Route 53

VPC

Public Subnet

App Load Balancer

Internet Gateway

NAT Gateway

Private Subnet

App Layer

EKS
Model API

Async Task

Pre/Post Processing

SQS
Async Jobs

Invoke

NLP

ML Services

SageMaker
Train/Inference

Comprehend
NLP

Data Layer

DynamoDB
Metadata

S3
Data Lake

IAM

IAM

KMS

KMS

Metrics/Logs

Metrics/Logs

IAM

IAM

KMS

KMS

Metrics/Logs

Metrics/Logs

IAM

IAM

KMS

KMS

Security & Monitoring

IAM Roles

KMS

CloudWatch

AWS AI/ML Application Architecture
This architecture uses S3 for data storage, SageMaker for model training/inference, EKS for scalable app serving, and integrates security, monitoring, and queueing for a robust AI pipeline.