont ofrece cinco formas de distribuir cargas y beneficios:
  - Distribución equitativa
  - Necesidad individual
  - Esfuerzo individual
  - Contribución social
  - Mérito

## Enlaces relacionados

[Ética de la IA de IBM](#)



# Principales inquietudes actuales de la IA

Varias son las cuestiones recurrentes en las conversaciones éticas en torno a las tecnologías de IA. Entre ellas, destacamos:

## Singularidad tecnológica

Aunque este tema ofrece mucha atención al público, muchos investigadores no están preocupados por la idea de que la IA supere la inteligencia humana en un futuro cercano o inmediato. Esto también se conoce como superinteligencia, que Nick

Bostrum define como "cualquier intelecto que supera ampliamente a los mejores cerebros humanos en prácticamente todos los campos, incluida la creatividad científica, la sabiduría general y las habilidades sociales". A pesar de que la IA fuerte y la superinteligencia no son inminentes en la sociedad, su idea plantea algunas preguntas interesantes si consideramos el uso de sistemas autónomos, como los vehículos autónomos. No es realista pensar que un automóvil sin conductor nunca se metería en un accidente de coche, pero ¿quién es responsable bajo esas circunstancias? ¿Deberíamos ir tras vehículos autónomos, o limitarnos a la integración de esta tecnología para crear sólo vehículos semiautónomos que promuevan la seguridad entre los conductores? La pregunta aún no está resuelta, pero estos son los tipos de debates éticos que se están produciendo a medida que se desarrollan nuevas tecnologías innovadoras en el campo de la IA.

## Impacto de la IA en los puestos de trabajo

Aunque gran parte de la percepción pública de la inteligencia artificial se centra en la pérdida de empleos, esta preocupación probablemente debería reformularse. Con cada nueva tecnología disruptiva, vemos que la demanda del mercado para roles de trabajo específicos cambia. Por ejemplo, cuando vemos a la industria automotriz, muchos fabricantes, como GM, están cambiando para centrarse en la producción de vehículos eléctricos para alinearse con las iniciativas ecológicas. La industria energética no va a desaparecer, pero la fuente de energía está cambiando de una economía basada en el combustible a otra eléctrica. La inteligencia artificial debe ser vista de una manera similar, donde la inteligencia artificial cambiará la demanda de empleos a otras áreas. Se necesitarán personas para gestionar estos sistemas a medida que los datos crezcan y cambien cada día. Aún es necesario que haya recursos para abordar problemas más complejos dentro de las industrias que son más propensas a ser afectadas por turnos de demanda de trabajo, como el servicio al cliente. El importante aspecto de la inteligencia artificial y su efecto en el mercado de trabajo ayudará a las personas a pasar a estas nuevas áreas de demanda del mercado.

## Privacidad

La privacidad tiende a ser discutida en el contexto de la privacidad de los datos, la protección de datos y la seguridad de los datos, y estas preocupaciones han permitido a los legisladores hacer más avances en los últimos años. Por ejemplo, en 2016, la legislación de GDPR se creó para proteger los datos personales de las personas en la Unión Europea y el Espacio Económico Europeo, dando a las personas un mayor control de sus datos. En los Estados Unidos, los estados individuales están

desarrollando políticas, como la Ley de Privacidad del Consumidor de California (CCPA), que requieren que las empresas informen a los consumidores sobre la recopilación de sus datos. Esta reciente legislación ha obligado a las empresas a replantearse cómo almacenan y utilizan datos de identificación personal (PII). Como resultado, las inversiones dentro de la seguridad se han convertido en una prioridad cada vez mayor para las empresas, ya que buscan eliminar cualquier vulnerabilidad y oportunidad de vigilancia, hackeo y ciberataques.

## Sesgo y discriminación

Los casos de sesgo y discriminación en varios sistemas inteligentes han planteado muchas preguntas éticas sobre el uso de la inteligencia artificial. ¿Cómo podemos protegernos contra el sesgo y la discriminación cuando los propios datos de entrenamiento pueden estar sesgados? Si bien las empresas suelen tener buenas intenciones en cuanto a sus esfuerzos de automatización, [Reuters](#) (enlace externo a ibm.com) destaca algunas de las consecuencias imprevistas de incorporar la IA en las prácticas de contratación. En su esfuerzo por automatizar y simplificar un proceso, Amazon involuntariamente sesgó los posibles candidatos a un trabajo por género para puestos técnicos libres y, en última instancia, tuvieron que desechar el proyecto. A medida que surgen eventos como este, [Harvard Business Review](#) (enlace externo a ibm.com) ha planteado otros problemas en torno al uso de la IA en prácticas de contratación como, por ejemplo, qué datos *deberían* poder usarse 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 el reconocimiento facial hasta los algoritmos de redes sociales.

A medida que las empresas se vuelven más conscientes de los riesgos de la IA, también se han vuelto más activas en cuanto al debate en torno a su ética y valores. Por ejemplo, el año pasado, el CEO de IBM, Arvind Krishna, afirmó que IBM ha limitado sus productos de reconocimiento y análisis facial, enfatizando que *"IBM se opone firmemente y no tolera los usos de ninguna tecnología, incluida la tecnología de reconocimiento facial ofrecida por otros proveedores, para la vigilancia masiva, el establecimiento de perfiles raciales, las violaciones de los derechos humanos y las libertades básicas, o cualquier propósito que no sea consistente con nuestros valores y los Principios de confianza y transparencia"*.

## Responsabilidad

Dado que no hay una legislación significativa para regular las prácticas de IA, no

existe ningún mecanismo de ejecución real para garantizar prácticas éticas de IA. Los incentivos actuales para que las empresas se adhieran a estas directrices son las repercusiones negativas de un sistema de IA poco ético en el resultado final. Para cubrir esta carencia, han emergido marcos éticos como parte de una colaboración entre expertos en ética e investigadores para gestionar la construcción y distribución de modelos de IA dentro de la sociedad. No obstante, por ahora, solo sirven como guía, y un [estudio](#) (enlace externo a ibm.com) (PDF, 1 MB) muestra que la combinación de la responsabilidad distribuida y la falta de previsión de las posibles consecuencias no necesariamente conduce a impedir que la sociedad sufra daños.

## Cómo establecer la ética de la IA

Dado que la inteligencia artificial no genera máquinas morales, los equipos han comenzado a reunir marcos y conceptos para abordar algunas de las preocupaciones éticas actuales y determinar el futuro del trabajo dentro de este campo. Si bien cada día se añade más estructura a estas pautas, existe cierto consenso en torno a incorporar lo siguiente:

- **Gestión:** las empresas pueden aprovechar su estructura organizativa existente para ayudar a gestionar la IA ética. Si una empresa está recopilando datos, es probable que ya haya establecido un sistema de gestión para facilitar la estandarización de los datos y el aseguramiento de la calidad. Es probable que los equipos legales y regulatorios internos ya se estén asociando con equipos de gestión para garantizar la conformidad con entidades gubernamentales, por lo que ampliar el alcance de este equipo para incluir IA ética es una extensión natural de sus prioridades actuales. Este equipo también puede gestionar la conciencia organizacional e incentivar a las partes interesadas a actuar de acuerdo con los valores y estándares éticos de la empresa.
- **Explicabilidad:** los modelos de machine learning, en particular los modelos de deep learning, se denominan con frecuencia "modelos de caja negra", ya que generalmente no está claro cómo un modelo llega a una decisión determinada. Según este [estudio](#) (enlace externo a ibm.com) (PDF, 1.8 MB), la explicabilidad busca eliminar esta ambigüedad en torno a la creación del modelo y los resultados del modelo mediante la generación de una "explicación comprensible para los humanos que exprese la lógica de la máquina". Este tipo de transparencia es fundamental para generar confianza en los sistemas de IA y así garantizar que las personas comprendan por qué un modelo llega a una decisión determinada. Si





podemos comprender mejor el por qué, estaremos mejor equipados para evitar los riesgos de la IA, como el sesgo y la discriminación.

Sin duda, lograr una IA ética será importante para alcanzar el éxito. Sin embargo, es importante tener en cuenta que tiene un tremendo potencial para impactar a la sociedad para siempre. Comenzamos a ver esto en su integración en áreas del cuidado de la salud, como la radiología. Esta conversación sobre la ética de la IA es para garantizar que, en nuestro intento de aprovechar esta tecnología para hacer el bien, evaluemos adecuadamente su potencial de hacer daño dentro de su diseño.

## Organizaciones de IA ética

Dado que los estándares éticos no son *la* principal preocupación de los ingenieros de datos y los científicos de datos en el sector privado, ha surgido una serie de organizaciones para promover la conducta ética en el campo de la inteligencia artificial. Si está buscando más información, las siguientes organizaciones y proyectos brindan recursos para implementar una IA ética:

- **AlgorithmWatch:** esta organización sin fines de lucro se enfoca en un proceso de decisión y algoritmo explicable y rastreable en programas de IA. Haga clic [aquí](#) (enlace externo a ibm.com) para descubrir más.
- **AI Now Institute:** esta organización sin fines de lucro de la Universidad de Nueva York investiga las repercusiones sociales de la inteligencia artificial. Haga clic [aquí](#) (enlace externo a ibm.com) para descubrir más.
- **DARPA:** la [Agencia de Proyectos de Investigación Avanzados de Defensa](#) (enlace externo a ibm.com) del Departamento de Defensa de EE. UU. se centra en promover la IA explicable y la investigación de IA.
- **CHAI:** el [Centro de Inteligencia Artificial Compatible con Humanos](#) (enlace externo a ibm.com) es una cooperación de varios institutos y universidades para promover una IA confiable y sistemas beneficiosos demostrables.
- **NASCAI:** la [Comisión de Seguridad Nacional de Inteligencia Artificial](#) (enlace externo a ibm.com) es una comisión independiente "que considera los métodos y medios necesarios para avanzar en el desarrollo de inteligencia artificial, machine learning y tecnologías asociadas para abordar de manera integral las necesidades de defensa y seguridad nacional de los Estados Unidos".



# Punto de vista de IBM en cuanto a la ética de la IA

IBM también ha establecido su propio punto de vista en cuanto a la ética de la IA, creando principios de confianza y transparencia para ayudar a los clientes a descubrir sus valores en el debate de la IA. IBM tiene tres principios básicos que determinan su enfoque en materia de datos e IA:

1. **El propósito de la IA es aumentar la inteligencia humana.** Esto significa que no buscamos reemplazar la inteligencia humana con IA, sino apoyarla. Dado que cada innovación tecnológica implica cambios en la oferta y la demanda de puestos de trabajo específicos, IBM se compromete a apoyar a los trabajadores en esta transición invirtiendo en iniciativas globales para promover la formación profesional en torno a esta tecnología.
2. **Los datos e insights pertenecen a quien los creó.** Los clientes de IBM pueden estar seguros de que ellos, y solo ellos, son los propietarios de sus datos. IBM no ha proporcionado ni proporcionará acceso gubernamental a los datos de los clientes para ningún programa de vigilancia, y mantiene su compromiso con la protección de la privacidad de sus clientes.
3. **Los sistemas de IA deben ser transparentes y explicables.** IBM cree que las empresas tecnológicas deben ser claras en cuanto a quién entrena sus sistemas de IA, qué datos se utilizan en ese proceso y, lo más importante, qué se incluye en las recomendaciones de sus algoritmos.

IBM también ha desarrollado un conjunto de áreas de enfoque para guiar la adopción responsable de tecnologías de IA. Estas incluyen:

- **Explicabilidad:** un sistema de IA debe ser transparente, especialmente con respecto al contenido de las recomendaciones de su algoritmo, ya que es relevante para varios stakeholders con diferentes objetivos.
- **Equidad:** se refiere al trato justo 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 prejuicios humanos y promover la inclusión.
- **Robustez:** los sistemas basados en IA deben defenderse activamente de los ataques de adversarios para minimizar los riesgos de seguridad e impulsar 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 proteger la privacidad y los derechos de datos de los consumidores y brindar 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 ética de la IA. Para descubrir más acerca de las opiniones de IBM sobre ética e IA, haga clic [aquí](#).

## Ética de la IA e IBM

IBM busca garantizar que sus productos se desarrollen y utilicen teniendo en cuenta principios y directrices éticas. Uno de los productos que IBM ofrece a sus clientes es [IBM® Watson Studio](#), que mejora la supervisión y conformidad de estándares éticos de IA.

[IBM® Watson Studio](#) en [IBM® Cloud Pak for Data](#) ayuda a supervisar y gestionar modelos para utilizar una IA confiable. Una empresa puede visualizar y realizar un seguimiento de los modelos de IA en producción, validar y probar modelos para mitigar los riesgos regulatorios y aumentar la visibilidad del ciclo de vida de la IA. Regístrese para obtener un ID de IBM (IBMid) y [crear hoy su cuenta de IBM® Cloud](#) sin costo.

Para conocer más acerca del punto de vista de IBM acerca de la ética de la IA, haga clic [aquí](#).

## Soluciones relacionadas

---

## IBM® Watson Studio

Cree, ejecute y gestione modelos de IA. Prepare datos y cree modelos en cualquier nube mediante el uso de código fuente abierto o modelado visual. Prevea y optimice sus resultados.

[Explore Watson Studio](#) →

---

## IBM® Cloud Pak for Data

Cloud Pak for Data es una plataforma de datos abierta y ampliable que proporciona un tejido de datos (data fabric) para que todos los datos estén disponibles para la IA y analítica, en cualquier nube.

[Explore IBM® Cloud Pak for Data](#) →

---

## Ética de la IA de IBM

El enfoque multidisciplinario y multidimensional de la ética de la IA de IBM

[Explore la ética de la IA de IBM](#) →

# Dé el siguiente paso



IBM busca garantizar que sus productos se desarrollen y utilicen teniendo en cuenta principios y directrices éticas. Uno de los productos que IBM ofrece a sus clientes es IBM® Watson Studio, que mejora la supervisión y la conformidad de estándares éticos de IA. IBM® Watson Studio en IBM® Cloud Pak for Data ayuda a supervisar y gestionar modelos para utilizar una IA confiable. Una empresa puede visualizar y realizar un seguimiento de los modelos de IA en producción, validar y probar modelos para mitigar los riesgos regulatorios y aumentar la visibilidad del ciclo de vida de la IA.

**Descubra más acerca de IBM® Watson Studio**

