General

What is Amazon Bedrock?

Amazon Bedrock is a fully managed service that offers a choice of high-performing foundation models (FMs) from leading AI companies including AI21 Labs, Anthropic, Cohere, Meta, Stability AI, and Amazon, along with a broad set of capabilities that you need to build generative AI applications, simplifying development while maintaining privacy and security. With Amazon Bedrock's comprehensive capabilities, you can easily experiment with a variety of top FMs, customize them privately with your data using techniques such as fine tuning and retrieval-augmented generation (RAG), and create managed agents that execute complex business tasks—from booking travel and processing insurance claims to creating ad campaigns and managing inventory—all without writing any code. Since Amazon Bedrock is serverless, you don't have to manage any infrastructure, and you can securely integrate and deploy generative AI capabilities into your applications using the AWS services you are already familiar with.

Which FMs are available on Amazon Bedrock?

Amazon Bedrock customers can choose from some of the most cutting-edge FMs available today. This includes Anthropic's Claude, AI21 Labs' Jurassic-2, Stability AI's Stable Diffusion, Cohere's Command and Embed, Meta's Llama 2, and the Amazon Titan language and embeddings models.

Why should I use Amazon Bedrock?

There are five reasons to use Amazon Bedrock for building generative AI applications.

- Choice of leading foundation models: Amazon Bedrock offers an easy-to-use developer experience to work with a broad range of high-performing FMs from Amazon and leading AI companies like AI21 Labs, Anthropic, Cohere, Meta, and Stability AI. You can quickly experiment with a variety of FMs in the playground, and use a single API for inference regardless of the models you choose, giving you the flexibility to use FMs from different providers and keep up to date with the latest model versions with minimal code changes.
- Easy model customization with your data: Privately customize FMs with your own data through a visual interface without writing any code. Simply select the training and validation data sets stored in Amazon Simple Storage Service (Amazon S3) and, if required, adjust the hyperparameters to achieve the best possible model performance.
- Fully managed agents that can invoke APIs dynamically to execute tasks: Build agents that execute complex business tasks—from booking travel and processing insurance claims to creating ad campaigns, preparing tax filings, and managing your inventory—by dynamically calling your company systems and APIs. Amazon Bedrock's fully managed agents extend the reasoning capabilities of FMs to break down tasks, create an orchestration plan, and execute it.

- Native support for RAG to extend the power of FMs with proprietary data: With Knowledge Bases for Amazon Bedrock, you can securely connect FMs to your data sources for retrieval augmentation—from within the managed service—extending the FM's already powerful capabilities and making it more knowledgeable about your specific domain and organization.
- Data security and compliance certifications: Amazon Bedrock offers several capabilities to support security and privacy requirements, and has achieved HIPAA eligibility and GDPR compliance. With Amazon Bedrock, your content is not used to improve the base models and is not shared with third-party model providers. Your data in Amazon Bedrock is always encrypted in transit and at rest, and you can optionally encrypt the data using your own keys. You can use AWS PrivateLink with Amazon Bedrock to establish private connectivity between your FMs and your Amazon Virtual Private Cloud (Amazon VPC) without exposing your traffic to the Internet.

How can I get started with Amazon Bedrock?

With the serverless experience of Amazon Bedrock, you can quickly get started. Navigate to Amazon Bedrock in the AWS console and try out the FMs in the playground. You can also create an agent and test it in the console. Once you've identified your use case, you can easily integrate the FMs into your applications using AWS tools without having to manage any infrastructure.

How does Amazon Bedrock work with other services?

Amazon Bedrock leverages AWS Lambda for invoking actions, Amazon S3 for training and validation data, and Amazon CloudWatch for tracking metrics.

What are the most common use cases for Amazon Bedrock with Amazon Lex?

You can get started with use cases quickly

- Create new pieces of original content, such as short stories, essays, social media posts, and web page copy.
- Search, find, and synthesize information to answer questions from a large corpus of data.
- Create realistic and artistic images of various subjects, environments, and scenes from language prompts.
- Help customers find what they're looking for with more relevant and contextual product recommendations than word matching.
- Get a summary of textual content such as articles, blog posts, books, and documents to get the gist without having to read the full content.

What is Amazon Bedrock Chat Playground?

Amazon Bedrock offers a playground that allows you to experiment with various FMs using a conversational chat interface. You can provide a prompt and use a web interface inside the AWS Management Console to supply a prompt and use the pretrained models to generate text or images, or alternatively use a fine-tuned model that has been adapted for your use case.

In which AWS Regions is Amazon Bedrock available?

For a list of AWS Regions where Amazon Bedrock is available, see Amazon Bedrock endpoints and quotas in the Amazon Bedrock Reference Guide.

How do I customize a model on Amazon Bedrock?

You can easily fine-tune FMs on Amazon Bedrock. To get started, provide the training and validation dataset, configure hyperparameters (epochs, batch size, learning rate, warmup steps) and submit the job. Within a couple of hours, your fine-tuned model can be accessed with the same API (InvokeModel).

Can I train a model and deploy it on Amazon Bedrock?

Amazon Bedrock is a managed service that you can use to access foundational models. You can fine-tune a model and use it with the Amazon Bedrock API.

Agents for Amazon Bedrock

What are agents for Amazon Bedrock?

Agents for Amazon Bedrock are fully managed capabilities that make it easier for developers to create generative AI-based applications that can complete complex tasks for a wide range of use cases and deliver up-to-date answers based on proprietary knowledge sources. With just a few clicks, agents for Amazon Bedrock automatically break down tasks and create an orchestration plan–without any manual coding. The agent securely connects to company data through an API, automatically converting data into a machine-readable format, and augmenting the request with relevant information to generate the most accurate response. Agents can then automatically call APIs to fulfill a user's request. For example, a manufacturing company might want to develop a generative AI application that automates tracking inventory levels, sales data, supply chain information and can recommend optimal reorder points and quantities to maximize efficiency. As fully managed capabilities, agents for Amazon Bedrock remove the undifferentiated lifting of managing system integration and infrastructure provisioning, allowing developers to use generative AI to its full extent throughout their organization.

How can I connect FMs to my company data sources?

You can securely connect FMs to your company data sources using agents for Amazon Bedrock. With a knowledge base, you can use agents to give FMs in Amazon Bedrock access to additional data that helps the model generate more relevant, context-specific, and accurate responses without continually retraining the FM. Based on user input, agents identify the appropriate knowledge base, retrieve the relevant information, and add the information to the input prompt, giving the model more context information to generate a completion.

What are some use cases for agents for Amazon Bedrock?

Agents for Amazon Bedrock can help you increase productivity, improve your customer service experience, or automate DevOps tasks.

How do agents for Amazon Bedrock help improve developer productivity?

With agents, developers have seamless support for monitoring, encryption, user permissions, and API invocation management without writing custom code. Agents for Amazon Bedrock automate the prompt engineering and orchestration of user-requested tasks. With fully managed agents, you don't have to worry about provisioning or managing infrastructure and can take applications to production faster.

Security and Privacy

Is the content processed by Amazon Bedrock moved outside the AWS Region where I am using Amazon Bedrock?

Any customer content processed by Amazon Bedrock is encrypted and stored at rest in the AWS Region where you are using Amazon Bedrock.

Are user inputs and model outputs made available to third-party model providers?

No. User inputs and model outputs are not shared with third-party model providers.

How can I securely use my data to customize FMs available through Amazon Bedrock?

With Amazon Bedrock, you can privately customize FMs, retaining control over how your data is used and encrypted. Amazon Bedrock makes a separate copy of the base foundational model and trains this private copy of the model. Your data including prompts, information used to supplement a prompt, FM responses, and customized FMs remain in the Region where the API call is processed.

How does Amazon Bedrock ensure my data used in fine-tuning remains private and confidential?

When you use Amazon Bedrock to customize a model, Amazon Bedrock can fine-tune the model for a particular task without having to annotate large volumes of data. Then, Amazon Bedrock makes a separate copy of the base foundation model that is accessible only by you and trains this private copy of the model. None of your content is used to train the original base models. You can configure your Amazon VPC settings to access Amazon Bedrock APIs and provide model fine-tuning data in a secure manner. Your data is encrypted in transit (TLS1.2) and at rest through service-managed keys.

What security and compliance standards does Amazon Bedrock support?

Amazon Bedrock offers several capabilities to support security and privacy requirements and is compatible with common compliance standards including GDPR and HIPAA. As with all AWS services, you get the standard AWS Identity and Access Management (IAM) controls for authentication and AWS CloudTrail for auditing API activity. All your data is encrypted at rest using your own AWS Key Management Service (AWS KMS) keys, which provides full control and visibility into how your data and custom models are being stored and accessed. With PrivateLink, you can pass your data on AWS to Amazon Bedrock exclusively through AWS and never through public internet. Amazon Bedrock can also attach its training instances to your Amazon VPC in order to read from and write data to Amazon S3.

Will AWS and third-party model providers use customer inputs to or outputs from Amazon Bedrock to train Amazon Titan or any third-party models?

No, AWS and the third-party model providers will not use any inputs to or outputs from Bedrock to train Amazon Titan or any third-party models.

SDK

What SDKs are supported for Amazon Bedrock?

Amazon Bedrock supports SDKs for runtime services. iOS and Android SDKs, as well as Java, JS, Python, CLI, .Net, Ruby, PHP, Go, and CPP support both text and speech input.

What SDKs support streaming functionality?

Streaming is supported on all the SDKs.

Billing and Support

How much does Amazon Bedrock cost?

Please see the **Amazon Bedrock Pricing Page** for current pricing information.

What support is provided for Amazon Bedrock?

Depending on your AWS support contract, Amazon Bedrock is supported under Developer Support, Business Support and Enterprise Support plans.

How can I track the input and output tokens?

You can use CloudWatch metrics to track the inputs and output tokens.