

AWS BED ROCK

EMBEDDINGS





Introduction

AWS

Amazon Web Services (AWS) is the world's most comprehensive and widely adopted cloud platform, offering over 200 fully featured services from data centers globally. It enables organizations to build, deploy, and scale applications with high availability and security. AWS serves diverse industries with solutions for computing, storage, machine learning, and more.

Services by AWS

Compute: Amazon EC2, AWS Lambda, Elastic Beanstalk

Storage: Amazon S3, EBS, Glacier

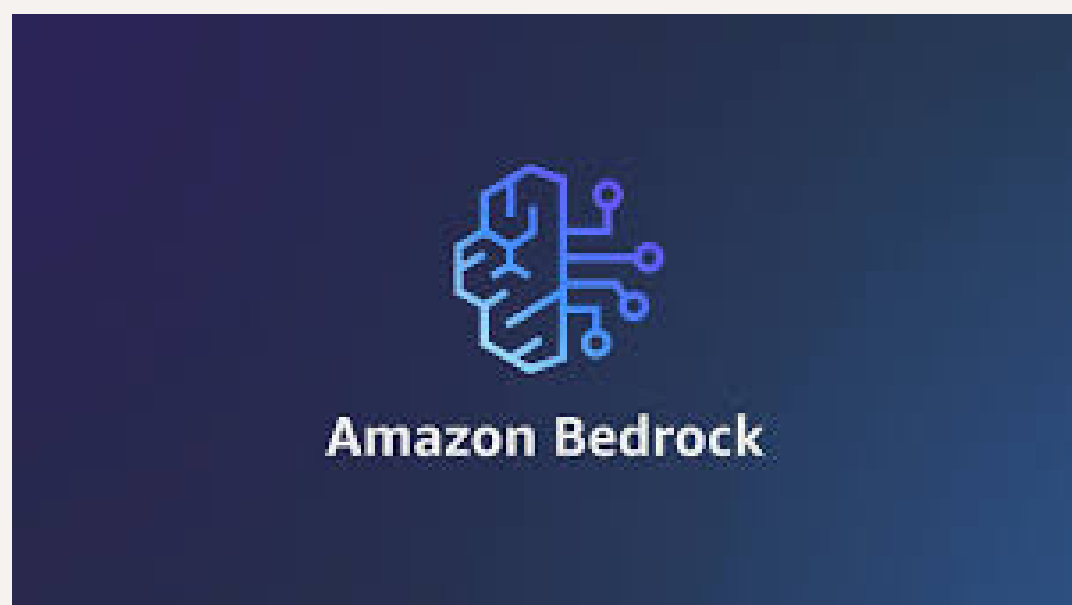
Databases: Amazon RDS, DynamoDB, Aurora

AI/ML: Amazon SageMaker, Bedrock, Rekognition

Serverless: AWS Lambda, Fargate

Monitoring: CloudWatch, AWS X-Ray

Services for **Gen AI Applications**





Foundation Models Explored





Amazon titan text premeir

Maximum Tokens: 3,072 tokens

Applications: Summarization, classification, Q&A, and information extraction.

Inference Parameters:

Temperature: Controls randomness in output (range: 0.0 to 1.0).

Strength: General-purpose model with high accuracy across text-based tasks.

Anthropic Claude 3 opus

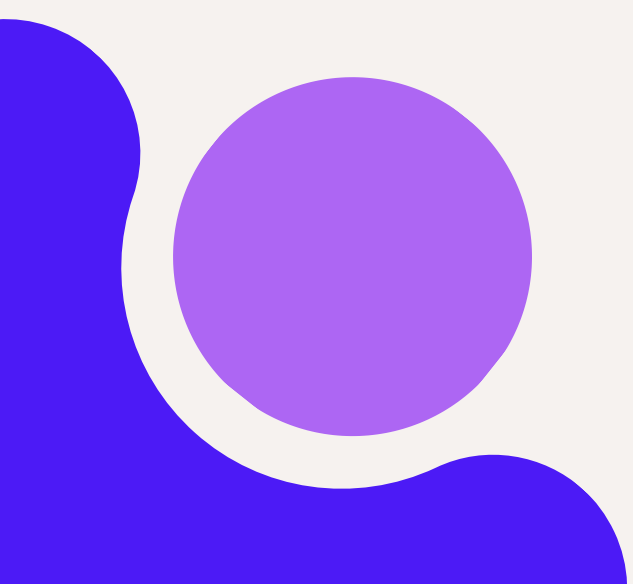
- Maximum Token Context: 200,000 tokens
- Applications: Long-form text analysis, multi-step problem-solving, and coding.
- Inference Parameters:
 - Temperature (adjusts randomness).
 - Stop Sequences (stops text generation at specific markers).
- Strength: Handles complex, large-context tasks with industry-leading efficiency.

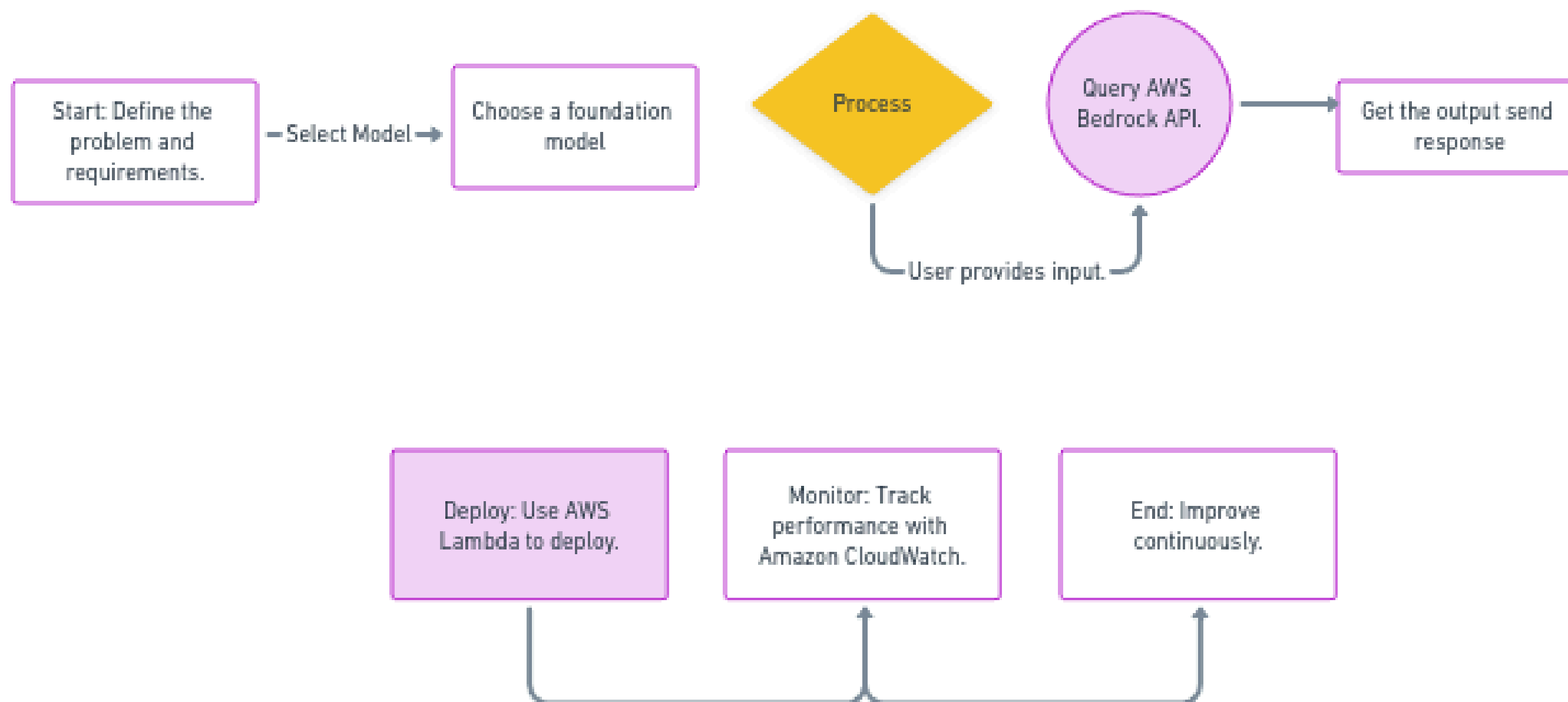
Mistral 7B

- Parameters: 7 billion
- Architecture: Grouped Query Attention (GQA) and Sliding Window Attention (SWA) for faster inference.
- Applications: Text summarization, multilingual translation, and NLP tasks.
- Strength: Open-source, efficient, and highly accessible for custom use cases.



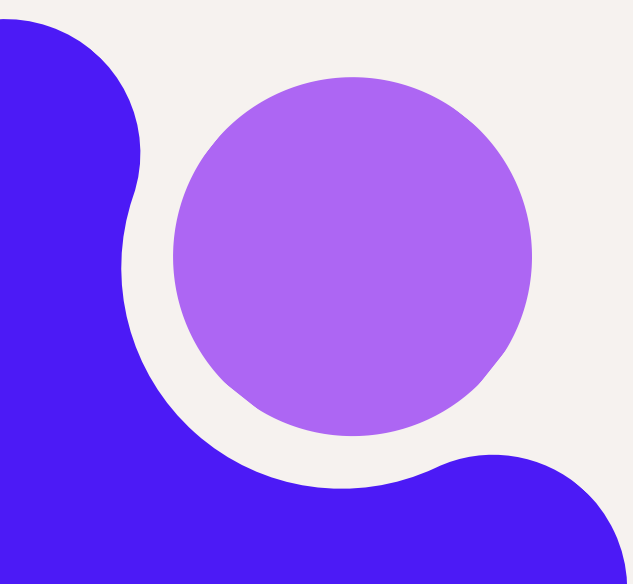
Working of **AWS bedrock**







Consider this Scenario



Health care Chatbot

A hospital wants a chatbot to assist patients and healthcare staff with:

- Providing health-related advice.
- Scheduling doctor appointments.
- Offering information about medications and treatments.



Work flow

1. User: "What are the symptoms of diabetes?"
2. Chatbot uses Bedrock to query the Amazon Titan model.
3. Retrieves information from medical databases and clinical guidelines.
4. Responds: "Common symptoms of diabetes include increased thirst, frequent urination, and fatigue. Would you like to schedule a consultation?"



Conclusion

- Explored Amazon Bedrock and its foundation models like Amazon Titan, Anthropic Claude, and Mistral AI.
- Learned to build workflows integrating Bedrock API, Lambda functions, and API Gateway.
- Discovered techniques for model customization using fine-tuning and Retrieval Augmented Generation (RAG).
- Gained knowledge of deploying applications with AWS Lambda and monitoring them using Amazon CloudWatch.

Thank you

AWS empowers you to **build** faster, **scale** smarter, and **innovate** without limits.