



# Generative AI (生成系AI) on AWS

インダストリー事業開発部

山口 賢人



# 機械学習は転換点に

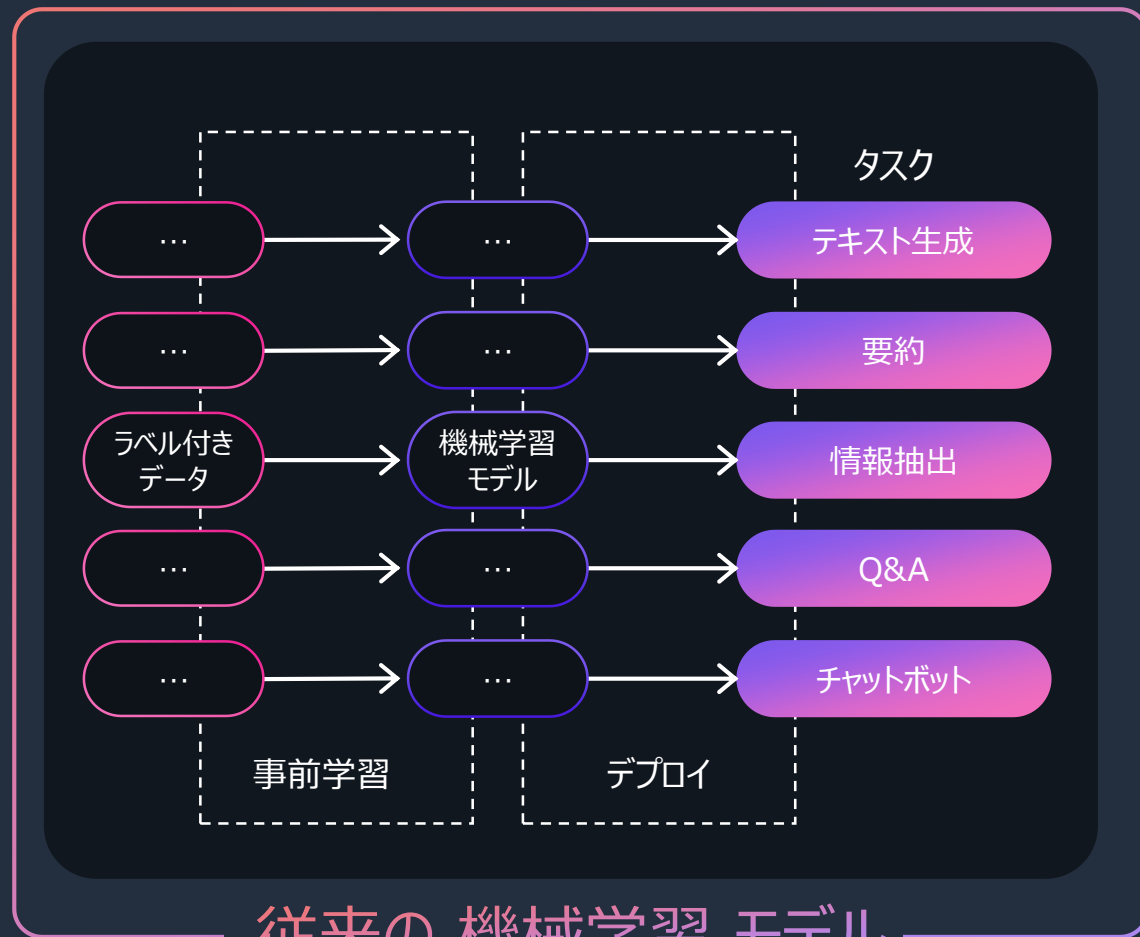
**主要要素:** 計算能力の拡大 | データの増大 | モデルの高度化

# 生成系 AI ( Generative AI )

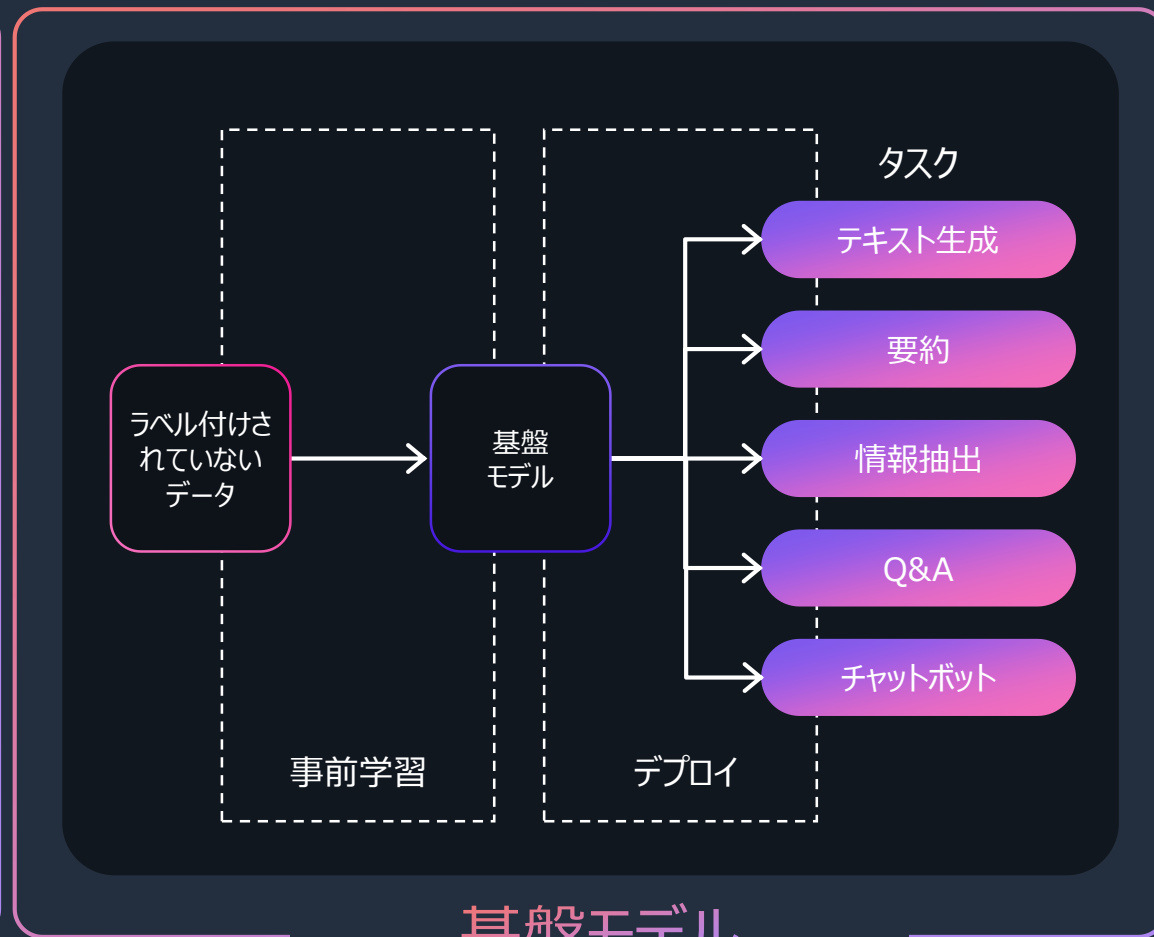
会話、ストーリー、画像、動画、音楽など、  
新しいコンテンツやアイデアを創造

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

# なぜ基盤モデルなのか



従来の機械学習モデル



基盤モデル

# 基盤モデルの例

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

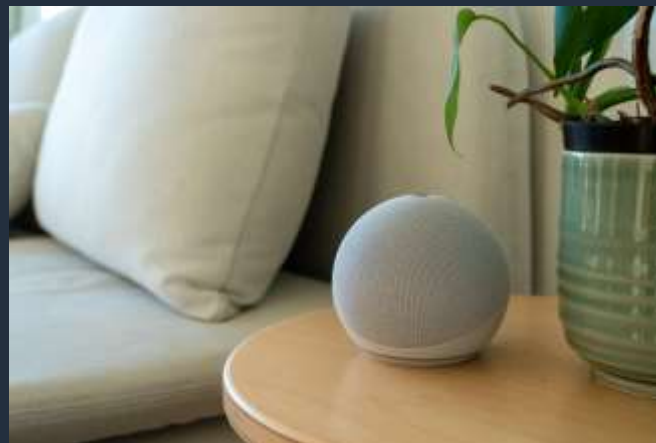


# 生成系 AI による Amazon のイノベーション



## 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

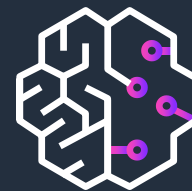


Titan Text



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**言語に関わるタスクを自動化**

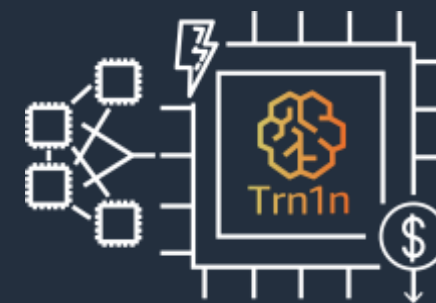
Amazon Titan Embeddings では、**レコメンデーションの精度向上**を実現

**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



# Amazon SageMaker

基盤モデルの大規模開発にも対応



## 人間参加型AI (Human-in-the-loop)

データセット品質を高め、データラベリングのオーバーヘッドと時間を削減



## マネージド・サービス

マネージド・インフラによるモデル トレーニング環境の完全制御



## 分散型トレーニング

分散型トレーニング完了を最大40%まで高速化



## 推論

学習モデルのリアルタイム・非同期推論を最高のコストパフォーマンスで実現



## MLOps

機械学習ライフサイクル全般に渡るプロセスを自動化、標準化

# Amazon SageMaker JumpStart

既存の基盤モデルに  
アクセスして開発・  
運用を始める



## Getting started with Amazon SageMaker JumpStart

Amazon SageMaker JumpStart is a machine learning (ML) hub that can help you accelerate your ML journey. Explore how you can get started with built-in algorithms with pretrained models from model hubs, pretrained foundation models, and prebuilt solutions to solve common use cases. To get started, see documentation or example notebooks that you can quickly execute.

Reset Filters

Q foundation models X

Product Type

Sort By

Popularity

Text Tasks

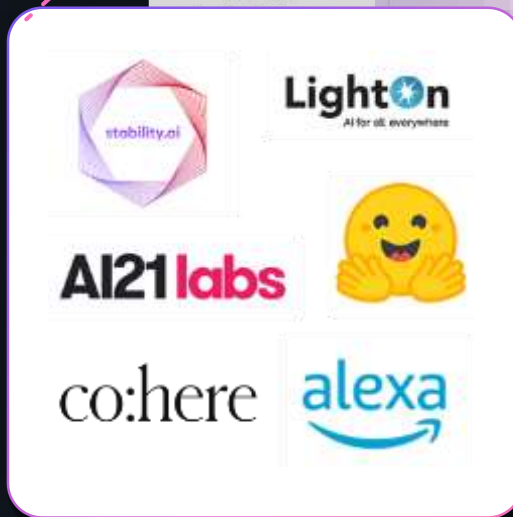
- End-to-end Solution
- Text Classification
- Text Embedding
- Text Generation
- Text Summarization
- Named Entity Recognition
- Question Answering

FOUNDATION MODEL PREVIEW  
Text Generation  
Proprietary Models  
Various Providers  
Models from AI21 Labs, Cohere, and LightOn in preview. Sign-up for preview with JumpStart in us-east-1 or eu-west-1 SageMaker Console.

FOUNDATION MODEL FEATURED  
Text to Image  
stability.ai  
Stable Diffusion 2  
Stabilityai  
Model ID: model-txt2img-stabilityai-stable-diffusion-v2. This is a text-to-image model from Stability AI and downloaded from HuggingFace. It takes a textual description as

FOUNDATION MODEL FEATURED  
Text Generation  
alexa  
AlexaTM (20b)  
Pytorch  
Model ID: pytorch-textgeneration1-alexa20b. AlexaTM 20B is a multitask, multilingual, large-scale sequence-to-sequence (seq2seq) model, trained on a mixture of Common Crawl

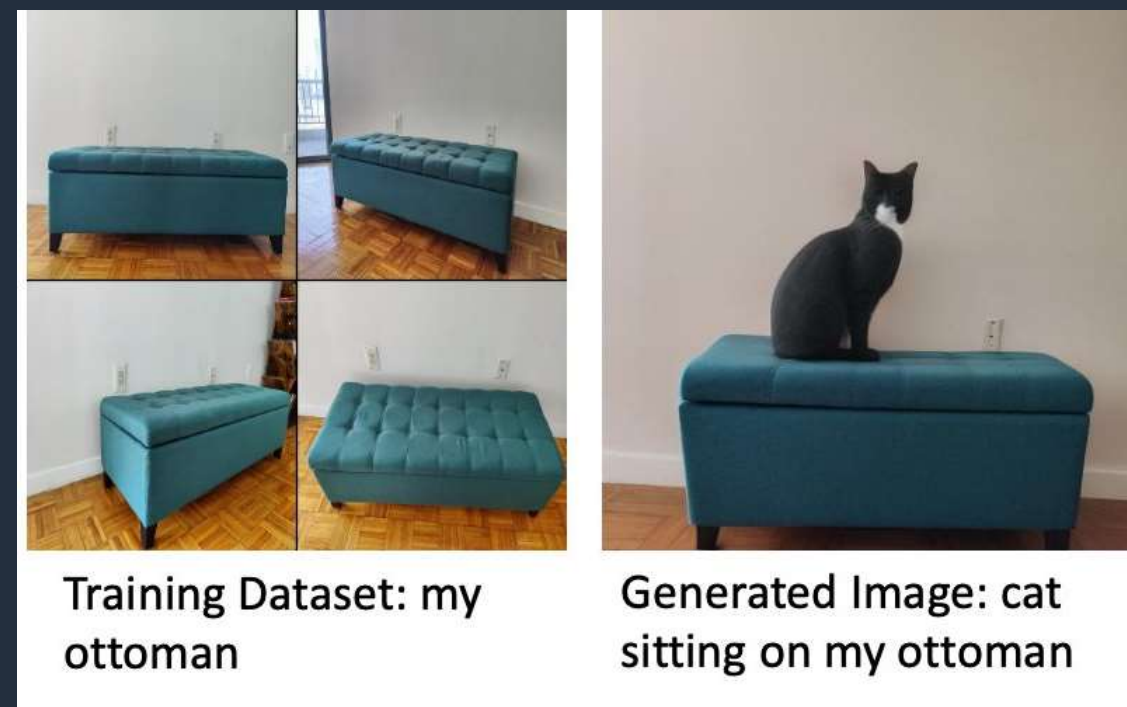
FOUNDATION MODEL FEATURED  
Text Generation  
Bloom 1b7  
Huggingface  
Model ID: huggingface-textgeneration-bloom-1b7. This is a Text Generation model built upon a Transformer model from Hugging Face. It takes a text string as input and predicts next words in the sequence. This model has BigScience Responsible AI License v1.0. Please read the [terms] (<https://huggingface.co/spaces/b>)



# Stable Diffusion on SageMaker JumpStart

Stable Diffusion をクラウド上でファインチューニングし API としてデプロイ

- ファインチューニングは数枚の画像で対象物を学習させることが可能
- ファインチューニングあたりおよそ 20分 で実行可能



[たった数枚の画像で Stable Diffusion をファインチューニングできる効率的な Amazon SageMaker JumpStart の使い方](#)

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

# M&E 業界での活用例

## • Image & Video

- 超解像 (アップスケーリング) / フレーム補完 / レストレーション
- アセット生成 (2D から 3D へ、キャラクター生成、マーケティングマテリアル生成など)
- ローカライゼーションや画像ベースでのビデオ編集
- “Render from rough”: スケッチからの絵コンテ生成、テクスチャのない3Dモデルからのレンダリング

## • Audio

- 音楽生成や音楽制作支援
- AI ボイスによる音声生成 (オーディオブックやポッドキャストコンテンツ)

## • Text

- 読書支援、執筆支援
- スクリプト要約、EPG メタ作成
- スポーツ統計情報からのストーリー作成
- 高度なアセット検索

# AWS が提供できる価値



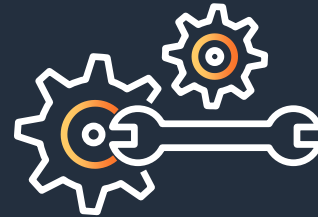
素早い生成系 AI  
アプリケーション開発

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



# 参考

## Generative AI（生成系 AI）へのアプローチ - 今、見直される AI 活用と変革を迎える生成系 AI 戦略

### アジェンダ

- 転換点を迎える AI/ML ユースケース
- 今変革を迎える生成系 AI
- AWS で作る自社用基盤モデル
- AWS で簡単に作る生成系 AI モデル（デモ）



# Thank you!