



REPORTE DE
TECNOLOGÍA:
**INTELIGENCIA
ARTIFICIAL**

Copyright © 2024 Banco Interamericano de Desarrollo (BID). Esta obra se encuentra sujeta a una licencia Creative Commons CC BY 3.0 IGO (<https://creativecommons.org/licenses/by/3.0/igo/legalcode>). Se deberá cumplir los términos y condiciones señalados en el enlace URL y otorgar el respectivo reconocimiento al BID.

En alcance a la sección 8 de la licencia indicada, cualquier mediación relacionada con disputas que surjan bajo esta licencia será llevada a cabo de conformidad con el Reglamento de Mediación de la OMPI. Cualquier disputa relacionada con el uso de las obras del BID que no pueda resolverse amistosamente se someterá a arbitraje de conformidad con las reglas de la Comisión de las Naciones Unidas para el Derecho Mercantil (CNUDMI). El uso del nombre del BID para cualquier fin distinto al reconocimiento respectivo y el uso del logotipo del BID, no están autorizados por esta licencia y requieren de un acuerdo de licencia adicional.

Note que el enlace URL incluye términos y condiciones que forman parte integral de esta licencia.

Las opiniones expresadas en esta obra son exclusivamente de los autores y no necesariamente reflejan el punto de vista del BID, de su Directorio Ejecutivo ni de los países que representa.



AUTORES

Lucía Latorre
Valentín Muro
Eduardo Rego

SUPERVISOR

Mariana Gutierrez

COLABORADORES

Ignacio Cerrato
José Daniel Zarate

TechLab

Los TechReports son una iniciativa del Laboratorio de Tecnologías Emergentes del departamento de TI del BID, conocido como TechLab, que se encarga de explorar, experimentar y difundir información sobre nuevas tecnologías para conocer su impacto en el Grupo BID y la región de ALC.



Agradecimientos: El equipo del BID desea agradecer a todas las personas que participaron en entrevistas y brindaron información clave para este documento.

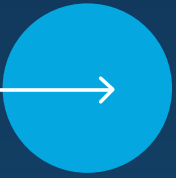


TABLA DE CONTENIDO

● RESUMEN EJECUTIVO	5
● DEFINICIÓN	6
● TÉCNICAS DE ENTRENAMIENTO Y APLICACIONES	7
Procesamiento de datos y aprendizaje automático	8
Procesamiento de Lenguaje Natural (NLP) y comunicaciones	10
Computer Vision	11
Inteligencia Artificial Generativa	12
● CASOS DE USO DE INTELIGENCIA ARTIFICIAL	14
Proyectos de interés	14
● RECOMENDACIONES	16
Para su desarrollo	16
Para el uso seguro y responsable	17
Sobre las consideraciones éticas	18
Sobre seguridad y privacidad	19
● REFERENCIAS	20



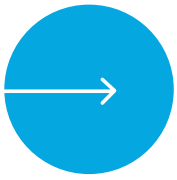
RESUMEN EJECUTIVO

La Inteligencia Artificial (IA) ha evolucionado significativamente desde su concepción, y actualmente sus aplicaciones son diversas, transformando industrias de manera profunda.

Este informe proporciona una visión integral de la IA, desde sus fundamentos hasta sus aplicaciones prácticas, abordando cuestiones como su definición, evolución e implementación. Además, se exploran distintas aplicaciones, como el aprendizaje automático, procesamiento de lenguaje natural, visión por computadora e IA generativa, con ejemplos y casos de uso específicos en sectores como salud, logística, medio ambiente y seguridad.

Resulta evidente que la expansión de la IA está cambiando cada vez más aspectos de la vida cotidiana y de diversas industrias. Sin embargo, su implementación responsable requiere una comprensión integral de sus principios éticos y técnicos, junto con una consideración constante de su impacto en la sociedad y la privacidad de los individuos. Finalmente, se ofrecen recomendaciones para su desarrollo responsable y se discuten consideraciones éticas como la privacidad y la transparencia.





DEFINICIÓN

La **inteligencia artificial (IA)** es un campo de estudio especialmente difícil de definir¹. Sin embargo, se lo puede definir como aquel que fusiona ciencias de la computación con robustos conjuntos de datos para facilitar la resolución de problemas², o bien, de manera más clásica como la capacidad de una computadora digital o un robot controlado por computadora para realizar tareas comúnmente asociadas con seres inteligentes³.

Luego del trabajo fundacional de Alan Turing en lo que bautizó “inteligencia mecánica”⁴, el término “inteligencia artificial” fue acuñado por el informático John McCarthy durante la Conferencia de Dartmouth⁵ en 1955, y definida como:

“la ciencia e ingeniería de crear máquinas inteligentes, especialmente programas de computadora inteligentes. Se relaciona con la tarea similar de utilizar computadoras para comprender la inteligencia humana, pero la IA no está limitada a métodos que sean biológicamente observables”⁶.



Con el paso de los años, la IA ha experimentado numerosos ciclos de sobreexpectación (hype cycle)⁷ en todos sus campos de estudio, pero con la aparición de nuevos modelos de acceso masivo de IA generativa a fines de 2022, como ChatGPT, muchas de sus discusiones recuperaron vigor⁸.

La IA comprende áreas como el aprendizaje automático (machine learning) y el aprendizaje profundo (*deep learning*); el reconocimiento de patrones; la toma de decisiones y la comprensión del lenguaje natural (NLP), entre muchas otras apuntan a la resolución de problemas de diversa naturaleza.

Por último, un objetivo central para algunas de las empresas principales de la disciplina es la creación de una Inteligencia Artificial General (AGI, en inglés), con proyectos en curso de OpenAI, DeepMind y Anthropic, entre otras. El cronograma para su desarrollo es objeto de un debate continuo⁹, con algunas opiniones que sugieren la posibilidad de lograrlo en años o décadas, mientras que otros estiman que podría llevar un siglo o más, e incluso hay quienes dudan de su viabilidad. Existe también controversia en torno a la definición precisa de AGI y si los modelos de lenguaje grandes (LLMs) modernos, como GPT-4, constituyen formas preliminares e incompletas de AGI. También es un tema recurrente en la ciencia ficción y en los estudios prospectivos sobre el futuro de la IA.

1. (Wang, 2019)

2. [What is Artificial Intelligence \(AI\) ? | IBM](#)

3. (Copeland, 2024)

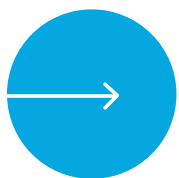
4. (Turing, 1950)

5. (McCarthy et al., 2006)

6. (McCarthy, 2007)

7. [Gartner Hype Cycle Research Methodology](#)

8. (Inter-American Development Bank, 2023; Mollick, 2022)



TÉCNICAS DE ENTRENAMIENTO Y APLICACIONES

Las soluciones IA son programas y herramientas de software que emplean técnicas de entrenamiento para llevar a cabo tareas específicas, desde simples y repetitivas hasta complejas funciones cognitivas que demandan capacidades comparables a las humanas.

PROCESAMIENTO DE DATOS Y APRENDIZAJE AUTOMÁTICO

El **aprendizaje automático o machine learning (ML)** se enfoca en utilizar información para imitar el proceso de aprendizaje humano, mejorando su precisión con cada iteración. Es esencial en la ciencia de datos y utiliza métodos estadísticos para entrenar algoritmos y realizar predicciones, así como para descubrir patrones en proyectos de minería de datos.¹⁰

El **aprendizaje profundo o deep learning (DL)** es un tipo de algoritmos de ML que consiste en redes neuronales artificiales (ANN, por sus siglas en inglés) con tres o más capas, que intentan imitar el funcionamiento del cerebro humano para “aprender” de grandes cantidades de datos.¹¹ Con frecuencia, los términos “machine learning” y “deep learning” se confunden, pero difieren en su enfoque de aprendizaje. Mientras que el DL puede emplear conjuntos de datos no estructurados, el ML suele requerir preprocesamiento de datos para aprovechar la estructura y etiquetado inicial de la información.¹²

Ambos modelos tienen su origen en el análisis de grandes volúmenes de datos, conocido como **big data**; estos conjuntos se caracterizan por su volumen, velocidad y variedad.¹³ El término refiere a colecciones de datos extremadamente grandes y diversas,¹⁴ que continúan creciendo exponencialmente con el tiempo, compuestas por datos estructurados, no estructurados y semiestructurados.¹⁵

El **aprendizaje por refuerzo (reinforcement learning)**, a diferencia del aprendizaje supervisado, no requiere ejemplos etiquetados ni la corrección explícita de sus errores. En cambio, se enfoca en equilibrar la exploración de nuevas posibilidades con el aprovechamiento del conocimiento existente que imita el proceso humano de prueba y error para conseguir objetivos o aprender.¹⁶

9. (McLean et al., 2023)

10. [What is Machine Learning? | IBM](#)

11. [What is Deep Learning? | IBM](#)

12. [What is Machine Learning? | IBM](#)

13. Big Data es otro concepto difícil de definir. (Favaretto et al., 2020)

14. [Big Data Defined: Examples and Benefits | Google Cloud](#)

15. Los datos estructurados son aquellos organizados en un formato predefinido y consistente, como tablas o bases de datos relacionales; los datos no estructurados carecen de un formato predefinido y no se pueden organizar fácilmente en una base de datos tradicional, como texto libre, imágenes o videos; y los datos semiestructurados tienen algún grado de estructura, pero no se ajustan completamente a un formato predefinido, como documentos XML o JSON.

16. <https://aws.amazon.com/what-is/reinforcement-learning/>

En cuanto a la **analítica predictiva (predictive analytics)** y la **minería de datos (data mining)** son disciplinas relacionadas que se enfocan en analizar grandes conjuntos de datos para extraer información valiosa y predecir eventos futuros. Mientras que la **analítica predictiva** utiliza técnicas estadísticas y de aprendizaje automático para hacer predicciones sobre tendencias y resultados futuros, la **minería de datos** se centra en descubrir patrones y relaciones ocultas en los datos para generar conocimientos accionables. Ambas disciplinas son fundamentales para la toma de decisiones informadas y la optimización de procesos en diversas industrias.

Estas aplicaciones ganan cada vez más presencia en diversas industrias como la salud, las finanzas, el comercio minorista y la manufactura. Entre ellas se encuentran algunos ejemplos:



Ejemplos y soluciones disponibles de aprendizaje automático

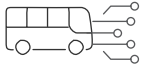


Asistencia en salud. Los avances recientes en el uso de ML en la industria de la salud han permitido aliviar la carga de los médicos y mejorar la precisión y calidad de la atención. Se han utilizado en roles de apoyo para identificar tendencias, desarrollar modelos de predicción de enfermedades y aumentar la eficiencia en instituciones médicas mediante la organización de registros electrónicos, detección de anomalías en muestras y cirugías asistidas por robots.¹⁷

17. (Habehh & Gohel, 2021)



Motores de recomendación. Los algoritmos de aprendizaje profundo pueden analizar enormes colecciones de datos de uso de las plataformas para extraer patrones y ofrecer recomendaciones personalizadas. Es el caso de los algoritmos de sugerencia de Netflix¹⁸ o de Spotify,¹⁹ las predicciones de tráfico de Google Maps²⁰ o las de clima en The Weather Channel.²¹



Conducción autónoma. Ciertos vehículos utilizan redes neuronales y ML para detectar obstáculos y tomar decisiones de conducción en tiempo real, basándose en datos de sensores como cámaras y radares. Por ejemplo, los vehículos de Waymo²², Tesla²³ y Aurora²⁴, que buscan lograr, eventualmente, la conducción completamente autónoma.



Detección de fraudes financieros. Mediante el análisis de grandes conjuntos de datos financieros, los algoritmos de análisis predictivo pueden identificar patrones sospechosos y prevenir posibles fraudes. Existen soluciones de IBM, Kount, FICO y Feedzai, entre otros²⁵.



Manufactura. La industria manufacturera ha sido lenta en la adopción de nuevas tecnologías, y la inteligencia artificial no es una excepción²⁶. Sin embargo, en consecuencia de los avances rápidos en algoritmos y más datos disponibles debido a sensores baratos y la fabricación inteligente, el uso de ML en la manufactura crecerá rápidamente, en aplicaciones como mantenimiento predictivo, gestión de almacenes y la mejora de calidad en procesos complejos como la fabricación de semiconductores²⁷.

PROCESAMIENTO DE LENGUAJE NATURAL (NLP) Y COMUNICACIONES

El **procesamiento del lenguaje natural (NLP)** es un campo que combina informática y lingüística, centrado en capacitar a las computadoras para comprender y trabajar con el lenguaje humano.

Esto implica analizar conjuntos de datos de texto o habla mediante métodos de aprendizaje automático basados en reglas o en probabilidades, para que las computadoras comprendan el contenido y contexto de los documentos. Sus principales desafíos incluyen el reconocimiento y comprensión del significado de los componentes lingüísticos que forma una frase u oración, así como la generación de lenguaje natural.

18. [Research Areas: Machine Learning](#)

19. [Machine Learning - Spotify Research](#)

20. [Google Maps I01: How AI helps predict traffic and determine routes](#)

21. [IBM's The Weather Company Continues to Be the World's Most Accurate Forecaster Overall](#)

22. [AutoML: Automating the design of machine learning models for autonomous driving](#)

23. [Autopilot | Tesla](#)

24. [Aurora Innovation](#)

25. (Abdallah et al., 2016)

26. (Stackpole, 2023)

27. (Wuest et al., 2016)



Ejemplos y soluciones disponibles de NLP



Asistentes virtuales. Aplicaciones diseñadas para interactuar de manera conversacional²⁹, ofreciendo asesoramiento personalizado. Es el caso de Siri (Apple), Assistant (Google) o Alexa (Amazon), pero también de chatbots personalizados configurados a medida con plataformas como Amazon Lex o Citibot³⁰.



Reconocimiento de voz. Permite que una computadora interprete y comprenda el lenguaje humano hablado. Se utiliza para convertir voz en texto y como entrada para los asistentes virtuales. Es utilizado por muchos asistentes virtuales, pero también como método de entrada mediante dictado en todos los sistemas operativos actuales, donde además cumple una importante función de accesibilidad.



Análisis de sentimientos en redes sociales. El análisis de datos de redes sociales puede utilizarse para predecir el sentimiento de los usuarios hacia una marca o producto específico, por ejemplo para evaluar el desempeño de cierta marca o producto a partir de los comentarios en redes sociales. Existen muchos proveedores de estos servicios, entre los que se destacan Brandwatch y Lexalytics, aunque existen también ofertas de IBM y Microsoft, entre otros.

28. (Chowdhary, 2020)

29. (Inter-American Development Bank, 2023)

30. [What is a Bot? - Types of Bots Explained - AWS](#)

COMPUTER VISION

La **visión artificial o visión por computadora (computer vision)** es la capacidad de interpretar y comprender de forma automática el mundo visual, utilizando imágenes, videos y modelos de aprendizaje profundo para identificar y clasificar objetos con precisión³¹.

Los motores comunes incluyen el reconocimiento óptico de caracteres (OCR), el reconocimiento inteligente de caracteres (ICR)³² y los más avanzados incorporan redes neuronales convolucionales (CNN)³³.



Ejemplos y soluciones disponibles de Computer Vision



Reconocimiento biométrico. Se utiliza para identificar y autenticar a individuos mediante características físicas únicas como el rostro, la voz o las huellas dactilares. Estos sistemas se emplean en seguridad como FaceID (Apple)³⁴ o sistemas de pago biométrico, pero también en redes sociales como el caso de los efectos o filtros de TikTok³⁵ o Instagram³⁶. Es una de las implementaciones de IA más controversiales de la actualidad³⁷, que ha llevado a varias iniciativas legislativas alrededor del mundo³⁸.

31. [What is Computer Vision? | IBM](#)

32. A diferencia del OCR, ICR puede reconocer caracteres manuscritos y no solo impresos.

33. [What is OCR? - Optical Character Recognition Explained - AWS](#)

34. [Usar Face ID en el iPhone o iPad Pro - Soporte técnico de Apple](#)

35. [Face Info | Learn | TikTok Effect House](#)

36. [Introducing Face Filters & More on Instagram | Instagram Blog](#)

37. (Chokshi, 2019)

38. (Bu, 2021)



Inspección visual automática. Utiliza tecnologías de visión artificial, ML y CNN para detectar defectos o anomalías, garantizando la calidad del proceso. El caso de Cognex³⁹ que se centra en soluciones industriales y de manufactura, o el de NVIDIA que abarca diversas áreas, como la aplicación médica con su plataforma Clara⁴⁰.

INTELIGENCIA ARTIFICIAL GENERATIVA

La **IA generativa (GenAI)**⁴¹ utiliza ML para crear nuevos contenidos a partir de un conjunto de datos de entrenamiento, pudiendo generar texto, imágenes, videos, audio, código o renderizaciones 3D, entre otros⁴². También puede ser útil para mejorar la interacción de empresas con clientes, consultar grandes volúmenes de información y automatizar tareas repetitivas⁴³.



39. [Cognex](#)

40. [NVIDIA Clara | AI-powered Solutions for Healthcare](#)

41. <https://publications.iadb.org/en/tech-report-generative-ai>

42. [Reporte IA Generativa BID](#)

43. [¿Qué es la IA generativa y cuáles son sus aplicaciones? Google Cloud](#)

Ejemplos y soluciones disponibles de IA Generativa



Generación de contenido. La generación de imágenes, texto, audio, vídeos, datos sintéticos y modelos 3D ofrece aplicaciones en campos como la publicidad y el márketing, pero también en atención médica, seguridad, edición y diseño. El campo ha ganado amplia atención debido a la popularidad de aplicaciones como ChatGPT de OpenAI⁴⁴, Copilot de Microsoft⁴⁵ y Gemini (ex Bard) de Google⁴⁶.

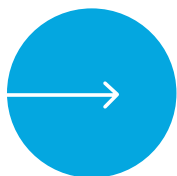


Diseño asistido. Herramientas de diseño, prototipado, creación de logotipos o generación de arte visual utilizan GenAI para ofrecer resultados personalizados que son evaluados antes de la producción. Por ejemplo, Autodesk utiliza IA en varias de sus soluciones⁴⁷ para optimizar el diseño y la simulación de prototipos, agilizando el proceso de fabricación, mientras que Adobe ha incorporado IA en muchos de sus productos⁴⁸.



Optimización de procesos. La GenAI resulta útil en la generación de código, agilizando la programación y detectando errores. Ejemplos incluyen GitHub Copilot⁴⁹ y Gemini⁵⁰ o Vertex AI de Google⁵¹, que sugieren código y ofrecen asistencia a desarrolladores.

46. [Gemini](#)
47. [Autodesk AI | Artificial Intelligence](#)
48. [Generative AI - Adobe Sensei](#)
49. [GitHub Copilot - Your AI pair programmer](#)
50. [Bard now helps you code](#)
51. [Generación de código con IA | Google Cloud](#)



CASOS DE USO DE INTELIGENCIA ARTIFICIAL

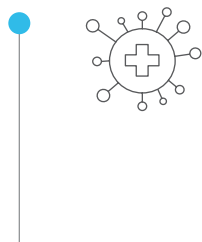
El 37% de las organizaciones ya han implementado IA de alguna manera, lo que representa un aumento del 270% en los últimos años⁵².

Este crecimiento es evidencia de su importancia, ya que se proyecta que los ingresos del mercado mundial de software de IA alcanzarán alrededor de USD 126 mil millones para 2025⁵³. Para ese mismo año, se espera que el 95% de las interacciones con los clientes sean impulsadas por IA⁵⁴.



PROYECTOS DE INTERÉS

En el desarrollo de la IA, se emplea un enfoque interdisciplinario que abarca áreas como matemáticas, ciencias de la computación, lingüística, psicología, entre otras.



Diagnóstico médico avanzado. En La IA se está utilizando para analizar radiografías, resonancias magnéticas o tomografías, y para ayudar a médicos en el diagnóstico de enfermedades. Por ejemplo, proyectos como IDx-DR⁵⁵ utilizan IA para detectar la retinopatía diabética en imágenes de retina, o Watson for Oncology⁵⁶ (IBM), un sistema de soporte de decisiones que ayuda a oncólogos a tomar decisiones informadas sobre el tratamiento del cáncer.

52. [Gartner Survey Shows 37 Percent of Organizations Have Implemented AI in Some Form](#)

53. [Global AI software market size 2018-2025 | Statista](#)

54. [What Makes Emerging Technologies The Future Of Customer Experience?](#)

55. [IDx-DR - Healthvisors](#)

56. [IBM Watson for Oncology](#)



Predicción de desastres naturales. Los sistemas de IA pueden analizar grandes cantidades de datos, como imágenes satelitales, datos climáticos e históricos para predecir y gestionar mejor los desastres naturales. Por ejemplo, el modelo GraphCast⁵⁷ (Google) y su proyecto “DeepMind for Flood Forecasting” para predecir inundaciones con mayor precisión⁵⁸.



Optimización logística. La IA se utiliza para optimizar rutas de transporte, gestionar inventarios y predecir la demanda en cadenas de suministro. Un ejemplo es el uso de algoritmos para optimizar la entrega de paquetes y reducir los tiempos de espera en fechas críticas en Amazon⁵⁹, junto a la optimización de su modelo AWS Stock Depletion Engine para prevenir el desperdicio de productos perecederos⁶⁰.



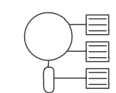
Conservación del medio ambiente. La IA se utiliza en proyectos de conservación para monitorear y proteger el medio ambiente y la vida silvestre. Por ejemplo, los modelos que se utilizan para rastrear y proteger más de 50 especies amenazadas o en peligro de extinción en el proyecto Wild Me⁶¹.



Seguridad y vigilancia. Los sistemas de IA se utilizan en sistemas de seguridad para analizar comportamientos sospechosos en entornos públicos y privados. Por ejemplo, la tecnología de Pangiam junto a la Administración de Seguridad del Transporte de los Estados Unidos (TSA) se utilizó para identificar posibles artículos prohibidos en el equipaje de mano⁶², como la incorporación del reconocimiento facial TSA PreCheck⁶³ para la verificación de identidad de los pasajeros.



Creación de contenido multimedia. La IA se utiliza en la creación de contenido multimedia, como música, arte y diseño gráfico. Por ejemplo, el modelo de AIVA utiliza IA para componer música original para clientes como Globant y Vodafone⁶⁴.



Investigación de nuevos materiales. La IA se utiliza en la investigación y desarrollo de materiales. Por ejemplo, utilizando ML para simular y predecir las propiedades de nuevos materiales, como metales más ligeros y más resistentes, es el caso del proyecto Material Genome Initiative y el descubrimiento de nuevos metales polares y el diseño de diodos orgánicos emisores de luz⁶⁵.

57. [GraphCast: AI model for faster and more accurate global weather forecasting - Google DeepMind](#)

58. [The Technology Behind our Recent Improvements in Flood Forecasting](#)

59. [5 ways Amazon is using AI to improve your holiday shopping and deliver your package faster](#)

60. [AWS Stock Depletion Engine to prevent waste of perishable products | Amazon Supply Chain and Logistics](#)

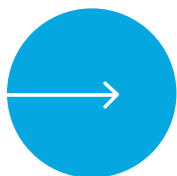
61. Wild Me

62. [TSA Trials Pangiam's AI-Driven Solution To Improve Airport Threat Detection](#)

63. [TSA PreCheck®: Touchless Identity Solution | Transportation Security Administration](#)

64. [AIVA](#)

65. (de Pablo et al., 2019)



RECOMENDACIONES



La IA ha emergido como una herramienta poderosa que impulsa la innovación en diversos campos, desde la atención médica hasta la gestión empresarial. Sin embargo, su implementación conlleva una serie de responsabilidades que deben abordarse para garantizar su uso seguro y responsable. Esto requiere un enfoque integral que aborde tanto los aspectos técnicos como éticos de su desarrollo y aplicación.

Garantizar la minimización o ausencia de consecuencias perjudiciales o no deseadas durante el desarrollo de proyectos de inteligencia artificial requiere una comprensión exhaustiva del papel de los principios de responsabilidad durante todas las etapas: diseño, implementación y mantenimiento de las aplicaciones de IA⁶⁶.

La **Inteligencia Artificial Responsable** (RAI por sus siglas en inglés) podría ser definida como la práctica de diseñar, desarrollar e implementar sistemas de IA con el propósito de empoderar tanto a empleados como a empresas, mientras que impacta de manera equitativa en clientes y en la sociedad en general. Esta práctica permite a las compañías generar confianza y expandir el uso de la IA con seguridad y certeza⁶⁷.

PARA SU DESARROLLO

Para el desarrollo responsable de la IA, resulta necesario considerar aspectos como la eficacia, robustez, y escalabilidad del modelo al evaluar un sistema de IA, estableciendo controles adecuados para evitar cuestiones como fugas de información, sobreajuste y subajuste iterando la mejora continua y así fomentar el desarrollo seguro y responsable de la IA como una herramienta beneficiosa.

66. [Here's why organizations should commit to responsible AI | World Economic Forum](#)

67. [Responsible AI | AI Ethics & Governance | Accenture](#)



Infraestructura escalable. Garantizar la capacidad de procesamiento mediante el uso de infraestructuras flexibles y actualizables. Empresas como Netflix utilizan infraestructuras en la nube, como Amazon Web Services (AWS), para ejecutar algoritmos de IA que personalizan recomendaciones de contenido a millones de usuarios adaptándose a las fluctuaciones en la demanda y garantizando el rendimiento incluso en momentos de alto tráfico⁶⁹.



Gobernanza. Establecer políticas y procedimientos para garantizar la integridad de los datos utilizados en los modelos de IA simplificando el control de accesos y aumentando la transparencia. La herramienta de Amazon SageMaker mejora el control y la visibilidad sobre los modelos de ML, permitiendo capturar y compartir información, manteniéndose informado sobre comportamientos como sesgos⁷⁰.



Validación y verificación. Antes de implementar un sistema de IA en un entorno de producción, es necesario realizar pruebas exhaustivas para validar su funcionamiento y verificar su precisión y confiabilidad. Por ejemplo, ReLM es una interfaz de prueba consultable para LLMs, permitiendo explicar y medir el comportamiento directamente sobre grandes conjuntos de datos⁷¹.

PARA EL USO SEGURO Y RESPONSABLE

Es esencial que las organizaciones implementen la IA de manera responsable para minimizar los riesgos asociados. La auditoría de IA implica evaluar, mitigar y asegurar la seguridad, legalidad y ética de un algoritmo, identificando riesgos tanto técnicos como de gobernanza y recomendando medidas para mitigarlos⁷².

Al evaluar un sistema de IA, es crucial considerar la eficacia, robustez, sesgo, explicabilidad y privacidad. Esto implica establecer controles adecuados para la gestión de datos sensibles y crear protocolos de seguridad para proteger la integridad de los sistemas contra ataques cibernéticos. Además, seguir buenas prácticas que promuevan la transparencia y la equidad, colocando a las personas y sus objetivos en el centro del diseño de sistemas, y respetando valores como la equidad, confiabilidad y transparencia⁷³.



Evaluaciones de riesgos. Realizar evaluaciones de riesgos periódicas para identificar posibles amenazas a la seguridad de los sistemas de IA y tomar medidas preventivas. Por ejemplo, **Counterfit**, una herramienta de código abierto de Microsoft, facilita a las organizaciones realizar evaluaciones de riesgos de seguridad en sus sistemas de IA. Su objetivo es garantizar la robustez, fiabilidad y confianza de los algoritmos utilizados en los negocios⁷⁴.

68. [Uso responsable de la IA para las políticas públicas: manual de ciencia de datos](#)

69. [Netflix Case Study](#)

70. [ML Governance with Amazon SageMaker](#)

71. (Kuchnik et al., 2023)

72. [Here's why organizations should commit to responsible AI | World Economic Forum](#)

73. [Microsoft Responsible AI Standard, v2](#)

74. [AI security risk assessment using Counterfit | Microsoft Security Blog](#)



Abordaje de sesgos: Es esencial para garantizar la equidad y precisión en los sistemas automatizados. Se deben evaluar periódicamente los datos de entrenamiento, identificar patrones discriminatorios y aplicar técnicas de corrección adecuadas⁷⁵. Por ejemplo, **Google Translate** busca reducir sesgos de género en traducciones, utilizando un enfoque de tres pasos: identificar consultas ambiguas, generar traducciones específicas de género y verificar su calidad antes de mostrarlas⁷⁶.



Mecanismos de explicabilidad: Establecer estos mecanismos sirven para comprender cómo funcionan los sistemas de IA y cómo se toman las decisiones⁷⁷. Por ejemplo, Microsoft muestra en su compendio de documentación de **Azure OpenAI** los detalles sobre cómo se procesan, usan y almacenan los datos que proporciona el usuario en el servicio⁷⁸.

SOBRE LAS CONSIDERACIONES ÉTICAS

Los sistemas de IA generan dilemas éticos que impactan áreas como la toma de decisiones, el empleo, la atención médica, la educación y más. Estos sistemas pueden exacerbar sesgos existentes y desafiar las concepciones sobre las capacidades humanas⁷⁹.

A largo plazo, podrían incluso cuestionar la singularidad y la autonomía humanas, planteando interrogantes sobre el autoconocimiento, la interacción social y el valor humano⁸⁰.

Los sistemas de IA deben empoderar a los seres humanos, garantizar la seguridad, respetar la privacidad, ser transparentes, evitar el sesgo, beneficiar a la sociedad y al medio ambiente, y ser auditables para garantizar la responsabilidad y proporcionar reparación cuando sea necesario⁸¹.



Respeto a las personas. Reconocer la autonomía individual y exigir que se protejan a aquellos con autonomía limitada. Es importante el consentimiento informado, asegurando que los usuarios estén plenamente conscientes de los riesgos y beneficios y tengan la libertad de retirarse en cualquier momento⁸².



Responsabilidades. Los desarrolladores y usuarios de sistemas de inteligencia artificial deben asumir la responsabilidad de las decisiones y acciones tomadas por estos sistemas. Se deben establecer mecanismos claros de rendición de cuentas para garantizar que se puedan identificar y corregir posibles errores o mal uso⁸³.

75. [Fairness](#)

76. [Providing Gender-Specific Translations in Google Translate](#)

77. [Explainability](#)

78. [Data, privacy, and security for Azure OpenAI Service](#)

79. (UNESCO, 2022)

80. Ídem.

81. [Directrices éticas para una IA fiable | Configurar el futuro digital de Europa](#)

82. [User data rights](#)

83. [Accountability](#)



Evaluación de impacto. Evaluar sistemáticamente el impacto ético de los sistemas de IA en términos de equidad, justicia y respeto a los derechos individuales y comunitarios desde un enfoque integral: desde fuerza laboral inclusiva hasta evaluación de sesgos y monitoreo continuo⁸⁴. **Fujitsu** propone una guía junto a herramienta que ayuda a identificar riesgos éticos, facilitando el proceso de evaluación de riesgos y haciéndolo más eficiente⁸⁵.

SOBRE SEGURIDAD Y PRIVACIDAD

La privacidad suele debatirse en el contexto de la privacidad de datos, la protección de datos y la seguridad de los mismos, y estas preocupaciones han permitido a los responsables de formular políticas avanzar más en este ámbito en los últimos años⁸⁶.



Protección de datos sensibles. Los modelos de ML pueden usar datos sensibles (personales, financieros, médicos u otros) para hacer predicciones, lo que requiere cumplir con normas legales, regulatorias y sociales para garantizar la privacidad de las personas⁸⁷.



Seguridad del modelo. Esto implica asegurar que los sistemas se comporten según lo previsto (estabilidad) y abordar los desafíos como la imprevisibilidad de escenarios y el equilibrio de restricciones vs. flexibilidad. La investigación de seguridad en ML se centra en amenazas como el envenenamiento de datos (*data poisoning*), la recuperación de datos sensibles, el robo de modelos y ejemplos adversariales⁸⁸.



Cifrado y Anonimización. Implementar medidas como enmascaramiento, hashing y encriptación es crucial para proteger la privacidad del usuario y evitar la divulgación no autorizada⁸⁹. Apple utiliza técnicas como el **Secure Enclave** y un ID aleatorio en Siri para garantizar la privacidad y seguridad de las interacciones del usuario con el dispositivo⁹⁰.

84. [Responsible AI Practices](#)

85. [AI Ethics : Fujitsu Global](#)

86. [AI Ethics | IBM](#)

87. [Responsible AI Practices](#)

88. [Responsible AI Practices](#)

89. [Personal Identifiable Information \(PII\) Anonymization](#)

90. [Privacy - Features - Apple](#)



REFERENCIAS

- 1** Abdallah, A., Maarof, M. A., & Zainal, A. (2016). Fraud detection system: A survey. *Journal of Network and Computer Applications*, 68, 90–113. <https://doi.org/10.1016/j.jnca.2016.04.007>
- 2** Bu, Q. (2021). The global governance on automated facial recognition (AFR): Ethical and legal opportunities and privacy challenges. *International Cybersecurity Law Review*, 2(1), 113–145. <https://doi.org/10.1365/s43439-021-00022-x>
- 3** Chokshi, N. (2019, mayo 15). Facial Recognition's Many Controversies, From Stadium Surveillance to Racist Software. *The New York Times*. <https://www.nytimes.com/2019/05/15/business/facial-recognition-software-controversy.html>
- 4** Chowdhary, K. R. (2020). Natural Language Processing. En K. R. Chowdhary (Ed.), *Fundamentals of Artificial Intelligence* (pp. 603–649). Springer India. https://doi.org/10.1007/978-81-322-3972-7_19
- 5** Copeland, B. J. (2024). artificial intelligence. *Encyclopedia Britannica*. <https://www.britannica.com/technology/artificial-intelligence>
- 6** de Pablo, J. J., Jackson, N. E., Webb, M. A., Chen, L.-Q., Moore, J. E., Morgan, D., Jacobs, R., Pollock, T., Schlom, D. G., Toberer, E. S., Analytis, J., Dabo, I., DeLongchamp, D. M., Fiete, G. A., Grason, G. M., Hautier, G., Mo, Y., Rajan, K., Reed, E. J., ... Zhao, J.-C. (2019). New frontiers for the materials genome initiative. *Npj Computational Materials*, 5(1), Article 1. <https://doi.org/10.1038/s41524-019-0173-4>
- 7** Favaretto, M., Clercq, E. D., Schneble, C. O., & Elger, B. S. (2020). What is your definition of Big Data? Researchers' understanding of the phenomenon of the decade. *PLOS ONE*, 15(2), e0228987. <https://doi.org/10.1371/journal.pone.0228987>
- 8** Habehh, H., & Gohel, S. (2021). Machine Learning in Healthcare. *Current Genomics*, 22(4), 291–300. <https://doi.org/10.17713/cgen.2021.22040291>
- 9** Inter-American Development Bank. (2023). Tech Report Generative AI. Inter-American Development Bank. <https://doi.org/10.18235/0005105>
- 10** Kuchnik, M., Smith, V., & Amvrosiadis, G. (2023). Validating Large Language Models with ReLM (arXiv:2211.15458). arXiv. <https://doi.org/10.48550/arXiv.2211.15458>
- 11** McCarthy, J. (2007). WHAT IS ARTIFICIAL INTELLIGENCE? <https://www-formal.stanford.edu/jmc/whatisai.pdf>

- 12** McCarthy, J., Minsky, M. L., Rochester, N., & Shannon, C. E. (2006). A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence, August 31, 1955. *AI Magazine*, 27(4), Article 4. <https://doi.org/10.1609/aimag.v27i4.1904>
- 13** McLean, S., Read, G. J. M., Thompson, J., Baber, C., Stanton, N. A., & Salmon, P. M. (2023). The risks associated with Artificial General Intelligence: A systematic review. *Journal of Experimental & Theoretical Artificial Intelligence*, 35(5), 649–663. <https://doi.org/10.1080/0952813X.2021.1964003>
- 14** Mollick, E. (2022, diciembre 14). ChatGPT Is a Tipping Point for AI. *Harvard Business Review*. <https://hbr.org/2022/12/chatgpt-is-a-tipping-point-for-ai>
- 15** Stackpole, B. (2023, junio 28). For AI in manufacturing, start with data | MIT Sloan. Ideas Made to Matter | MIT Sloan. <https://mitsloan.mit.edu/ideas-made-to-matter/ai-manufacturing-start-data>
- 16** Turing, A. M. (1950). I.—COMPUTING MACHINERY AND INTELLIGENCE. *Mind*, LIX(236), 433–460. <https://doi.org/10.1093/mind/LIX.236.433>
- 17** UNESCO. (2022). Recommendation on the Ethics of Artificial Intelligence. UNESCO. <https://unesdoc.unesco.org/ark:/48223/pf0000381137>
- 18** Wang, P. (2019). On Defining Artificial Intelligence. *Journal of Artificial General Intelligence*, 10(2), 1–37. <https://doi.org/10.2478/jagi-2019-0002>
- 19** Wuest, T., Weimer, D., Irgens, C., & Thoben, K.-D. (2016). Machine learning in manufacturing: Advantages, challenges, and applications. *Production & Manufacturing Research*, 4(1), 23–45. <https://doi.org/10.1080/21693277.2016.1192517>



 **BID**
Mejorando vidas