



[¿Qué es la informática en la nube?](#) / [Centro de conceptos de computación en la nube](#)
/ [IA generativa](#) / [Gen-AI](#)

¿Qué es la IA generativa?

[Cree una cuenta de AWS](#)



Explorar servicios de IA generativa

Cree, despliegue y ejecute aplicaciones de inteligencia artificial (IA) generativa en AWS



Explorar la IA generativa en AWS

Innove más rápido con el conjunto más completo de servicios de IA generativa



Ver capacitaciones en IA generativa

Comience su capacitación en IA generativa con contenido creado por expertos de AWS



¿Qué es la IA generativa?

La inteligencia artificial generativa (IA generativa) es un tipo de inteligencia artificial que puede crear ideas y contenidos nuevos, como conversaciones, historias, imágenes, videos y música. Las tecnologías de IA intentan imitar la inteligencia humana en tareas informáticas no tradicionales, como el reconocimiento de imágenes, el *natural language processing* (NLP, procesamiento de lenguaje natural) y la traducción. La IA generativa es el siguiente paso en la inteligencia artificial. Puede entrenarla para que aprenda lenguaje humano, lenguajes de programación, arte, química, biología o cualquier tema complejo. Reutiliza los datos de entrenamiento para resolver nuevos problemas. Por ejemplo, puede aprender vocabulario en inglés y crear un poema a partir de las palabras que procesa. Su organización puede utilizar la IA generativa para diversos fines, como los chatbots, la creación de medios y el desarrollo y diseño de productos.

¿Por qué es importante la IA generativa?

Las aplicaciones de IA generativa como ChatGPT han captado la atención y la imaginación generalizadas. Pueden ayudar a reinventar la mayoría de las experiencias y aplicaciones de los clientes, crear nuevas aplicaciones nunca antes vistas y ayudar a los clientes a alcanzar nuevos niveles de productividad.

Según Goldman Sachs, la IA generativa podría impulsar un aumento del 7 por ciento (o casi 7 billones de dólares) en el producto interno bruto (PIB) mundial. También anticipan que podría aumentar el crecimiento de la productividad en 1,5 puntos porcentuales en 10 años.

A continuación, presentamos algunos beneficios más de la IA generativa.

Acelera la investigación

Los algoritmos de la IA generativa pueden explorar y analizar datos complejos de nuevas formas. Por lo tanto, los investigadores pueden descubrir nuevas tendencias y patrones que, de otro modo, no serían evidentes. Estos algoritmos pueden resumir contenido, esbozar múltiples vías de solución, proporcionar muchas ideas y crear documentación detallada a partir de notas de investigación. Esta es la razón por la que la IA generativa mejora drásticamente la investigación y la innovación.

Por ejemplo, los sistemas de IA generativa se utilizan en la industria farmacéutica para generar y optimizar secuencias de proteínas y acelerar significativamente el descubrimiento de fármacos.

Mejora las experiencias de los clientes

inteligencia artificial que respondan con mayor precisión a los clientes para resoluciones en el primer contacto. Pueden aumentar la participación de los clientes al presentar ofertas seleccionadas y comunicaciones de forma personalizada.

Optimiza los procesos empresariales

Con la IA generativa, su empresa puede optimizar los procesos empresariales mediante aplicaciones de machine learning (ML) e IA en todas las líneas de negocio. Puede aplicar esta tecnología en todas las líneas de negocio, incluidas la ingeniería, el marketing, el servicio al cliente, las finanzas y las ventas.

Por ejemplo, esto es lo que la IA generativa puede hacer para la optimización:

- Extraer y resumir datos de cualquier fuente para funciones de búsqueda de conocimiento.
- Evaluar y optimizar diferentes escenarios para reducir costos en áreas como marketing, publicidad, finanzas y logística.
- Generar datos sintéticos para crear datos etiquetados para el aprendizaje supervisado y otros procesos de ML.

Aumenta la productividad de los empleados

Los modelos de IA generativa pueden aumentar los flujos de trabajo de los empleados y actuar como asistentes eficientes para todos los miembros de su organización. Pueden hacer de todo, desde la búsqueda hasta la creación, de una manera similar a los humanos.

La IA generativa puede aumentar la productividad de diferentes tipos de trabajadores:

- Respalda las tareas creativas mediante la generación de varios prototipos en función de determinadas entradas y restricciones. También puede optimizar los diseños existentes en función de la retroalimentación de humanos y las restricciones específicas.
- Genera nuevas sugerencias de código de software para las tareas de desarrollo de aplicaciones.
- Respalda la administración mediante la generación de informes, resúmenes y proyecciones.
- Genera nuevos guiones de ventas, contenido de correo electrónico y blogs para los equipos de marketing.

Puede ahorrar tiempo, reducir costos y mejorar la eficiencia en toda su organización.

¿Cómo funciona la IA generativa?

[Foundation models](#) (FM, modelos fundacionales) son modelos de ML entrenados en un amplio espectro de datos generalizados y sin etiquetar. Son capaces de realizar una amplia variedad de tareas generales.

Los FM son el resultado de los últimos avances de una tecnología que ha estado evolucionando durante décadas. En general, un modelo fundacional utiliza relaciones y patrones aprendidos para predecir el siguiente elemento de una secuencia.

Por ejemplo, con la generación de imágenes, el modelo analiza la imagen y crea una versión de ella más nítida y definida. Del mismo modo, con texto, el modelo predice la siguiente palabra de una cadena de texto en función de las palabras anteriores y su contexto. Luego, selecciona la siguiente palabra mediante el uso de técnicas de distribución de probabilidad.

Modelos de lenguaje de gran tamaño

Los [large language models](#) (LLM, modelos de lenguaje de gran tamaño) son una clase de FM. Por ejemplo, los modelos de *generative pre-trained transformer* (GPT, transformadores generativos preentrenados) de OpenAI son LLM. Los LLM se centran específicamente en tareas basadas en el lenguaje, como el resumen, la generación de texto, la clasificación, la conversación abierta y la extracción de información.

[Más información sobre GPT »](#)

Lo que hace que los LLM sean especiales es su capacidad para realizar múltiples tareas. Pueden hacerlo porque contienen muchos parámetros que los hacen capaces de aprender conceptos avanzados.

Un LLM como el GPT-3 puede considerar miles de millones de parámetros y tiene la capacidad de generar contenido a partir de muy pocas entradas. A través de su exposición previa al entrenamiento a datos a escala de Internet en todas sus diversas formas y en una miríada de patrones, los LLM aprenden a aplicar sus conocimientos en una amplia gama de contextos.

¿Cómo afectará la IA generativa a las industrias?

Si bien la IA generativa podría afectar a todos los sectores con el tiempo, algunos sectores están preparados para beneficiarse rápidamente de esta tecnología.

Servicios financieros

general.

- Las instituciones crediticias pueden acelerar la aprobación de préstamos para mercados desatendidos desde el punto de vista financiero, especialmente en países en desarrollo.
- Los bancos pueden detectar rápidamente el fraude en las reclamaciones, en las tarjetas de crédito y en los préstamos.
- Las empresas de inversión pueden utilizar la potencia de la IA generativa para brindar asesoramiento financiero personalizado y seguro a sus clientes a bajo costo.

Sanidad y ciencias biológicas

Uno de los casos de uso más prometedores de la IA generativa es acelerar el descubrimiento y la investigación de fármacos. La IA generativa utiliza modelos para crear nuevas secuencias de proteínas con propiedades específicas para diseñar anticuerpos, enzimas, vacunas y terapias génicas.

Las empresas de salud y ciencias de la vida pueden utilizar modelos generativos con el fin de diseñar secuencias de genes sintéticos para aplicaciones en biología sintética e ingeniería metabólica. Por ejemplo, pueden crear nuevas vías biosintéticas u optimizar la expresión génica con fines de biofabricación.

Por último, la IA generativa se puede utilizar para crear datos sintéticos de pacientes y de atención médica. Esto resulta útil para entrenar modelos de IA, simular ensayos clínicos o estudiar enfermedades raras sin acceso a grandes conjuntos de datos del mundo real.

Automoción y fabricación

Las empresas automotrices pueden utilizar la tecnología de la IA generativa con muchos propósitos, desde ingeniería hasta experiencias a bordo de vehículos y servicio al cliente. Por ejemplo, pueden optimizar el diseño de las piezas mecánicas para reducir la resistencia aerodinámica en los diseños de los vehículos o adaptar el diseño de los asistentes personales.

Las compañías automotrices utilizan la IA generativa para ofrecer un mejor servicio al cliente al proporcionar respuestas rápidas a las preguntas más frecuentes. Se pueden crear nuevos diseños de materiales, chips y piezas con IA generativa para optimizar los procesos de fabricación y reducir los costos.

La IA generativa también se puede utilizar para generar datos sintéticos para probar aplicaciones. Esto es especialmente útil para los datos que no suelen incluirse en los conjuntos de datos de prueba (como defectos o casos extremos).

Contenido multimedia y entretenimiento

-
- Los artistas pueden complementar y mejorar sus álbumes con música generada por IA para crear experiencias completamente nuevas.
 - Las organizaciones de medios pueden utilizar la IA generativa para mejorar las experiencias de su audiencia al ofrecer contenido y anuncios personalizados para aumentar sus ingresos.
 - Las compañías de juegos pueden usar la IA generativa para crear nuevos juegos y permitir a los jugadores crear avatares.

Telecomunicación

Los primeros casos de uso de la IA generativa en las telecomunicaciones se centran en reinventar la experiencia del cliente. La experiencia del cliente se define por las interacciones acumuladas de los suscriptores en todos los puntos de contacto del recorrido del cliente.

Por ejemplo, las organizaciones de telecomunicaciones pueden aplicar la IA generativa para mejorar el servicio al cliente con agentes conversacionales en vivo similares a los humanos. También pueden optimizar el rendimiento de la red mediante el análisis de los datos de la red para recomendar soluciones. Además, pueden reinventar las relaciones con los clientes con asistentes de ventas individuales personalizados.

Energía

La IA generativa es adecuada para tareas del sector energético que implican el análisis, el reconocimiento de patrones, la previsión y la optimización de datos sin procesar complejos. Las organizaciones de energía pueden mejorar el servicio al cliente mediante el análisis de los datos empresariales para identificar los patrones de uso. Con esta información, pueden desarrollar ofertas de productos específicas, programas de eficiencia energética o iniciativas de respuesta a la demanda.

La IA generativa puede ayudar a gestionar la red, aumentar la seguridad del sitio operativo y optimizar la producción de energía mediante la simulación de yacimientos.

¿Cómo funcionan los modelos de IA generativa?

Los modelos tradicionales de machine learning eran discriminativos o se centraban en clasificar los puntos de datos. Intentaban determinar la relación entre los factores conocidos y desconocidos. Por ejemplo, observan imágenes (datos conocidos como la disposición de los píxeles, la línea, el color y la forma) y las asignan a palabras (el factor desconocido).

Matemáticamente, los modelos funcionaban identificando ecuaciones que podían mapear numéricamente factores desconocidos y conocidos como variables x e y .

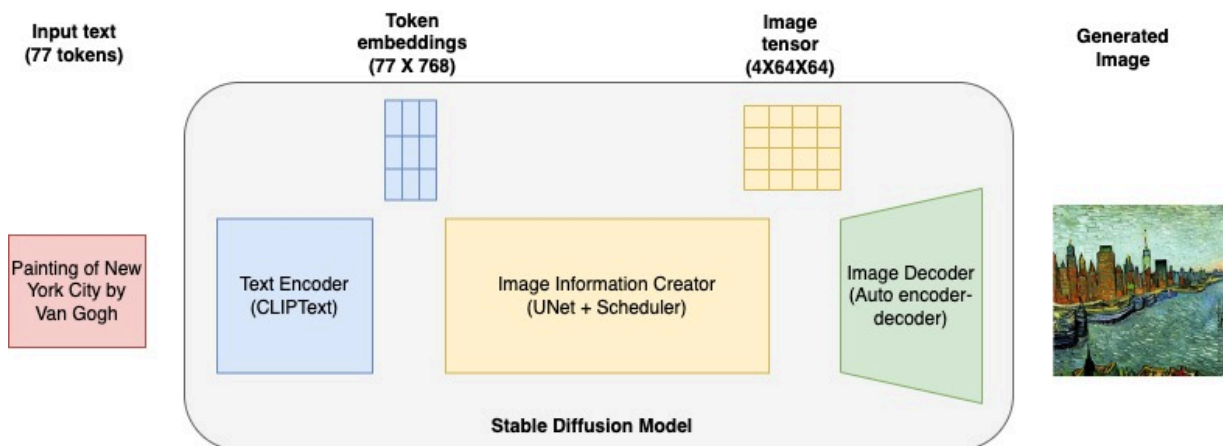
Por ejemplo, los modelos generativos analizan imágenes de animales para registrar variables como diferentes formas de orejas, formas de ojos, rasgos de la cola y patrones de piel. Aprenden las características y sus relaciones para comprender cómo son los diferentes animales en general. Luego, pueden recrear nuevas imágenes de animales que no estaban en el set de entrenamiento.

A continuación, ofrecemos algunas categorías amplias de modelos de IA generativa.

Modelos de difusión

Los modelos de difusión crean nuevos datos realizando, de forma iterativa, cambios aleatorios controlados en una muestra de datos inicial. Empiezan con los datos originales y añaden cambios sutiles (ruido), haciendo que pierdan la similitud con el original progresivamente. Este ruido se controla cuidadosamente para garantizar que los datos generados sigan siendo coherentes y realistas.

Tras añadir ruido en varias iteraciones, el modelo de difusión invierte el proceso. La eliminación de ruido inversa elimina gradualmente el ruido para producir una nueva muestra de datos que se parezca a la original.



Redes generativas adversativas

La [generative adversarial network](#) (GAN, red generativa adversativa) es otro modelo de IA generativa que se basa en el concepto del modelo de difusión.

Las GAN funcionan entrenando dos redes neuronales de manera competitiva. La primera red, conocida como *generador*, crea muestras de datos falsas añadiendo ruido aleatorio. La segunda red, denominada *discriminador*, trata de distinguir entre los datos reales y los datos falsos producidos por el generador.

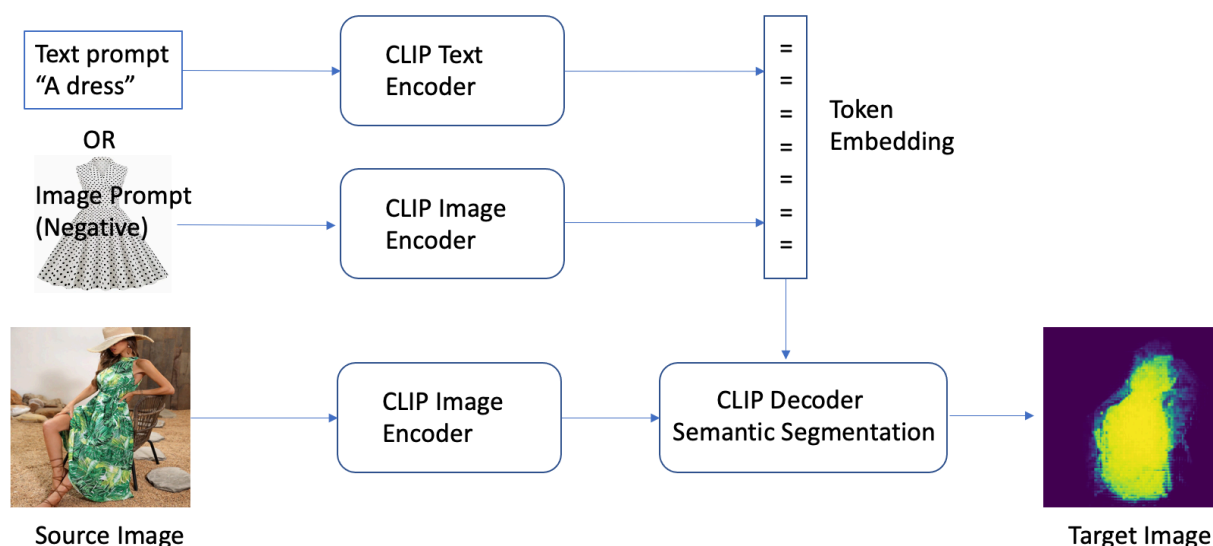
Las GAN se utilizan ampliamente para generar imágenes realistas, transferir estilos y realizar tareas de aumento de datos.

Autocodificadores variacionales

Los VAE aprenden una representación compacta de los datos llamada *espacio latente*. El espacio latente es una representación matemática de los datos. Puede considerarlo como un código único que representa los datos en función de todos sus atributos. Por ejemplo, si estudia rostros, el espacio latente contiene números que representan la forma de los ojos, la forma de la nariz, los pómulos y las orejas.

Los VAE utilizan dos redes neuronales: el *codificador* y el *decodificador*. La red neuronal del codificador mapea los datos de entrada a una media y una varianza para cada dimensión del espacio latente. Genera una muestra aleatoria a partir de una distribución gaussiana (normal). Este ejemplo es un punto en el espacio latente y representa una versión comprimida y simplificada de los datos de entrada.

La red neuronal del decodificador toma este punto muestreado del espacio latente y lo reconstruye en datos que se parecen a la entrada original. Las funciones matemáticas se utilizan para medir qué tan buena es la coincidencia entre los datos reconstruidos y los datos originales.



Modelos basados en transformadores

El modelo de IA generativa basado en transformadores se basa en los conceptos de codificador y decodificador de los VAE. Los modelos basados en transformadores añaden más capas al codificador para mejorar el rendimiento en las tareas basadas en texto, como la comprensión, la traducción y la escritura creativa.

La codificación de un elemento de secuencia depende no solo del elemento en sí, sino también de su contexto dentro de la secuencia.

Cómo funcionan los modelos basados en transformadores

Para entender cómo funcionan los modelos basados en transformadores, imagine una oración como una secuencia de palabras.

La autoatención ayuda al modelo a centrarse en las palabras relevantes a medida que procesa cada palabra. Para capturar diferentes tipos de relaciones entre palabras, el modelo generativo basado en transformadores emplea múltiples capas de codificación denominadas *cabezas de atención*. Cada cabezal aprende a prestar atención a diferentes partes de la secuencia de entrada. Esto permite que el modelo considere simultáneamente diferentes aspectos de los datos.

Cada capa también refina las incrustaciones contextuales. Las capas hacen que las incrustaciones sean más informativas y capturan todo, desde la sintaxis gramatical hasta los significados semánticos complejos.

¿Cómo evolucionó la tecnología de IA generativa?

Los modelos generativos primitivos se han utilizado durante décadas en estadística para ayudar en el análisis de datos numéricos. Las redes neuronales y el aprendizaje profundo fueron los precursores recientes de la IA generativa moderna. Los *variational autoencoders* (VAE, autocodificadores variacionales), desarrollados en 2013, fueron los primeros modelos generativos profundos que podían generar imágenes y voz realistas.

Los VAE introdujeron la capacidad de crear variaciones novedosas de varios tipos de datos. Esto llevó a la rápida aparición de otros modelos de IA generativa, como las redes generativas adversativas y los modelos de difusión. Estas innovaciones se centraron en generar datos que se parecieran cada vez más a datos reales, a pesar de haber sido creados artificialmente.

En 2017, se produjo un nuevo cambio en la investigación sobre la IA con la introducción de los transformadores. Los transformadores integraron a la perfección la arquitectura del codificador y el decodificador con un mecanismo de atención. Agilizaron el proceso de formación de modelos de lenguaje con una eficacia y versatilidad excepcionales. Modelos notables como el GPT surgieron como modelos fundamentales capaces de entrenarse previamente sobre extensos corpus de texto sin procesar y de adaptarlos para diversas tareas.

aplicaciones en todos los sectores. Las innovaciones recientes se centran en refinar los modelos para que funcionen con datos propietarios. Los investigadores también quieren crear textos, imágenes, vídeos y discursos que se parezcan cada vez más a los producidos por humanos.

¿Cuáles son las prácticas recomendadas en la adopción de la IA generativa?

Si su organización desea implementar soluciones de IA generativa, tenga en cuenta las siguientes prácticas recomendadas para mejorar sus esfuerzos.

Comience con las aplicaciones internas

Es mejor comenzar la adopción de la IA generativa con el desarrollo de aplicaciones internas, centrándose en la optimización de los procesos y la productividad de los empleados. Obtiene un entorno más controlado para probar los resultados y, al mismo tiempo, desarrollar habilidades y comprender la tecnología. Puede probar los modelos exhaustivamente e incluso personalizarlos en fuentes de conocimiento internas.

De esta forma, sus clientes tendrán una experiencia mucho mejor cuando finalmente utilice los modelos para aplicaciones externas.

Mejore la transparencia

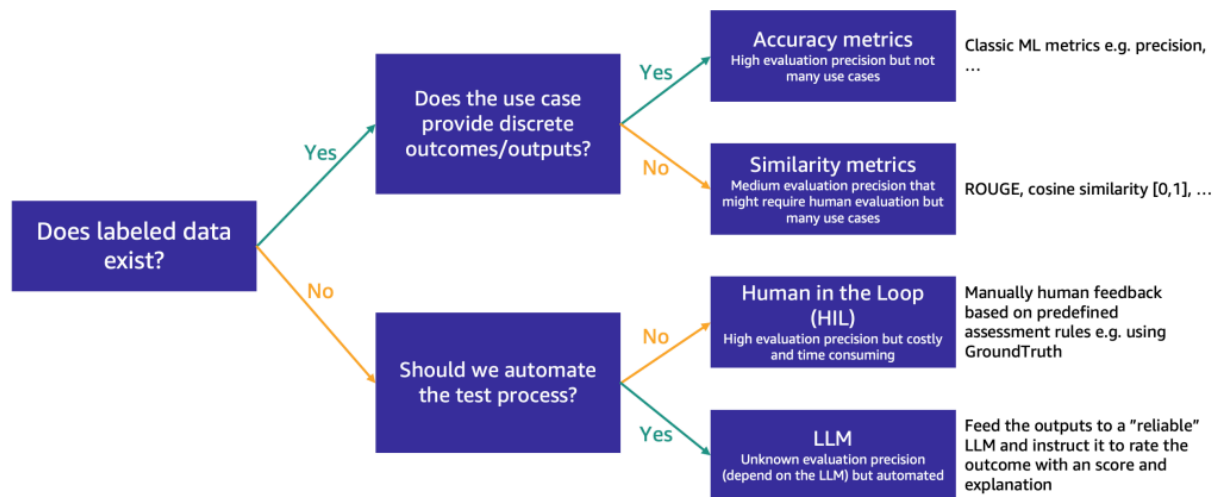
Comunique con claridad todas las aplicaciones y resultados de la IA generativa, de modo que sus usuarios sepan que están interactuando con IA y no con humanos. Por ejemplo, la IA puede presentarse como IA, o los resultados de búsqueda basados en la IA pueden estar marcados y resaltados.

De esta forma, sus usuarios pueden usar su propio criterio cuando interactúan con el contenido. También pueden ser más proactivos a la hora de abordar cualquier imprecisión o sesgo oculto que puedan tener los modelos subyacentes debido a las limitaciones de sus datos de entrenamiento.

Implemente la seguridad

Implemente barreras de protección para que sus aplicaciones de IA generativa no permitan el acceso inadvertido y no autorizado a datos confidenciales. Involucre a los equipos de seguridad desde el principio para que se puedan considerar todos los aspectos desde el principio. Por ejemplo, es posible que tenga que enmascarar los datos y eliminar la *personally identifiable*

Desarrolle procesos de prueba automatizados y manuales para validar los resultados y pruebe todo tipo de escenarios que pueda experimentar el sistema de IA generativa. Tenga diferentes grupos de probadores beta que prueben las aplicaciones de diferentes maneras y documenten los resultados. El modelo también mejorará continuamente a través de las pruebas, y usted tendrá más control sobre los resultados y las respuestas esperados.



¿Cuáles son las aplicaciones comunes de la IA generativa?

Gracias a la IA generativa, puede aprovechar el machine learning para su empresa con mayor rapidez y aplicarlo a un conjunto más amplio de casos de uso. Puede aplicar la IA generativa en todas las líneas de negocio, incluidas la ingeniería, el marketing, el servicio al cliente, las finanzas y las ventas. La generación de código es uno de los usos más prometedores de la IA generativa y, gracias a [Amazon Q Developer](#), un asistente que utiliza tecnología de IA generativa para el desarrollo de software, obtenemos grandes resultados en la productividad de los desarrolladores.

Además de la generación de código, hay muchas aplicaciones en las que puede utilizar la IA generativa para lograr un cambio radical en la experiencia del cliente, la productividad de los empleados, la eficiencia empresarial y la creatividad. Puede utilizar la IA generativa para mejorar la experiencia del cliente mediante capacidades como chatbots, asistentes virtuales, centros de atención inteligentes, personalización y moderación del contenido. También es posible aumentar la productividad de sus empleados con la búsqueda conversacional, la creación de contenido y el resumen de texto impulsados por IA generativa, entre otros. Además, la IA generativa le permite mejorar las operaciones comerciales con el procesamiento inteligente de documentos, los asistentes de mantenimiento, el control de calidad y la inspección visual y la generación de datos de entrenamiento sintéticos. Por último, puede

¿Cómo puede AWS ayudar a la IA generativa?

Amazon Web Services (AWS) facilita la creación y el escalado de aplicaciones de IA generativa para sus datos, casos de uso y clientes. Con la [IA generativa en AWS](#), obtiene seguridad y privacidad de nivel empresarial, acceso a FM líderes del sector, aplicaciones impulsadas por IA generativa y un enfoque centrado en los datos.

Elija entre una gama de [servicios de IA generativa](#) que respaldan a todo tipo de organizaciones en todas las etapas de la adopción y madurez de la IA generativa:

- La generación de código es una de las aplicaciones más prometedoras de la IA generativa. Con [Amazon Q Developer](#), un asistente con tecnología de IA generativa que se utiliza para el desarrollo de software, puede obtener grandes resultados en la productividad de los desarrolladores.
- [Amazon Bedrock](#) es otro servicio totalmente gestionado que ofrece una selección de FM de alto rendimiento y un amplio conjunto de capacidades. Puede experimentar fácilmente con varios de los principales FM, personalizarlos de forma privada con sus datos y crear agentes gestionados que ejecuten tareas empresariales complejas.
- También puede usar [Amazon SageMaker JumpStart](#) para descubrir, explorar e implementar FM de código abierto, o incluso crear uno propio. SageMaker JumpStart proporciona infraestructuras y herramientas gestionadas para acelerar la creación, el entrenamiento y la implementación de modelos escalables, fiables y seguros.
- [AWS HealthScribe](#) es un servicio que cumple con los requisitos de la HIPAA y que permite a los proveedores de software para la atención médica crear aplicaciones clínicas que generen automáticamente notas clínicas mediante el análisis de las conversaciones entre el paciente y el médico. AWS HealthScribe combina el reconocimiento de voz y la inteligencia artificial (IA) generativa para reducir la carga de la documentación clínica mediante la transcripción de las conversaciones entre el paciente y el médico y la generación de notas clínicas más fáciles de revisar.
- [Amazon Q en QuickSight](#) ayuda a los analistas empresariales a crear y personalizar imágenes con facilidad mediante comandos en lenguaje natural. Las nuevas [capacidades de creación de BI generativa](#) amplían las consultas en lenguaje natural de QuickSight Q más allá de responder a preguntas bien estructuradas (como «¿cuáles son los 10 productos más vendidos en California?») para ayudar a los analistas a crear rápidamente imágenes personalizables a partir de fragmentos de preguntas (por ejemplo, «los 10 mejores productos»), aclarar el propósito de una consulta formulando preguntas de seguimiento, afinar las visualizaciones y realizar cálculos complejos.

[Cree una cuenta](#) hoy mismo para comenzar a utilizar la IA generativa en AWS.



Descubra otros recursos relacionados con el producto

[Genere innovaciones más rápido con los servicios de IA generativa de AWS »](#)



Regístrese para obtener una cuenta gratuita

Obtenga acceso instantáneo al nivel Gratuito de AWS.

[Regístrese »](#)



Comenzar a crear en la consola

Comience a crear en la consola de administración de AWS.

[Iniciar sesión »](#)

Recursos para AWS

Introducción

Formación técnica y certificación

Biblioteca de soluciones de AWS

Centro de arquitectura

Preguntas frecuentes sobre cuestiones técnicas y productos

Informes de analistas

Socios de AWS

Información sobre AWS

¿Qué es AWS?

¿Qué es la informática en la nube?

Inclusión, diversidad e igualdad en AWS

¿Qué es DevOps?

¿Qué es un contenedor?

¿Qué es un lago de datos?

Seguridad en la nube de AWS

Desarrolladores en AWS Ayuda

[Centro de desarrolladores](#)[SDK y herramientas](#)[.NET en AWS](#)[Python en AWS](#)[Java en AWS](#)[PHP en AWS](#)[JavaScript en AWS](#)[Contáctenos](#)[Reciba ayuda de expertos](#)[Abra un ticket de soporte técnico](#)[AWS re:Post](#)[Centro de conocimientos](#)[Información general de AWS Support](#)[Legal](#)[Carreras en AWS](#)

Amazon es un empleador que ofrece igualdad de oportunidades: *minorías, mujeres, discapacitados, veteranos, identidad de género, orientación sexual y edad.*

Idioma

[عربي |](#)[Bahasa Indonesia |](#)[Deutsch |](#)[English |](#)[Español |](#)[Français |](#)[Italiano |](#)[Português |](#)[Tiếng Việt |](#)[Türkçe |](#)[Русский |](#)[ไทย |](#)[日本語 |](#)[한국어 |](#)[中文 \(简体\) |](#)[中文 \(繁體\)](#)

Privacidad

[|](#)

