
AWS BEDROCK UPDATES

STEVE MCDOWELL, CHIEF ANALYST
APRIL 23, 2024

CONTEXT

Amazon announced updates to its Bedrock generative AI platform that expands its capabilities while improving the user experience. These enhancements focus on helping developers create AI applications quickly and securely.

Let's take a look at what AWS announced.

BACKGROUND: WHAT IS AWS BEDROCK?

AWS Bedrock is Amazon's platform designed to facilitate the development and scaling of generative AI applications. It provides users with diverse foundation models, including first-party models from Amazon and third-party models from leading AI companies like Meta, Cohere, Anthropic, AI21 Labs, Mistral AI, and Stability AI. This extensive choice allows users to experiment with various LLMs and other foundation models to find the best fit for their applications.

Bedrock offers a range of tools that make building generative AI applications easier and more efficient. The platform supports model evaluation, allowing developers to benchmark and compare models against specific datasets and evaluation metrics. This helps them choose the most suitable models based on accuracy, latency, and cost projections.

BEDROCK UPDATES

Amazon introduced several significant enhancements to Bedrock that improve its generative AI capabilities, offering more model choices, easier customization, and stronger security.

The key enhancements to Bedrock include:

Expanded Model Choice

- **Llama 3 Models:** Amazon Bedrock now includes Meta's Llama 3 model family, providing more options for text generation, summarization, classification, sentiment analysis, and translation. Llama 3 8B is suitable for smaller deployments, while Llama 3 70B excels in content creation and conversational AI.
- **Cohere's Command R and Command R+:** Cohere's enterprise-grade FMs are now supported and designed for long-context tasks like retrieval-augmented generation (RAG) and multi-step automation. Command R+ is optimized for long-context tasks, while Command R suits large-scale production workloads.

New Tools for Building Generative AI Applications

- **Model Evaluation Tool:** This tool simplifies choosing the best foundation model (FM) by enabling benchmarking and comparison against specific datasets and evaluation metrics. Developers can assess models based on accuracy, latency, cost projections, and qualitative factors to ensure the selected model aligns with their project goals.
- **Amazon Titan Image Generator:** This tool allows customers to generate high-quality images from natural language prompts. To ensure transparency, it features built-in watermarking for ethical AI usage and tamper-resistant detection.
- **Custom Model Import:** This feature enables users to import their own custom models, such as those from Amazon SageMaker, into Amazon Bedrock. It seamlessly integrates custom models with Amazon Bedrock's capabilities, allowing users to leverage their existing AI investments.

Enhanced Privacy and Security

- **Guardrails for Amazon Bedrock:** This feature helps customers prevent harmful content and manage sensitive information. It provides content filtering across six categories, including hate, insults, sexual content, violence, misconduct, and prompt attack. Guardrails also offers robust personal identifiable information (PII) detection, allowing users to customize safety and privacy protections for their generative AI applications.

Improved Platform Experience

- **Knowledge Bases:** This feature supports multiple data sources and allows for a fully managed RAG experience. It facilitates the enrichment of responses with additional information and enhances search relevance.
- **Agents for Amazon Bedrock:** Agents allow developers to define specific tasks, workflows, and decision-making processes. They can create custom workflows that align with business needs and automate complex tasks.

ANALYSIS

AWS's enhancements to Amazon Bedrock expand its functionality and make it easier for users to build, customize, and secure generative AI applications. The broader selection of models and improved tools for building applications offer greater flexibility.

Enhanced security features like Guardrails ensure responsible AI practices. These improvements provide a scalable, secure, and customizable platform for generative AI applications.

These enhancements make Amazon Bedrock more versatile, user-friendly, and secure. By expanding model choice, offering customization tools, and reinforcing security measures, Amazon Bedrock becomes a more comprehensive platform for building and scaling generative AI applications.

The improved platform experience, focusing on flexibility and automation, allows users to experiment and deploy AI solutions quickly and efficiently, ultimately driving business value and innovation.

This is a strong set of enhancements and promises to keep Amazon competitive in the generative AI market.



© Copyright 2024 NAND Research.

NAND Research is a registered trademark of NAND Research LLC, All Rights Reserved.

This document may not be reproduced, distributed, or modified, in physical or electronic form, without the express written consent of NAND Research. Questions about licensing or use of this document should be directed to info@nand-research.com.

The information contained within this document was believed by NAND Research to be reliable and is provided for informational purposes only. The content may contain technical inaccuracies, omissions, or typographical errors. This document reflects the opinions of NAND Research, which is subject to change. NAND Research does not warranty or otherwise guarantee the accuracy of the information contained within.

NAND Research is a technology-focused industry analyst firm providing research, customer content, market and competitive intelligence, and custom deliverables to technology vendors, investors, and end-customer IT organizations.

Contact NAND Research via email at info@nand-research.com or visit our website at nand-research.com.