



# AWSの主要NLP/AI関連サービス

## 技術的な強み



### スケーラブルな計算リソース (GPU/TPU)

P4d/P5 インスタンス (A100/H100 GPU) をオンデマンドで利用。大規模モデルの学習・推論を低コストで実現。推論は Inf1/Inf 2、学習は Trn1/Trn2/Trn 3 で最大 50% コスト削減。AWS 独自設計チップで GPU に依存しない高効率 AI 基盤を構築



### 最新の基盤モデルへのアクセス

Claude Sonnet 4.6/ Claude Opus 4.6 ・ Amazon Nova 2 等を API 一本で切り替え。プロビジョンドスループットで安定推論



### RAG システム構築支援 / MLOps ・ LLMops

Amazon Bedrock Knowledge Bases + OpenSearch で RAG を即構築。SageMaker MLflow で実験追跡・モデルレジストリを一元管理

## 主要 NLP / AI サービス



### Amazon Bedrock

Claude ・ Nova 等の基盤モデルを API 経由で即時利用。RAG ・ エージェント構築を完全マネージドで提供



### Amazon SageMaker

モデル開発・トレーニング・デプロイを一元管理。GPU/TPU クラスタを自動スケールで活用可能



### AWS Tranium

学習専用カスタムシリコン。EC2 Trn1/Trn2 /Trn3 インスタンスで大規模 LLM のファインチューニングを最大 50% 低コストで実現



### AWS Inferentia

推論専用カスタムシリコン。EC2 Inf1/Inf2 インスタンスで高スループット・低レイテンシ推論をコスト効率よく実現



### Claude Code on Bedrock

Anthropic Claude によるターミナル型 AI コーディングエージェント。コードベース全体を理解し、実装・デバッグ・リファクタリングを自律実行

## 具体的なユースケース・事例



### 学術研究での活用

論文コーパス解析・文献マイニング・多言語要約。Amazon Bedrock で最新 LLM を即座に研究へ適用



### 日本企業での NLP 導入事例

カスタマーサポート自動化・社内文書検索・契約書レビュー。Amazon Bedrock で日本語対応 RAG を構築



### 研究から本番環境への展開

Amazon SageMaker Pipelines で実験管理→ CI/CD →本番デプロイを自動化。MLOps/LLMOps で継続的改善を実現



AWS の生成 AI



お問い合わせフォーム



Build on Trainium プログラム



AWS Cloud Credit for Research

## 今すぐ始める - 特典オファー

- ✓ 無料技術相談
- ✓ 無料 アーキテクチャレビュー
- ✓ AWS Cloud Credit for Research (研究クレジットプログラム)