



# 今変革を迎える生成系AI

ビルダーの為の生成系AI開発ツール

黒川 亮 Rio Kurokawa

プリンシパル AI/ML 事業開発マネジャー



人手不足

より  
大きな成果



接点の  
多様化

一貫した  
価値提供



持続  
可能性

ビジネス  
の拡大



**Solution**

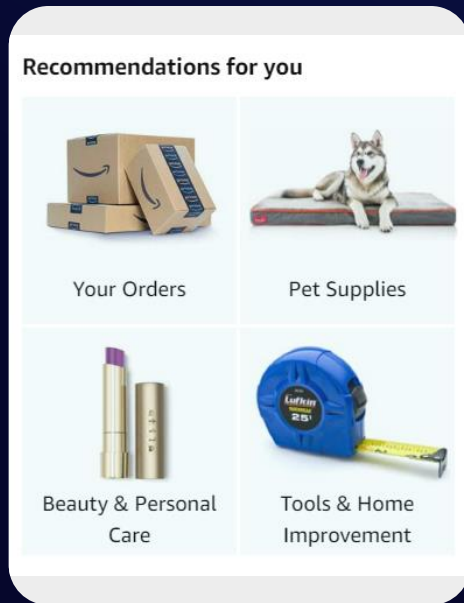
お客様との対話で生まれる AWS サービス



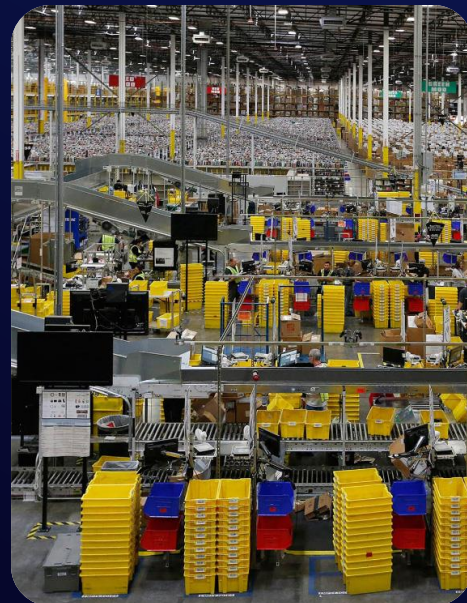
# 機械学習は転換点に

**主要要素:** 計算能力の拡大 | データの増大 | モデルの高度化

# 20 年以上にわたる Amazon の機械学習によるイノベーション



毎分 4,000 個  
Amazon.com での取引



毎日 160 万個  
のパッケージ発送



毎週 10 億件を超える Alexa  
でのインタラクション



2016 年 12 月 7 日～  
Prime Air ドローン配達

# AWSのミッション：全てのお客様に機械学習をお届けする

## AIサービス

医療AI			産業AI			異常検知		コード&DevOps				
Amazon HealthLake	Amazon Transcribe Medical	Amazon Comprehend Medical	AWS Panorama + Appliance	Amazon Monitron	Amazon Lookout for Equipment	Amazon Lookout for Vision	Amazon Lookout for Metrics	Amazon DevOps Guru	Amazon CodeGuru	Amazon CodeWhisperer		
画像・動画	音声		テキスト			検索	チャットボット	パーソナライゼーション	需要予測	不正検知	コンタクト・センター	強化学習
Amazon Rekognition	Amazon Polly	Amazon Transcribe	Amazon Comprehend	Amazon Translate	Amazon Textract	Amazon Kendra	Amazon Lex	Amazon Personalize	Amazon Forecast	Amazon Fraud Detector	Contact Lens Voice ID	AWS DeepRacer

## 機械学習サービス

	ラベリング	SageMaker Studio IDE (統合開発環境)												目検確認	
		データ準備	特徴量ストア	Auto ML	Spark/R	バイアス検出	ノートブック	アルゴリズム選択	モデル学習	パラメータ最適化	デバッグプロファイル	本番デプロイ	管理モニター		CI/CD
		SageMaker JumpStart													

## フレームワーク & インフラ

				Deep Learning AMIs & Containers	GPUs & CPUs	Elastic Inference	Trainium	Inferentia	FPGA





# 100,000 社を超えるお客様が AWS で機械学習に取り組んでいる



# 生成系 AI ( Generative AI )

会話、ストーリー、画像、動画、音楽など、  
新しいコンテンツやアイデアを創造

一般に**基盤モデル (FM: ファウンデーションモデル)**  
と呼ばれる膨大なデータに基づいて事前に  
トレーニングされた大規模モデルを搭載



# 基盤モデルの種類

インプット



基盤モデル



アウトプット

“心臓の健康と歩行の相関性について  
記事をまとめて”

Text-to-text

簡単な自然言語プロンプトから、複数アプリケーションのためにテキスト生成

“健康な心臓を保つには  
一日10,000歩のウォーキング  
が有効です”

“ハンド・ソープ”

Text-to-embeddings

検索や文書間の類似点検索などの用途向けに、テキストの数値表現を生成

以下を数値で表現  
“ハンドソープ・リフィル、  
ハンドソープ・ディスペンサー、  
ハンドソープ抗菌”

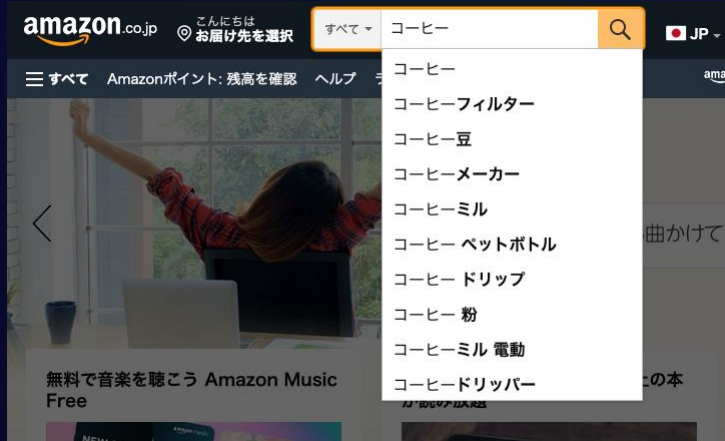
“宇宙飛行士が馬に乗っている写真”

Multimodal

自然言語プロンプトから、画像生成・編集



# 生成系 AI による Amazon のイノベーション



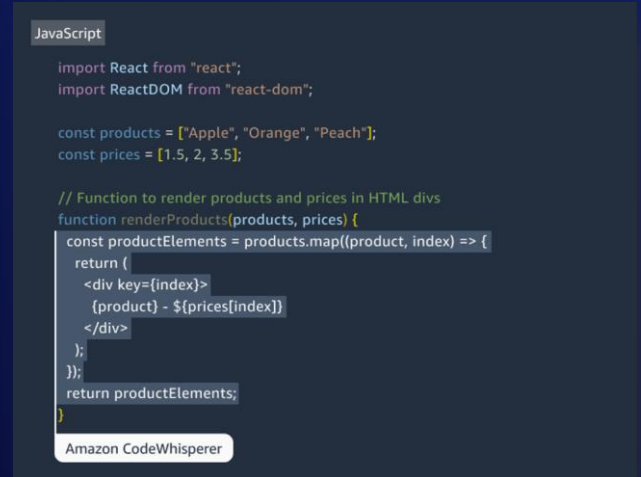
## Amazon の検索機能

基盤モデルを導入し  
検索結果を改善



## Alexa 教師モデル

最小限の入力情報から  
多言語による応答を実現



## Amazon CodeWhisperer

コメントとコードから  
推奨コードを自動生成

NEW



# Amazon Bedrock

Limited Preview

お客様の業務課題を解決する  
基盤モデルによる生成系 AI アプリを  
サーバレスで素早く開発可能に

# 複数の基盤モデルから用途に最適なものを選択

Amazon が提供

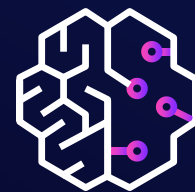


Titan Text

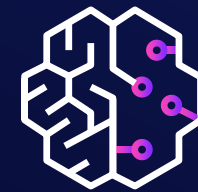


Titan  
Embeddings

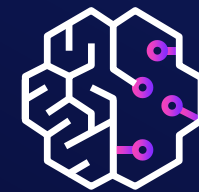
最先端スタートアップ企業が提供  
AI21 Labs, Anthropic, Stability AI



Jurassic-2



Claude



Stable  
Diffusion

# Amazon Titan

Amazon の 20 年以上に及ぶ機械学習の経験に基づいて構築



## Titan Text

自然言語処理 (NLP)



## Titan Embeddings

検索・レコメンデーション

Amazon Titan Text では、要約やテキスト生成などの言語に関するタスクを自動化

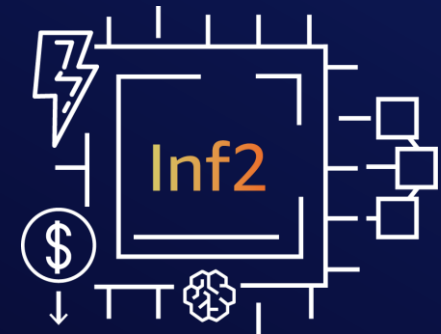
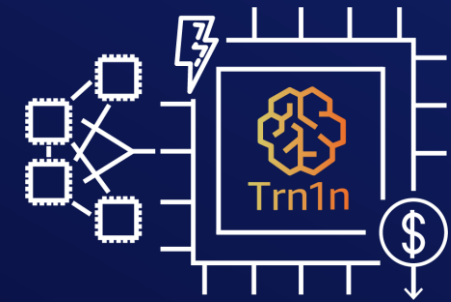
Amazon Titan Embeddings では、レコメンデーションの精度向上を実現

不適切・有害なコンテンツを制限することで、生成系 AI の責任ある利用をサポート

NEW

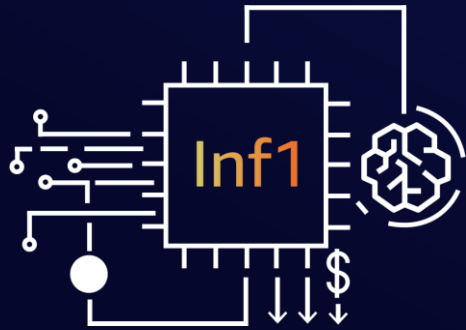
# Amazon EC2 Trn1n インスタンス Amazon EC2 Inf2 インスタンス

トレーニングと推論に  
高いコストパフォーマンスを発揮する  
専用設計の機械学習アクセラレータを搭載



# AWS の機械学習特化型アクセラレータ

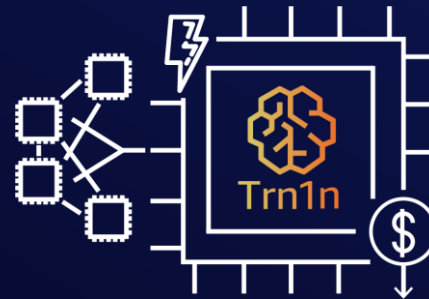
## AWS Inferentia



ディープラーニングモデルの  
推論処理を最も安価に実現

推論処理あたりのコストを  
最大 **70%** 削減\*

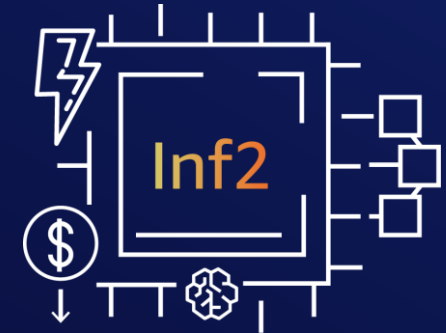
## AWS Trainium



大規模言語モデルや  
拡散モデルのトレーニングを  
高コスト効率・高性能で実行

トレーニングコストを  
最大 **50%** 削減\*

## AWS Inferentia2



大規模言語モデルや  
拡散モデルの推論を  
高性能・低コストで実行

最大 **40%** 高い  
コストパフォーマンス\*

A hand is shown placing a small flag on the peak of a mountain. A winding path leads up the mountain. The background is a dark blue gradient.

何を実現するのか？

競争力  
向上

新規事業  
の開拓

社会への  
貢献



# 自社の業務に最適な基盤モデルは何か？



## 広告

マーケティング支援、サムネイル自動生成



## 金融サービス

バックオフィス業務効率化、パーソナルアシスタント、不正検知学習用データ生成



## ヘルスケア

タンパク質構造過程、医薬品開発、医療画像の高解像度化



## 小売

仮想ファッションモデル、商品画像・商品説明文生成、3Dプロダクトデザイン



## メディア、エンターテインメント

ビデオゲーム生成、コンテンツの高解像度化、背景・アセット生成、イラスト着色



## エネルギー

プラント設計書の確認、予知保全学習用データ生成



## モビリティ

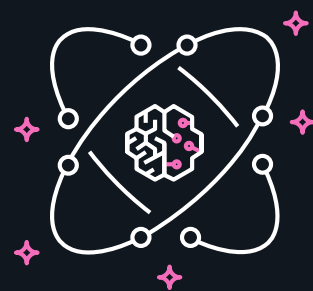
車内音声アシスタント、燃費向上のための設計部品



## ハードウェア事業

半導体デザイン

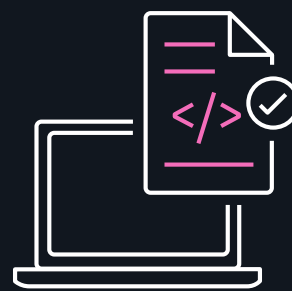
# 基盤モデルにおける課題



プロプライエタリ  
モデル使用

---

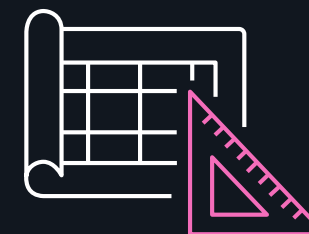
セキュリティ、データ管理、  
コスト、カスタマイズに課題



公開済み  
モデル・チューニング

---

運用に辿り着くまでに  
相当な分類作業要



スクラッチ  
モデル開発

---

コスト、開発期間要  
業務・機械学習  
専門スキル必須

# Why AWS ?



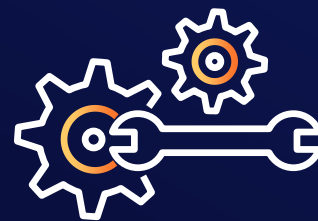
素早い生成系 AI  
アプリケーション開発

API を通じて  
基盤モデルを利用

インフラ管理が不要



利用用途に合った  
基盤モデルを  
選択可能



お客様のデータを  
外部に出すことなく  
基盤モデルを  
カスタマイズ可能



AWS の包括的な  
セキュリティ機能で  
データ保護を強化



AWS ツールを使い、  
信頼性高く、  
拡張性がある  
セキュアな  
AI アプリケーション  
のデプロイが可能

# 私達がお手伝いします

セキュリティ

ナレッジ

AWS認定  
人材育成

継続的な  
イノベーション

導入  
事例・実績

Amazon  
自身の経験

プログラム

運用実績

エコシステム

お客様  
支援の実績

AWS  
コミュニティ

グローバル  
インフラ  
ストラクチャ

# AWSで作る自社用基盤モデル

## AIサービス

医療AI			産業AI			異常検知	コード&DevOps					
Amazon HealthLake	Amazon Transcribe Medical	Amazon Comprehend Medical	AWS Panorama + Appliance	Amazon Monitron	Amazon Lookout for Equipment	Amazon Lookout for Vision	Amazon Lookout for Metrics	Amazon DevOps Guru	Amazon CodeGuru	Amazon CodeWhisperer		
画像・動画	音声	テキスト			検索	チャットボット	パーソナライゼーション	需要予測	不正検知	コンタクト・センター	強化学習	
Amazon Rekognition	Amazon Polly	Amazon Transcribe	Amazon Comprehend	Amazon Translate	Amazon Textract	Amazon Kendra	Amazon Lex	Amazon Personalize	Amazon Forecast	Amazon Fraud Detector	Contact Lens Voice ID	AWS DeepRacer

## 機械学習サービス



## フレームワーク & インフラ

				Deep Learning APIs & Containers	GPUs & CPUs	Elastic Inference	Trainium	Inferentia	FPGA

# AWSで簡単に作る生成系AIモデル

## AIサービス

医療AI			産業AI			異常検知		コード&DevOps				
Amazon HealthLake	Amazon Transcribe Medical	Amazon Comprehend Medical	AWS Panorama + Appliance	Amazon Monitron	Amazon Lookout for Equipment	Amazon Lookout for Vision	Amazon Lookout for Metrics	Amazon DevOps Guru	Amazon CodeGuru	Amazon CodeWhisperer		
画像・動画	音声	テキスト			検索	チャットボット	パーソナライゼーション	需要予測	不正検知	コンタクト・センター	強化学習	
Amazon Rekognition	Amazon Polly	Amazon Transcribe	Amazon Comprehend	Amazon Translate	Amazon Textract	Amazon Kendra	Amazon Lex	Amazon Personalize	Amazon Forecast	Amazon Fraud Detector	Contact Lens Voice ID	AWS DeepRacer

## 機械学習サービス

	ラベリング	SageMaker Studio IDE (統合開発環境)												目検確認
		データ準備	特徴量ストア	Auto ML	Spark/R	バイアス検出	ノートブック	アルゴリズム選択	モデル学習	パラメータ最適化	デバッグプロファイル	本番デプロイ	管理モニター	
		<b>SageMaker JumpStart</b>												

## フレームワーク & インフラ

				Deep Learning AMIs & Containers	GPUs & CPUs	Elastic Inference	Trainium	Inferentia	FPGA



明日への第一歩を踏み出しましょう

