Date: 15.09.2025

# AWS Bedrock

## Contents

## 1. Introduction

**Amazon Bedrock** is a fully managed AWS service designed for building, customizing, and deploying generative AI applications securely using the latest foundation models and robust orchestration tools.

## 2. Key Features of AWS Bedrock

### a. Foundation Model Choice

- Access a curated catalog of top foundation models from providers like Anthropic, Meta, Mistral, Stability AI, Cohere, AI21 Labs, DeepSeek, Luma, and Amazon.
- Supports text, image, audio, and multimodal generative AI tasks.
- Easily switch between models via API to adapt to changing needs.

### b. Model Customization

- Customize models with company data using fine-tuning and RAG for enterprise-specific outputs.
- Private data remains encrypted and is never used to train base models, ensuring privacy and compliance.
- Securely manage and deploy custom models for organization-specific use cases.

### c. Security, Privacy, and Responsible AI

- Built-in guardrails to block harmful content and enforce responsible AI practices.
- Automated reasoning reduces errors and hallucinations, supporting strong governance.
- Complies with industry standards: GDPR, HIPAA, SOC, ISO, CSA STAR, FedRAMP High, and more.
- End-to-end data encryption with robust identity and access management.

### d. Agents and Orchestration

- Agents orchestrate multi-step workflows, integrating systems, APIs, and Knowledge Bases.
- Automate workflows to build smart chatbots, assistants, and enterprise bots.
- Serverless, scalable architecture for rapid AI app development and deployment.

### e. Data Automation and Analytics

- Streamlined data handling for documents, images, video, and audio with pre-built automation for faster insights.
- Multimodal support enables analytics and automation across diverse media types.

### f. Ease of Integration and Cost Control

- Easy AWS integration for logging, monitoring, scaling, and security via CloudWatch, CloudTrail, and IAM.
- Flexible pricing with pay-as-you-go and provisioned throughput to optimize costs.
- Cost management tools provide real-time spending insights and forecasting.

Amazon Bedrock empowers enterprises to leverage generative AI with top-tier models, advanced orchestration, built-in security, and seamless integration with AWS infrastructure—all with a focus on privacy, compliance, and responsible AI deployment.

## 3. Core Concepts

**a. Foundation Models (FMs)**

Large pre-trained models powering generative AI tasks like text generation, image creation, and summarization.

- Users choose from various FMs to suit their needs from providers like: AI21 Labs, Amazon, Anthropic, Cohere, DeepSeek, Meta, Mistral AI, Stability AI, TwelveLabs
- FMs provide backend logic for agents, flows, and other components in prompt-based tasks and data processing.
- Pricing depends on the selected model and the operations performed per invocation.

**b. Knowledge Bases (KBs)**

Managed repositories of structured and unstructured data that enhance AI models and responses.

- Store business data, documents, procedures, and more.
- Agents and Flows dynamically query KBs for context-aware, accurate responses & tasks.
- Guardrails and policies ensure safe and reliable content retrieval and generation.
- Pricing depends on the resources required by KBs like storage buckets and web crawlers.

**c. Guardrails**

Configurable safety mechanisms that block or filter harmful or sensitive content in AI workflows.

- Include word filters, topic bans, sensitive data screens, image filters, and contextual rules.
- Can be customized per use case and layered across prompts and responses.
- Applied at workflow, agent, FM, or KB levels for comprehensive safety.
- Evaluate both inputs and outputs to block or mask content as configured.

**d. Flows**

Visual or programmatic orchestration tools in AWS Bedrock to define, connect, and deploy generative AI workflows.

- A Flow is composed of nodes, connections, and configurations.
  - **Node**: A step in the flow, like invoking an FM, querying a KB, or calling a Lambda function.
  - **Connections**: Solid lines pass data between nodes while dotted lines trigger actions based on conditions.
- Features a drag-and-drop builder for easy workflow assembly linking model invocations, real-time data, and external services like AWS Lambda.
- Fully deployable to production, integrating FMs, KBs, Agents, and Guardrails.

**e. Agents**

Orchestrators that manage complex user interactions across FMs and other Bedrock resources.

- An agent comprises:
  - Foundation Model for understanding and response.
  - Instructions defining tasks and logic.
  - Action Groups (APIs, Lambdas) for task execution.
  - Knowledge Bases for context.
  - Prompt Templates for customized prompts.
- Use guardrails for safety that can be deployed standalone service endpoints or within Flows for multi-step AI workflows.