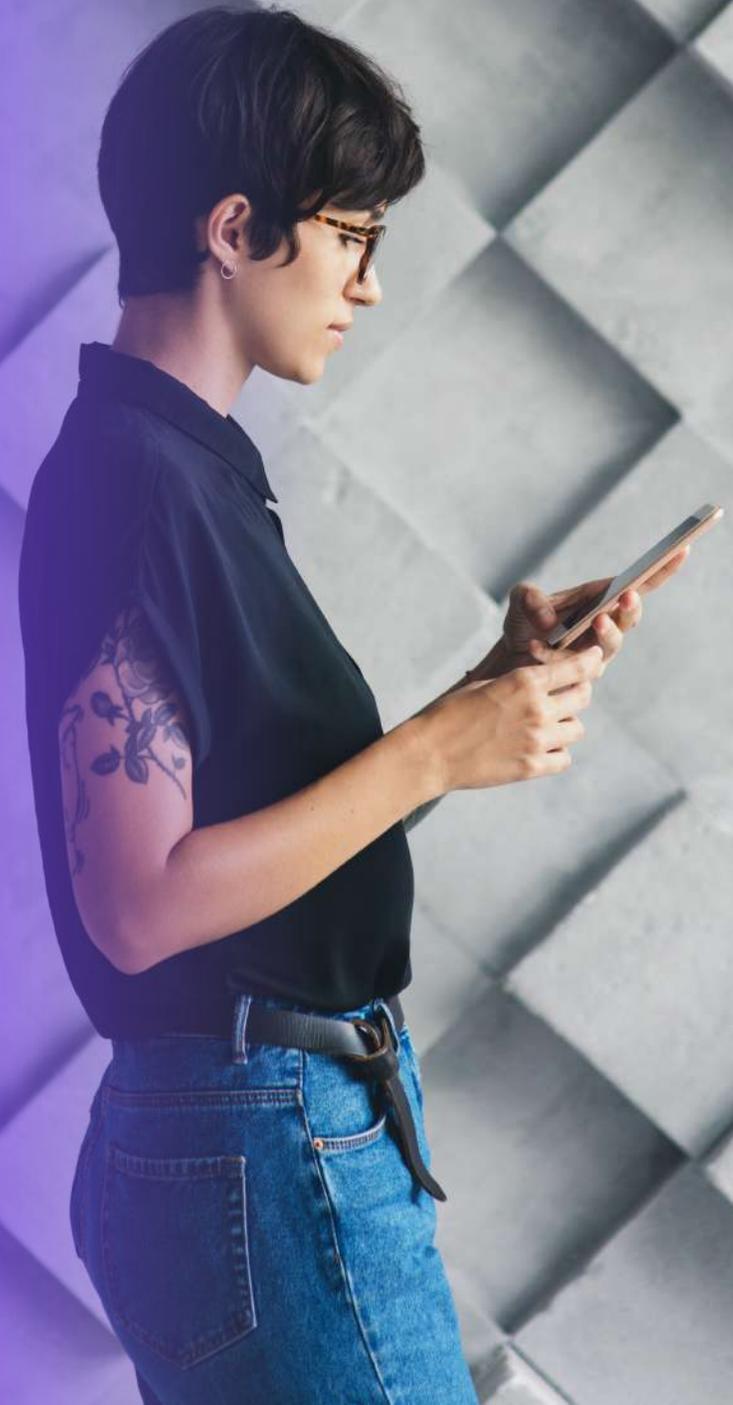




Democratización, operacionalización y fiabilidad: las tres claves del éxito de la IA

Liberar el potencial del machine learning en la era de la IA generativa



Índice

PARTE I - EL MERCADO

Introducción	4
Resumen	5

PARTE II - LAS CLAVES DEL ÉXITO

Clave n.º 1: Democratizar.....	7
Clave n.º 2: Operacionalizar	9
Clave n.º 3: Generar confianza	11
Los resultados satisfactorios comienzan aquí	15

PARTE I - EL MERCADO

El rápido desarrollo de la IA generativa está impulsando inversiones e innovaciones en IA cada vez mayores

INTRODUCCIÓN

Por qué la inteligencia artificial importa más que nunca

La inteligencia artificial (IA) existe desde hace décadas, pero con la gran disponibilidad de capacidad de computación escalable, la proliferación masiva de datos y el rápido avance de las tecnologías de machine learning (ML), las organizaciones de todos los sectores transforman sus negocios.

Con aplicaciones pensadas para el consumidor, como ChatGPT, que ofrecen resultados impresionantes y demuestran la potencia y sofisticación de los modelos de ML actuales de forma atractiva y clara, la **IA generativa** ha captado la atención y la curiosidad de todo el mundo. Nos encontramos en un verdadero punto de inflexión, y consideramos que la mayoría de las experiencias y aplicaciones de los clientes se reinventarán con la IA generativa.

Aunque las cifras específicas sobre el crecimiento de la IA generativa varían, el impacto de esta tecnología en la economía mundial no está en debate. Los efectos potenciales de estas cifras son asombrosos, tanto para las organizaciones que realizan las inversiones como para las que no. Según Goldman Sachs, la IA generativa podría impulsar un aumento del 7 % (o casi 7 billones de USD) del PIB mundial y estimular el crecimiento de la productividad en 1,5 puntos porcentuales en un periodo de 10 años. ¹ Además, Bloomberg prevé que el gasto en IA generativa alcanzará más de 109 000 millones de USD en 2030, una tasa de crecimiento anual compuesto (CAGR) del 34,6 % entre 2022 y 2030. ²

Ya sea que esté ampliando sus iniciativas de IA con herramientas generativas o que recién comience con la IA y el ML, una guía de referencia clara puede ayudarlo a desarrollar su estrategia y ayudar a garantizar resultados exitosos. Este libro electrónico expone los tres pilares estratégicos del éxito y ofrece recomendaciones prácticas que puede aplicar en su organización.



¹ "Generative AI could raise global GDP by 7%", Goldman Sachs, abril de 2023

² "Generative AI Market to be Worth \$109.37 Billion by 2030: Grand View Research, Inc", Bloomberg, enero de 2023



¿Qué es la IA generativa?

La IA generativa es un tipo de IA que puede crear nuevos contenidos e ideas, como conversaciones, historias, imágenes, videos y música. Como toda IA, la IA generativa está basada en modelos de ML, modelos muy grandes que se entrenan previamente a partir de grandes cantidades de datos, conocidos como modelos de base (FM). Los recientes avances en ML han permitido la aparición de modelos que contienen miles de millones de parámetros o variables.

Para dar una idea del cambio de escala, el mayor modelo entrenado previamente en 2019 fue de 330 millones de parámetros. Ahora, los modelos más grandes tienen más de 500 000 millones de parámetros, lo que supone un aumento de 1600 veces su tamaño en pocos años. Los FM actuales pueden realizar una amplia gama de tareas que abarcan múltiples dominios, como escribir entradas de blog, generar imágenes, resolver problemas matemáticos, entablar diálogos y responder a preguntas basadas en un documento. El tamaño y la naturaleza de uso general de los FM los diferencian de los modelos de ML tradicionales, que suelen realizar tareas específicas como analizar textos en busca de opiniones, clasificar imágenes y predecir tendencias.

Los FM pueden realizar muchas más tareas porque contienen un gran número de parámetros que les permiten aprender conceptos complejos. Además, gracias a su exposición previa a los datos a escala de Internet en todas sus formas y a sus innumerables patrones, los FM aprenden a aplicar sus conocimientos en una amplia gama de contextos. Aunque las capacidades y las posibilidades derivadas de los FM entrenados previamente son impresionantes, las organizaciones valoran aún más el hecho de que estos modelos generalmente eficaces también se pueden personalizar para realizar funciones específicas del dominio que son determinantes para sus negocios y que, sin embargo, solo utilizan una pequeña fracción de los datos y los cálculos necesarios para entrenar un modelo desde cero.

Los FM personalizados pueden crear una experiencia única para el cliente, personificando la voz, el estilo y los servicios de la empresa en una amplia variedad de sectores de consumo, como los servicios bancarios, el turismo y los servicios de salud. Por ejemplo, una empresa financiera que necesite generar automáticamente un informe diario de actividad para su difusión interna mediante todas las transacciones pertinentes puede personalizar el modelo con datos privados, que incluirán informes anteriores, de modo que el FM aprenda cómo deben leerse estos informes y qué datos se utilizaron para generarlos.

PARTE II - LAS CLAVES DEL ÉXITO

Tres claves que pueden ayudar a las organizaciones a llevar la IA a más partes de su negocio

CLAVE N.º 1

Democratizar

Para seguir el ritmo de las continuas innovaciones en IA, es importante que las aplicaciones de IA generativa sean sencillas y prácticas para todos. Esto se conoce como *democratización* y, para muchos clientes de Amazon Web Services (AWS), es el primer paso para aprovechar el potencial de estas potentes tecnologías. Las organizaciones buscan en AWS formas sencillas de encontrar y acceder a modelos de base (FM) de alto rendimiento que ofrezcan resultados extraordinarios y sean los más adecuados para diferentes tareas empresariales. El segundo paso se centra en garantizar la perfecta integración de los FM en las aplicaciones sin tener que administrar enormes clústeres de infraestructuras ni incurrir en costos significativos. El tercer y último paso simplifica el proceso de creación de aplicaciones diferenciadas sobre el FM base utilizando los datos de cada organización (pocos o muchos datos) y manteniendo al mismo tiempo la seguridad de esos datos.

La IA democratizada automatiza más las aplicaciones que usamos para vivir, trabajar y jugar, lo que permite más tiempo para concentrarse en actividades más valiosas.



Cientes como Intuit, Thomson Reuters, AstraZeneca, Ferrari, Bundesliga, 3M y BMW, así como miles de startups y organismos públicos de todo el mundo, transforman sus empresas, sectores y misiones gracias a la IA.



Un camino claro hacia la inteligencia artificial democratizada

AWS ha desempeñado un papel clave en la democratización de la IA y en hacerla accesible a cualquiera que quiera utilizarla, incluidos más de 100 000 clientes de todos los tamaños y sectores. Tenemos la cartera más amplia y detallada de servicios de IA y ML. AWS ha invertido e innovado para ofrecer la infraestructura más escalable y de mayor rendimiento para el entrenamiento y la inferencia de ML rentables; ha desarrollado **Amazon SageMaker**, la forma más sencilla para todos los desarrolladores de crear, entrenar e implementar modelos; y ha lanzado una amplia gama de **servicios de IA** que permiten a los clientes agregar capacidades de IA, como el reconocimiento de imágenes, la previsión o la búsqueda inteligente a las aplicaciones con una simple llamada de API.

Adoptamos el mismo enfoque que consiste en democratizar la IA generativa: trabajamos para sacar estas tecnologías del ámbito de la investigación y los experimentos, y ampliar su disponibilidad mucho más allá de un puñado de startups y grandes empresas tecnológicas bien financiadas.

Amazon Bedrock ofrece a los clientes la manera más fácil de crear y escalar aplicaciones basadas en IA generativa utilizando los FM, lo que democratiza el acceso para todos los creadores. Bedrock proporciona acceso a una gama de potentes FM para texto e imágenes, incluidos los FM de **Amazon Titan**, a través de un servicio administrado de AWS escalable, fiable y seguro. Con la experiencia sin servidor de Bedrock, los clientes de AWS pueden encontrar fácilmente el modelo adecuado para lo que están tratando de hacer, comenzar rápidamente, personalizar de forma privada los FM con sus propios datos, e integrarlos e implementarlos fácilmente en sus aplicaciones utilizando las herramientas y capacidades de AWS con las que están familiarizados, sin tener que administrar ninguna infraestructura.

Los clientes de Bedrock pueden elegir entre algunos de los FM más avanzados disponibles en la actualidad. Esto incluye la familia Jurassic-2 de modelos de lenguaje grandes multilingües (LLM) de AI21 Labs, capaces de seguir instrucciones de lenguaje natural para generar texto en español, francés, alemán, portugués, italiano y neerlandés. Claude, el LLM de Anthropic para el diálogo reflexivo, la creación de contenido, el razonamiento complejo,

la creatividad y la codificación, puede realizar una amplia variedad de tareas conversacionales y de procesamiento de textos y se basa en la IA constitucional y en el entrenamiento de inocuidad. El modelo de generación de texto por comandos de Cohere está entrenado para seguir los comandos del usuario y ser de utilidad instantánea en aplicaciones empresariales prácticas, como el resumen, la redacción, el diálogo, la extracción y la respuesta a preguntas. El modelo de comprensión de textos de Cohere, Embed, puede utilizarse para tareas de búsqueda, agrupación de clústeres o clasificación en más de 100 idiomas, lo que permite a las organizaciones buscar fácilmente por significado o categorizar textos. Bedrock también facilita el acceso al paquete de FM de texto a imagen de Stability AI, incluido Stable Diffusion (el más popular de su clase), que puede generar imágenes, arte, logotipos y diseños únicos, realistas y de alta calidad.

Una de las capacidades más importantes de Bedrock es lo fácil que resulta personalizar un modelo. Los clientes solo tienen que remitir a Bedrock algunos ejemplos etiquetados en **Amazon Simple Storage Service (Amazon S3)**, y el servicio puede ajustar el modelo para una tarea concreta sin tener que anotar grandes volúmenes de datos (basta con 20 ejemplos).

Mientras Bedrock democratiza el acceso a los FM, la IA generativa también puede utilizarse para democratizar el desarrollo de software. Por ejemplo: **Amazon CodeWhisperer** es un complemento de programación basado en IA que utiliza un FM desde dentro del sistema para mejorar radicalmente la productividad de los desarrolladores al generar sugerencias de código en tiempo real basadas en los comentarios de los desarrolladores en lenguaje natural y en código previo de su entorno de desarrollo integrado (IDE). Durante la presentación preliminar CodeWhisperer, organizamos un desafío de productividad y los participantes que usaron CodeWhisperer completaron las tareas un 57 % más rápido, en promedio, y tenían un 27 % más de probabilidades de completarlas correctamente que los que no usaron CodeWhisperer. Se trata de un gran avance en la productividad de los desarrolladores, y creemos que es solo el principio. Además, CodeWhisperer es gratuito para desarrolladores individuales, por lo que empezar es muy fácil.

Operacionalizar

Con el rápido crecimiento de la adopción de la IA, los equipos empresariales y técnicos se enfrentan al desafío de crear más. En esta prisa por aprovechar las tecnologías, las organizaciones rara vez se detienen para establecer herramientas y procesos estándar para el desarrollo de ML. A causa de esto, diferentes equipos con diferentes habilidades y requisitos a menudo usan herramientas completamente diferentes y desconectadas, lo que hace que la colaboración sea poco práctica, incluso imposible.

Por ejemplo, un equipo de I+D podría estar trabajando en una aplicación de visión artificial (CV) con marcos y algoritmos de última generación, mientras que los equipos de ventas y marketing podrían estar creando un modelo de regresión lineal para pronosticar la demanda de los clientes en una hoja de cálculo almacenada localmente. O bien los desarrolladores podrían estar codificando una aplicación de compras móviles y desear agregar un motor de recomendaciones para que la experiencia del cliente sea más personal.

Relativamente pocas organizaciones utilizan herramientas y prácticas de ML operativas (como infraestructura, IDE, depuradores, perfiladores, herramientas de colaboración, flujos de trabajo y herramientas de gestión de proyectos) que se

puedan conectar de forma segura. Esta realidad complica la gestión entre equipos de analistas empresariales, desarrolladores y científicos de datos y la coordinación con las herramientas y procesos de software existentes. En estos escenarios comunes, aumentar o reducir la escala se vuelve extremadamente difícil.

La buena noticia es que existe una forma comprobada de minimizar los riesgos y las complicaciones del ML a la vez que se proporcionan prácticas sencillas y replicables a los equipos, mediante la operacionalización del ML.

La operacionalización del ML proporciona las herramientas, la infraestructura y el soporte de operaciones para escalar. La operacionalización del ML comienza con la adquisición de datos y con actividades de modelado del equipo de ciencia de datos informado por una comprensión clara de los objetivos comerciales para la aplicación de ML y de todos los problemas de gobernanza y conformidad. MLOps garantiza que los equipos de ciencia de datos, producción y operaciones trabajen juntos sin problemas en una serie de flujos de trabajo de ML que estén lo más automatizados posible. La intervención humana se incorpora según sea necesario, lo que garantiza implementaciones fluidas, el monitoreo de los datos y el seguimiento del rendimiento del modelo.

Cómo AWS ayuda a los clientes a operacionalizar el machine learning

Amazon SageMaker, que ya sabemos que es un servicio poderoso para ayudar a democratizar el aprendizaje automático, es igualmente adecuado para la operacionalización. Automatiza y estandariza cada paso del flujo de trabajo de MLOps para ayudar a escalar sin límites los proyectos. Gracias a SageMaker, los clientes de AWS ejecutan millones de modelos con miles de parámetros y generan cientos de miles de predicciones.

SageMaker también ofrece un servicio de ML integral para el etiquetado de datos, la preparación de datos, la ingeniería de características, la formación, el alojamiento, la supervisión y los flujos de trabajo a los que se puede acceder mediante una única interfaz visual en **Amazon SageMaker Studio**. En comparación con los entornos de ML autoadministrados, la productividad de los equipos de ciencia de datos puede mejorar hasta 10 veces y el tiempo de desarrollo del modelo se reduce de meses a semanas. Y todas las capacidades de SageMaker se ofrecen en una infraestructura de alto rendimiento, bajo costo y completamente administrada en la nube.

Con las herramientas de SageMaker, los clientes de AWS consiguen una gran escala (y grandes ahorros):

- **Vanguard** ha automatizado por completo la configuración de sus entornos de ML y ahora implementa modelos de ML 20 veces más rápido
- **AstraZeneca** puede implementar nuevos entornos de ML en cinco minutos en lugar de un mes para generar conocimientos que mejoren la I+D y aceleren la comercialización de nuevos fármacos
- **NerdWallet** redujo los costos de entrenamiento en casi un 75 %, incluso mientras aumentaba la cantidad de modelos entrenados
- **Zendesk** redujo los costos de inferencia de ML en un 90 % mediante la implementación de miles de modelos por punto de conexión mediante el uso de los puntos de conexión de múltiples modelos de SageMaker
- **Mueller Water Products** redujo a la mitad el número de alertas falsas y maximizó el potencial para identificar eventos de fuga reales

La operacionalización también implica que debemos abarcar la amplitud y la profundidad de los casos de uso de IA, incluidos los centros de contacto inteligentes, el procesamiento inteligente de documentos (IDP), la moderación de contenido, la personalización, la búsqueda inteligente, la prevención de fraudes, la verificación de identidad, el mantenimiento predictivo, la IA para DevOps, la IA para la salud y la inteligencia empresarial (BI) con tecnología de ML. AWS ofrece servicios para todos estos casos de uso y más.

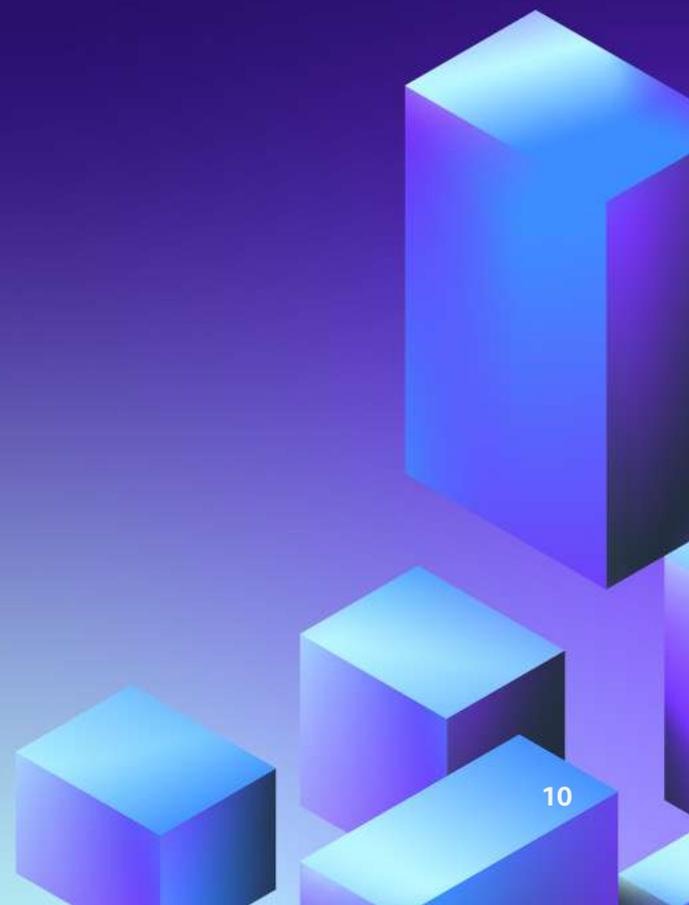


MUELLER

“Trabajar con AWS nos permite contar con un proceso de desarrollo más rápido”.

Kenji Takeuchi, SVP of Technology Solutions,
Mueller Water Products, 2021

[Lea más >](#)



CLAVE N.º 3

Generar confianza

Es muy importante que las organizaciones se ganen la confianza de sus clientes, socios y partes interesadas internas en relación con el uso de la IA generativa. Para crear y mantener esta confianza, las organizaciones deben realizar inversiones y tener en cuenta el uso responsable de la IA, la seguridad y la privacidad.

IA responsable

A medida que la IA generativa siga creciendo y evolucionando, será cada vez más importante adherirse a los principios del uso responsable de la IA para generar confianza y equilibrar la innovación potencial con los riesgos emergentes. Al abarcar un conjunto básico de conceptos (equidad, explicabilidad, solidez, seguridad, privacidad, transparencia y gobernanza), la IA responsable mitiga los riesgos mediante el uso transparente de datos y modelos. Se puede utilizar para aumentar el rendimiento de los modelos, mejorar la protección de datos y establecer mecanismos de detección y mitigación de sesgos en los sistemas de ML a fin de mejorar la equidad.

“El 94 % (de las empresas) tienen dificultades para operacionalizar todos los elementos clave de la IA responsable”.

Accenture, 2022

La IA responsable es una parte integral del ciclo de vida completo de la IA, que abarca desde el diseño inicial, el desarrollo y la infraestructura segura hasta la implementación y, en última instancia, el uso continuo. Es un proceso iterativo que requiere pruebas y auditorías continuas de posibles sesgos y precisión. Si bien la mayoría de las empresas han comenzado su camino hacia la IA responsable, la mayoría (el 94 %) tiene dificultades para operacionalizar todos los elementos clave de la IA responsable.³

Entonces, ¿cómo hacen las organizaciones para que la IA responsable pase de la teoría a la práctica? Comienzan educando a la próxima generación de líderes de ML para aumentar la equidad y mitigar el sesgo al traer perspectivas más diversas a la mesa, brindar recursos a fin de promover la educación y la formación, y garantizar la protección de datos y la privacidad. La IA responsable también requiere un esfuerzo multidisciplinario por parte de las empresas de tecnología, los políticos, los grupos comunitarios, los científicos y otros para abordar los nuevos desafíos a medida que surgen y trabajar para compartir las prácticas recomendadas y acelerar la investigación.



Cree una IA más responsable, segura y privada con AWS

Seguridad y privacidad

La seguridad y privacidad de datos es esencial para escalar la IA generativa de manera responsable. A la hora de personalizar y optimizar un modelo, las organizaciones deben saber dónde y cómo se utilizan sus datos. Deben tener la seguridad de que sus datos privados o su propiedad intelectual (PI) no se utilicen para entrenar a un modelo público y que los datos de sus clientes permanecen privados. Las organizaciones necesitan incorporar seguridad, escalabilidad y privacidad desde el principio con el fin de ser viables para sus aplicaciones empresariales.

Obtenga protecciones personalizadas con Amazon SageMaker

A medida que las organizaciones escalan su uso de las tecnologías de IA, pueden aprovechar los recursos de AWS para ayudar a implementar la IA responsable en todo el ciclo de vida del ML.

Las organizaciones pueden mitigar el sesgo y mejorar la explicabilidad con los servicios personalizados de AWS. **Amazon SageMaker Clarify** ayuda a mitigar el sesgo a lo largo del ciclo de vida del ML mediante la detección de posibles sesgos durante la preparación de los datos, después del entrenamiento del modelo y en el modelo implementado al examinar atributos específicos.

El monitoreo también es importante para mantener modelos de ML de alta calidad y garantizar predicciones precisas. El **Monitor de modelos de Amazon SageMaker**

detecta automáticamente y emite alertas cuando los modelos implementados en producción generan predicciones poco precisas.

Para mejorar la gobernanza, SageMaker proporciona **herramientas** personalizadas, como SageMaker Role Manager, SageMaker Model Cards y SageMaker Model Dashboard, que ofrecen un control más estricto y una visibilidad más completa de los modelos de ML. Los clientes de AWS pueden configurar usuarios con permisos de mínimo privilegio en cuestión de minutos; además, pueden capturar, recuperar y compartir fácilmente información importante del modelo y mantenerse informados sobre el comportamiento de este, como el sesgo, todo en un mismo lugar.

Mejorar la seguridad y la privacidad con Amazon Bedrock

Cuando los clientes utilizan Amazon Bedrock para personalizar un modelo, Bedrock puede ajustarlo para una tarea concreta sin tener que anotar grandes volúmenes de datos. A continuación, Bedrock hace una copia independiente del FM base a la que solo puede acceder el cliente y entrena esta copia privada del modelo.

Los clientes de AWS también pueden configurar sus ajustes de Amazon Virtual Private Cloud para acceder a las API de Bedrock y proporcionar datos de ajuste de modelos de forma segura. Los datos de los clientes siempre se cifran en tránsito (TLS1.2) y en reposo mediante claves administradas por el servicio.

AWS tiene en cuenta el uso responsable de la IA en cada fase del proceso de desarrollo de su modelo de base (FM) integral. A lo largo del diseño, el desarrollo, la implementación y el funcionamiento del FM, tenemos en cuenta:



Precisión



Equidad



**Propiedad intelectual y
derechos de autor**



Uso adecuado



Toxicidad



Privacidad

Para resolver estos problemas, incorporamos soluciones a nuestros procesos de adquisición de datos de formación, a los propios FM y a la tecnología que utilizamos para procesar previamente las solicitudes de los usuarios y para procesar los resultados posteriormente. Para todos nuestros FM, invertimos activamente para mejorar nuestras características y aprender de los clientes a medida que experimentan con nuevos casos de uso.

Recursos

Escuche a un académico de Amazon hablar sobre los nuevos desafíos y soluciones para desarrollar un IA generativa de forma responsable ›

Obtenga más información sobre los nuevos compromisos de la Casa Blanca, las organizaciones tecnológicas y la comunidad de IA para progresar en el uso responsable y seguro de la IA ›

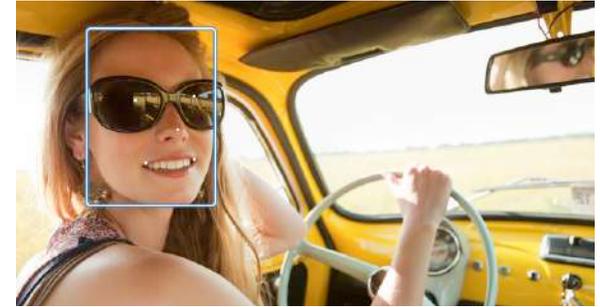
Consulte tres recursos esenciales para una IA más responsable:



1. En la guía *Responsible Use of Machine Learning* se ofrecen consideraciones y recomendaciones para el uso y desarrollo de los sistemas de ML de manera responsable en las tres fases principales de sus ciclos de vida: 1) diseño y desarrollo, 2) implementación y 3) uso continuo. [Más información >](#)



2. La educación continua sobre los últimos desarrollos en ML es una parte importante del uso responsable. AWS ofrece lo último en educación sobre ML en su camino de aprendizaje a través de programas como [AWS Machine Learning University \(MLU\) Bias and Fairness Course](#), [el programa Training & Certification](#) y [el programa AI and ML Scholarship](#).



3. Las Tarjetas de servicio de IA de AWS proporcionan transparencia y documentan los casos de uso previstos y las consideraciones de equidad para nuestros servicios de IA de AWS. Descubra las tarjetas de servicio de IA: comparación de rostros de [Amazon Rekognition](#), [Amazon Textract AnalyzeID](#) y [Amazon Transcribe - Batch \(inglés-EE. UU.\)](#).



AWS se compromete a seguir desarrollando la IA y el ML de forma responsable.

[Más información >](#)

Los resultados satisfactorios comienzan aquí

Más de 100 000 clientes han elegido AWS para IA a fin de crear nuevas experiencias del cliente, optimizar sus negocios, aumentar el ingenio de sus empleados, ayudar a mejorar la calidad de sus productos y mucho más. Eso se debe a que AWS lo respalda sin importar dónde se encuentre en su camino de ML, con las soluciones que necesita para escalar sin límites.

Comuníquese con los expertos de AWS

- El equipo global de **AWS Professional Services** está compuesto por expertos que pueden ayudarlo a lograr los resultados empresariales que desea en torno a la nube de AWS.
- El **Centro de innovación de IA generativa de AWS** lo conecta con expertos en IA y ML de AWS para que le ayuden a visualizar, diseñar y lanzar nuevos productos, servicios y procesos de IA generativa.

Colabore con un socio oficial de AWS

- Los **socios de AWS** están en una posición única para ayudar a los clientes de AWS a acelerar el camino a la nube de AWS.

Constrúyalo usted mismo con soluciones comprobadas

- El **Explorador de casos de uso de IA** lo ayuda a descubrir los principales casos de uso de IA, historias de clientes y rutas de implementación en función de sus objetivos empresariales.
- La **Biblioteca de soluciones de AWS** ofrece soluciones creadas por AWS y socios de AWS para una amplia variedad de casos de uso.
- Los **servicios de IA de AWS** le permiten agregar fácilmente inteligencia a las aplicaciones, sin necesidad de conocimientos de ML.
- **Amazon SageMaker** permite a los usuarios crear, entrenar e implementar modelos de ML para cualquier caso de uso con infraestructura, herramientas y flujos de trabajo completamente administrados.
- **Amazon Bedrock** hace accesibles a través de una API los FM de Amazon y de las principales startups de IA, como AI21 Labs, Anthropic, Cohere y Stability AI.

Desde las empresas más grandes del mundo hasta las startups, se crea más IA en AWS que en cualquier otro lugar.

Obtenga más información sobre cómo su empresa puede ofrecer resultados exitosos de IA y ML ›

