



# 創薬における機械学習の最新動向 – 生成系 AI がもたらすイノベーション

森下 裕介

アマゾンウェブサービスジャパン合同会社

ソリューションアーキテクト

2023年10月25日

# 自己紹介

森下 裕介 / Morishita, Yusuke

yusmoris@amazon.com

アマゾン ウェブ サービスジャパン合同会社  
エンタープライズ技術本部 ソリューションアーキテクト

好きな AWS サービス: Amazon Bedrock



Amazon SageMaker



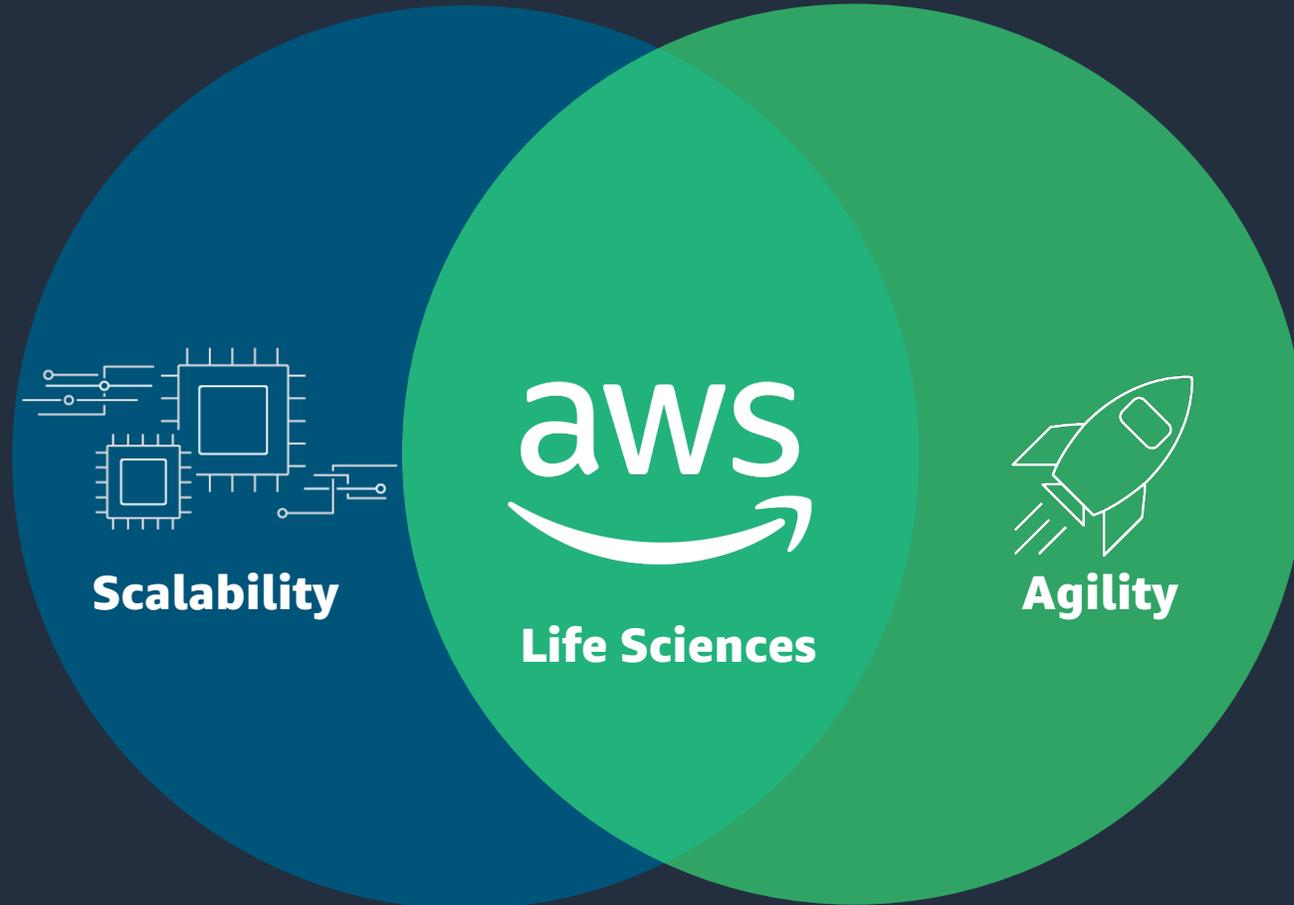
# 本日のセッションのゴール

- 創薬研究領域における機械学習活用について AWS がもたらす価値や取り組みを理解していただく
- 生成系AIの創薬研究領域でのユースケースと、AWSがこの領域でどのように活用いただけるかイメージを持っていただく

# 創薬領域における機械学習 on AWS



# 創薬における AWS の提供する価値



**AWS のミッション：全てのお客様に機械学習をお届けする**



ライフサイエンスの発展を  
支えた結果、**上位 10 社中 9社** が  
AI / 機械学習 に AWS を  
ご活用いただいています

 Pfizer

 BAYER

 moderna



 PHILIPS

 Medtronic

 teva

 Baxter

 Insilico  
Medicine

 GILEAD

 aidoc

 AstraZeneca

 aizon

 Roche

 Amazon  
SageMaker

 Numerate

 NOVARTIS

 Bristol Myers Squibb

 janssen  
PHARMACEUTICAL COMPANY OF  
JOHNSON & JOHNSON

2002

2017

2023

# 創薬研究における機械学習の活用領域例



Research



Genomics /  
Precision Medicine



Drug Design



Image Analysis



Knowledge Graph  
/ GNN



Quantum  
Computing



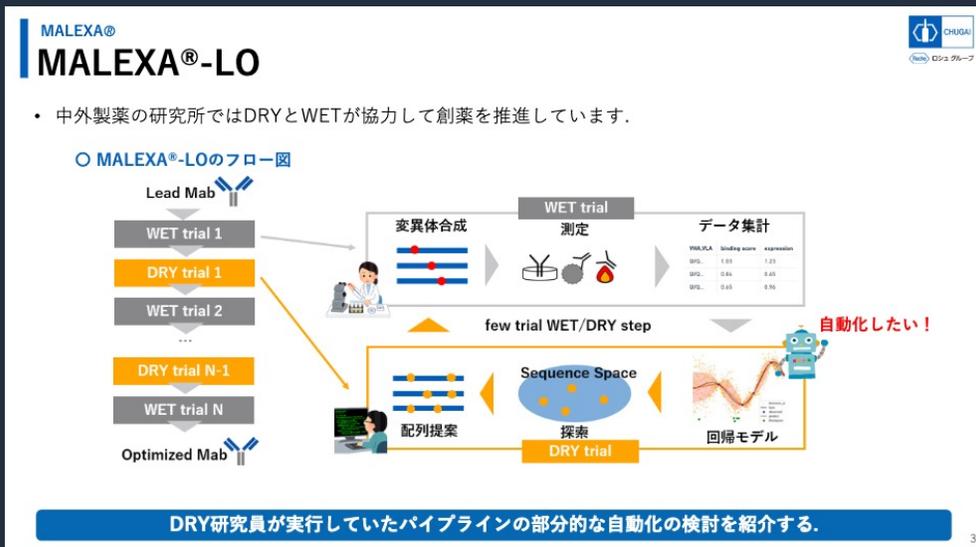
Connected Lab



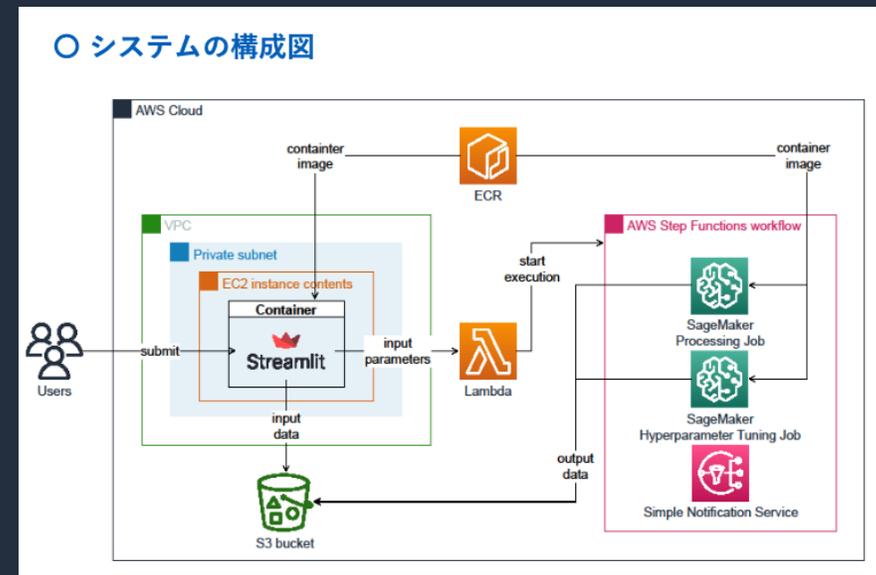
# AWS 事例：中外製薬様

## 機械学習による抗体創薬支援技術 MALEXA における MLOps の実践

Drug Design



- 機械学習による 抗体医薬リード最適化技術 “MALEXA-LO”
- GUI でのデータ入力に際し自動で機械学習パイプラインを実行、結果を共有する仕組みを内製で構築
- 計算負荷の高い機械学習の処理に対して自動でリソースをスケーリング



# AWS 事例：アストラゼネカ様

## 前臨床試験のための腎臓画像解析

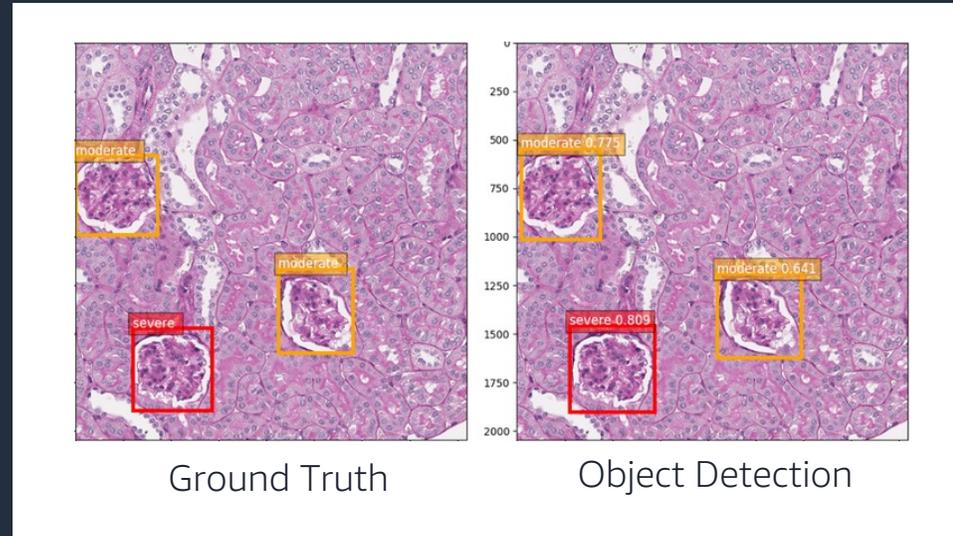


Image Analysis



### ビジネス課題

- 臨床試験前の腎臓損傷評価（目視作業）の効率化
- 腎臓内の糸球体(glomeruli)を事前に定義した損傷クラスに分類



### ML活用方法

- Amazon SageMaker Ground Truth で学習データを収集・ラベリング
- カスタムモデルで糸球体を抽出→ラベル付与

### 得られた成果

- 科学的知見を得るまでのプロセスを効率化
- 人手による分析の精度改善
- 作業コスト削減

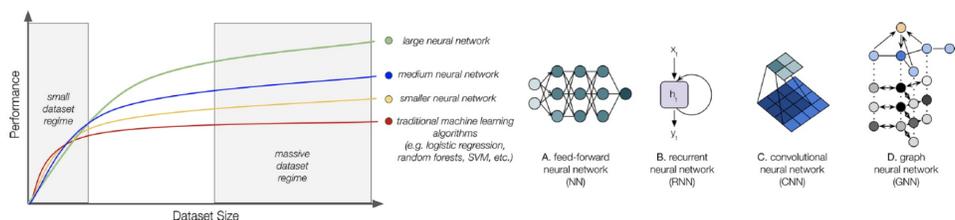


# AWS 事例：日本たばこ産業様

## グラフニューラルネットワーク (GNN) による物性予測



### 創薬における機械学習



適切な機械学習手法は  
データサイズに依存

Life Science分野における  
様々な機械学習手法

↑データサイズ = ↑計算時間

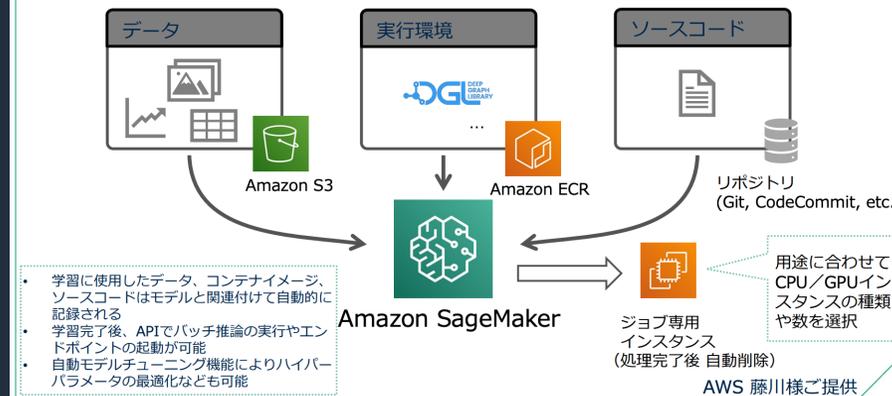
本発表ではGNNに注目

J. Med. Chem. 2020, 63, 16, 8705–8722

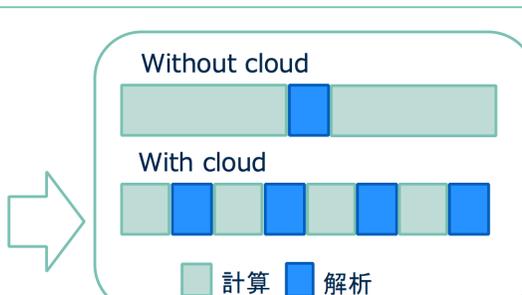
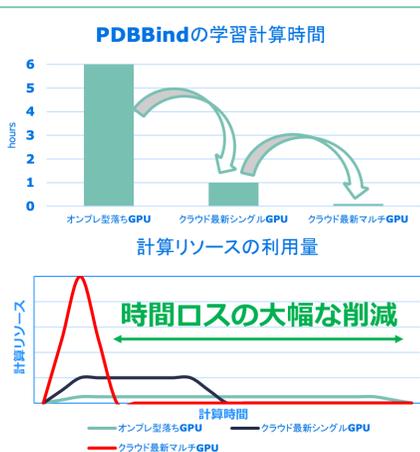
### Amazon SageMaker の学習ジョブ



必要なリソースを自由に組み合わせて学習ジョブを実行可能



### AWSを利用した DGL LifeSci



- ⌘ 計算サイクルの高速化
- ⌘ 環境構築までのAWSサポート
- ⌘ **true human in the loop**

- GNN (グラフニューラルネットワーク) を創薬に活用。リガンド結合親和性予測を実施
- ライフサイエンス領域向けの GNN ライブラリと AWS の機械学習プラットフォームを利用
- クラウド GPU を活用した機械学習で大幅な効率化

## ここまでのまとめ

クラウドのスケールABILITY・俊敏性を活用した  
特定のタスクに特化した  
高精度の機械学習モデルの開発・運用



近年 **生成系 AI** が劇的に進化

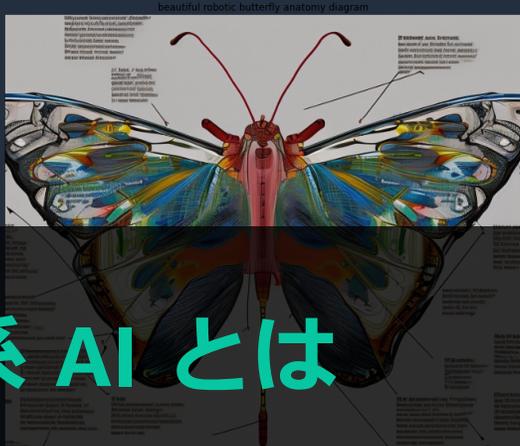
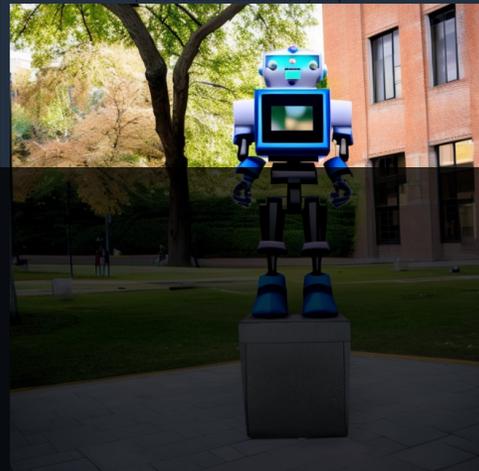
# 生成系 AI とは



A golden retriever wearing glasses and a hat in a portrait painting



photo of a statue of a robot in university courtyard



beautiful robotic butterfly anatomy diagram



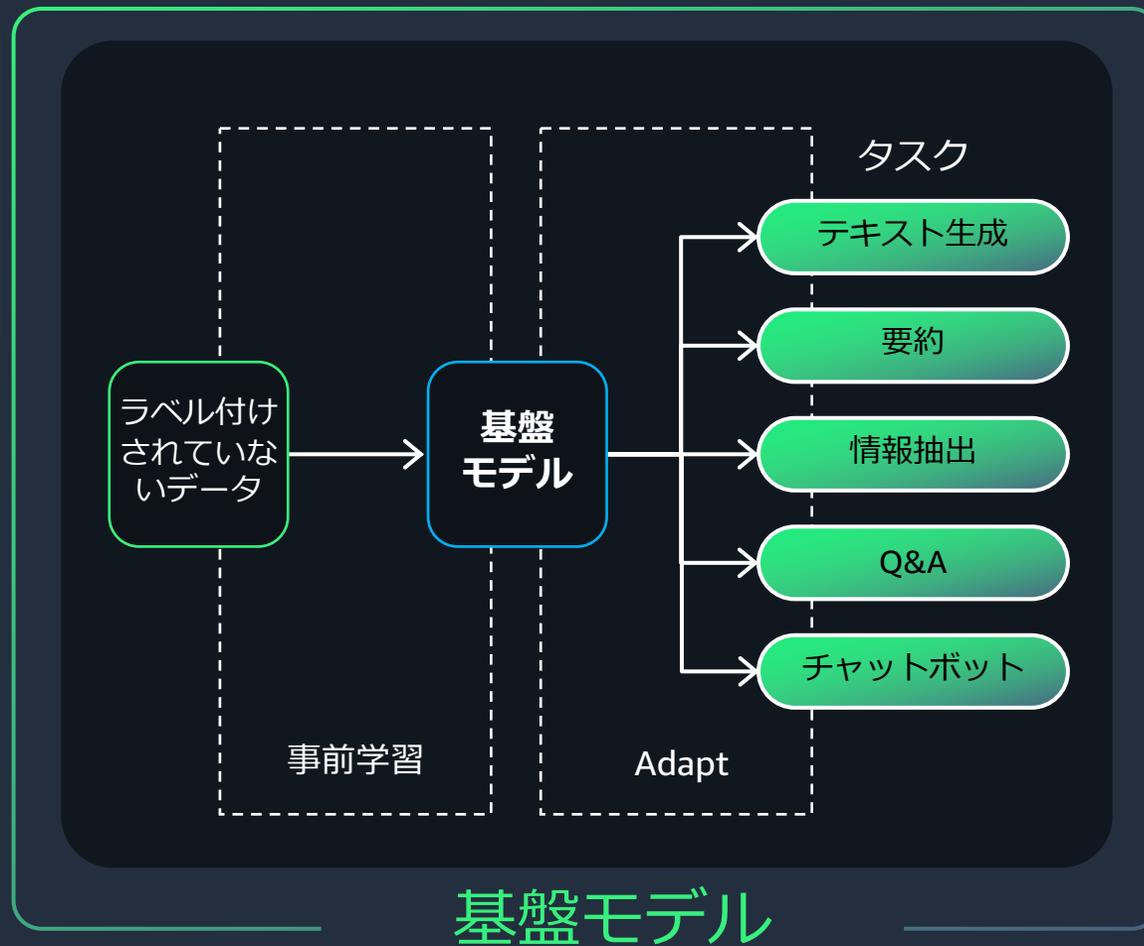
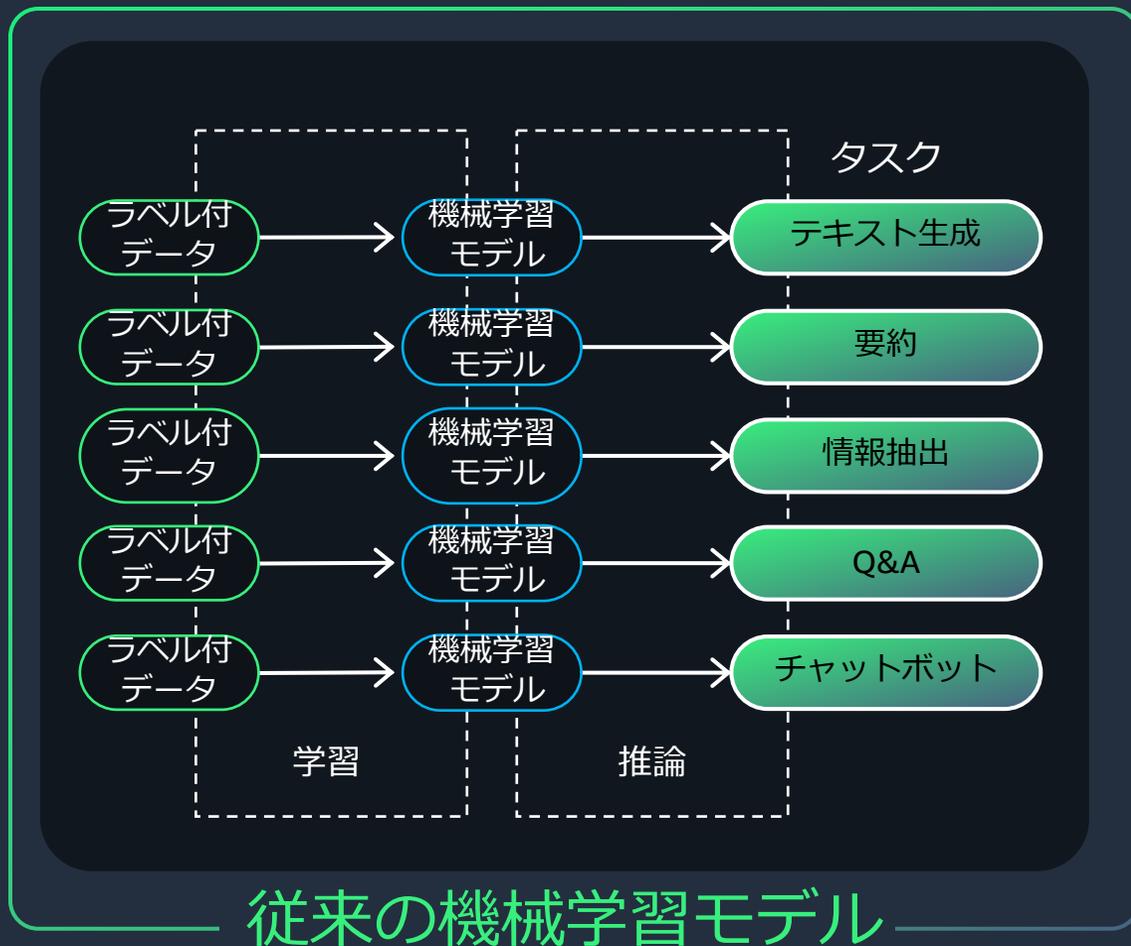
a robot thinking (how does a brain of a human)

# 生成系 AI とは

- 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 (Foundation Model) 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利用



# 従来と異なる「基盤モデル」というアプローチ



# 製薬バリューチェーンにおける生成系 AI の幅広いユースケース



## 研究開発

タンパク質  
フォールディング  
タンパク質設計



## 臨床開発

臨床試験実施計画書  
患者集団、施設  
の最適化  
合成対象群(SCA)



## 製造

予知保全  
リソース最適化



## 商用事務 医療事務

患者の治療結果  
予測  
コンテンツ生成



## 患者サポート

患者ケア  
コンシェルジュ  
患者と治験の  
マッチング

# 創薬研究領域における 生成系 AI on AWS



# 創薬研究領域における生成系 AI のイノベーション

汎用的な生成系 AI を活用した  
研究活動の生産性向上

創薬ドメインに特化した  
生成系 AI を活用した  
解析やドラッグデザインの  
高度化

# 創薬領域における生成系 AI のイノベーション

汎用的な生成系 AI を活用した  
研究活動の生産性向上

創薬ドメインに特化した  
生成系 AI を活用した  
解析やドラッグデザインの  
高度化

# 創薬研究活動におけるユースケース例 – 論文要約

## ユースケース (生成系AI)

- 🏠 ホーム
- 💬 チャット
- 📄 要約
- 🔍 校正
- ✉️ メール生成
- 🔗 CS 業務効率化

## ツール (非生成系AI)

- 🔍 Kendra 検索

## 会話履歴

- 📄 GenerativeAIの概要
- 📄 Generative AI概要
- 📄 AAAプラン更新とエンタープラ

## リンク

- 🌐 Bedrock
- 🔗 GitHub

aw

[→ サインアウト

## 要約

### 要約したい文章

"Highly accurate protein structure prediction with AlphaFold "

#### Abstract

Proteins are essential to life, and understanding their structure can facilitate a mechanistic understanding of their function. Through an enormous experimental effort, the structures of around 100,000 unique proteins have been determined but this represents a small fraction of the billions of known protein sequences. Structural coverage is bottlenecked by the months to years of painstaking effort required to determine a single protein structure. Accurate computational approaches are needed to address this gap and to enable large-scale structural bioinformatics. Predicting the three-dimensional structure that a protein will adopt based solely on its amino acid sequence—the structure prediction component of the ‘protein folding problem—has been an important open research problem for more than 50 years. Despite recent progress, existing

#### ▼ 追加コンテキスト - *Optional*

日本語の文章で要約してください。アルゴリズムのところを詳しく説明してください。

クリア

実行

# 創薬研究活動における基盤モデル活用のユースケース事例

## 基盤モデル単体で活用する場合

### 論文要約

Prompt 例

「以下の論文を3行で要約して ....」

### 研究レポート生成

「以下のフォーマットに合うようにレポートを作成して ....」

### コード生成

「PyTorchで〇〇するコードを書いて」

### 翻訳

Prompt 例

「以下の文章を日本語に翻訳して ....」

### レポート校正

「以下の文章の中に誤字脱字や不適切な文法があったら指摘して ....」

### 情報抽出

「以下の論文で、〇〇についてはどう書かれていますか？ ...」

# クラウド = “ビルディングブロック”

## AWS はやりたいことをご自身で実現する Self Service Platform



1つのサービスやツールでは自由度と実装コストの両立に限界



- **“Building Block”** = 複数のサービスを適材適所で組み合わせ、やりたいことを最小の手間で実現。  
AWS が提供しているのは皆様のやりたいことを支えるサービス  
≡ インフラの“パーツ” (アプリケーション)



AWSサービスは200以上



# Amazon Bedrock

最も簡単に 基盤モデルを活用し  
生成系 AI アプリケーションを  
構築できるサービス



API を介して基盤モデルを利用することで  
生成系 AI アプリケーションの開発を加速  
インフラの管理は不要



お客様の業務用途に適した基盤モデルを選択  
Amazon、AI21 Labs、  
Anthropic、Cohere、Meta、Stability AI



自社データを使用して  
基盤モデルを非公開で  
ファインチューニング

# Amazon Bedrock : 主要な基盤モデルを幅広くサポート

amazon

## Amazon Titan

テキストの要約、生成、分類、自由形式のQ&A、情報抽出、埋め込み、検索

**AI21 labs**

## Jurassic-2

スペイン語、フランス語、ドイツ語、ポルトガル語、イタリア語、オランダ語でのテキスト生成用の多言語LLM

**ANTHROPIC**

## Claude 2

誠実で責任感のあるAIシステムのトレーニングに関する研究に基づく、会話、質問応答、ワークフロー自動化のためのLLM

cohere

## Command

ビジネスアプリケーション用のテキスト生成モデルと、検索、クラスタリング、分類用の埋め込みモデル (100以上の言語に対応)

Meta

*Llama 2 (近日公開予定)*

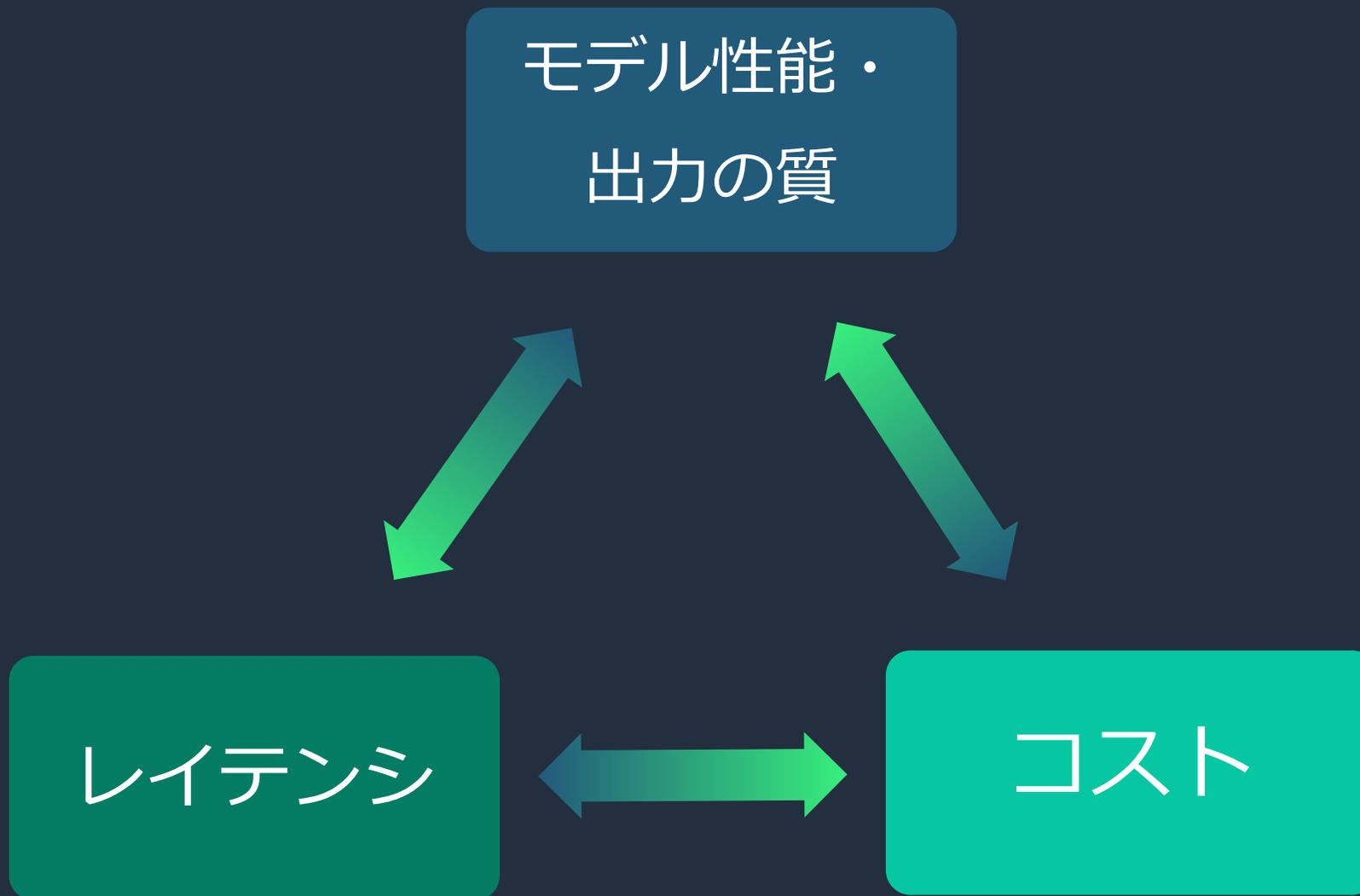
対話のユースケースや言語タスクに最適な、微調整されたモデル

**stability.ai**

## Stable Diffusion

ユニークでリアルで高品質な画像生成、アート、ロゴ、デザインの生成

# 基盤モデルの選択指針：個々のゴールにあわせたバランスが重要



# Amazon Bedrock なら お客様データは プライベートかつセキュア



お客様のデータが  
利用される基盤モデルの学習に  
利用されることはありません

お客様のデータはすべて作成  
されたリージョンに留まります



全てのデータは暗号化され  
転送、保存されます

オンプレミス - クラウド間の  
閉域接続での利用も可能



GDPR (一般データ保護規則)  
HIPAA コンプライアンス等  
標準規格に準拠



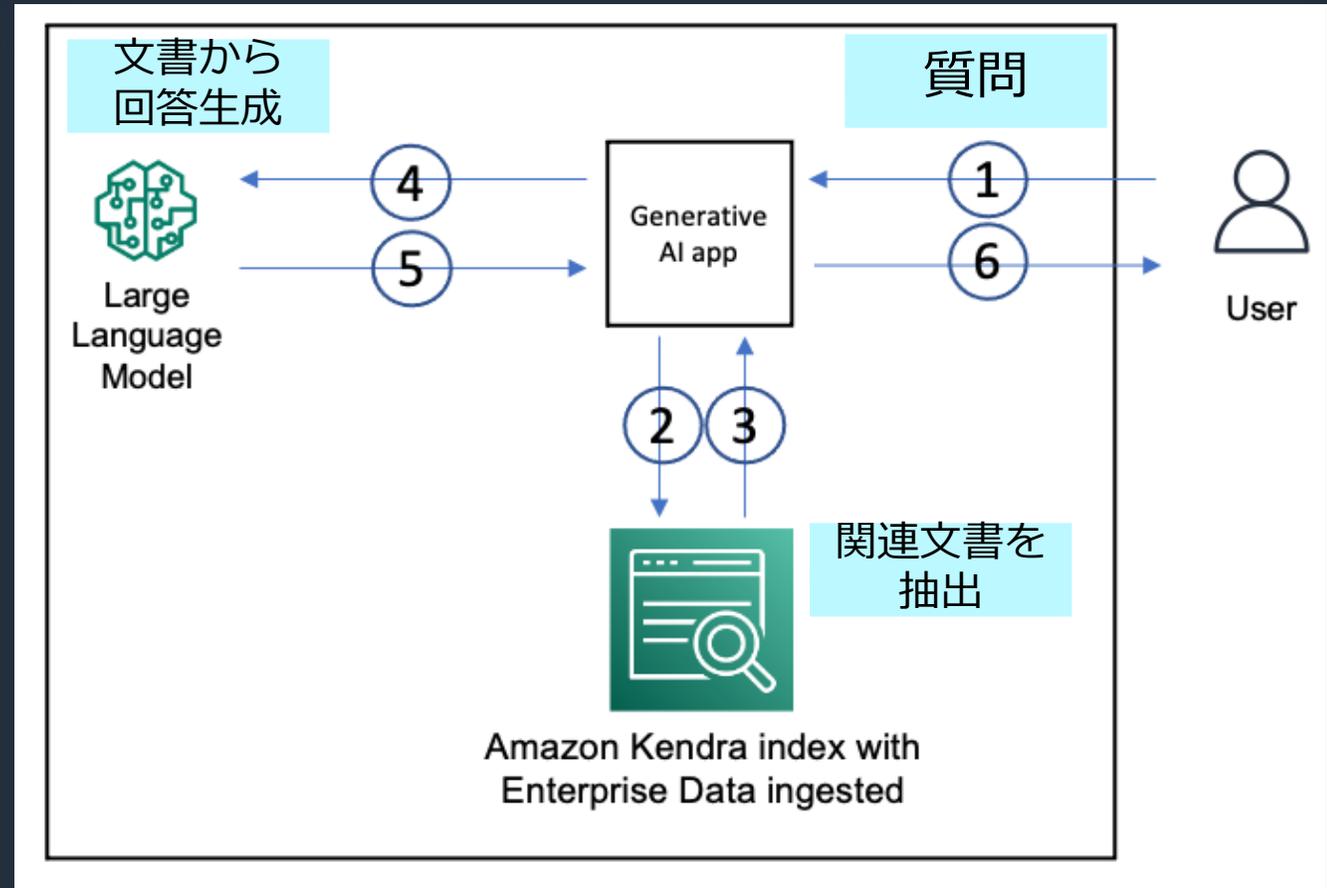
お客様ご自身のデータの活用は  
お客様の競争力・差別化につながります



# 社内のドキュメントや研究データと基盤モデルを組み合わせる

## 検索拡張生成 (RAG; Retrieval Augmented Generation)

- 社内のノウハウやデータや外部の情報などの膨大な情報から、**ユーザーの関心ごとに関する情報を抽出し、その要約を生成**
- 生成系 AI の弱点である**ハルシネーション (事実ではない回答を出力する)**を抑制し、回答の精度を向上
- **基盤モデルのファインチューニング不要**で自社のデータに基づいた回答が可能
- 文書検索の仕組み (Amazon Kendraなど) と生成系 AI (LLM) の機能を組み合わせ、組織内のコンテンツに基づく会話型UXを提供する最先端の生成系 AI アプリケーションを作成



※ RAG実装の一例。Amazon Bedrock の機能(Preview)だけでカバーすることもできます。

<https://aws.amazon.com/jp/blogs/news/quickly-build-high-accuracy-generative-ai-applications-on-enterprise-data-using-amazon-kendra-langchain-and-large-language-models/>



# RAG によって正しい情報のみに基づいた回答が可能に

## 基盤モデルのみ



Amazon Bedrockのセキュリティについて教えてください。  
なぜ Amazon Bedrock が安全に利用できるのかわかるように説明してください。



Amazon BedrockはAmazonが提供するマネージド型のブロックチェーンサービスです。Bedrockは高いセキュリティを実現するために以下のような特徴があります。

- 独自のコンセンサスアルゴリズムを使用しているため、51%攻撃などのリスクを軽減できます。
- ノードの運用とメンテナンスはAmazonが管理しています。
- すべてのトランザクションは暗号化されています。
- Amazonの厳格な物理的セキュリティ対策が適用されています。
- アクセス制御や監査ログなど、企業向けのセキュリティ機能が用意されています。
- 定期的なセキュリティ監査とペネトレーションテストが実施されています。

このよう  
め、重要  
の取り組

- 学習データに含まれていない情報は知り得ない  
→ 事実ではない情報が含まれる
- 回答の根拠がわからない

社内データやドキュメントにある  
情報に基づいた正確な回答

## RAG (= 文書検索 + 基盤モデル)



Amazon Bedrockのセキュリティについて教えてください。  
なぜ Amazon Bedrock が安全に利用できるのかわかるように説明してください。



Amazon Bedrockは、AWSの安全なインフラストラクチャ上に構築されています。

Amazon Bedrockは、機密データを保護するために、次のセキュリティ対策を講じています。

- すべての通信はTLSによって暗号化されています。
- 顧客のデータは、AWSの安全なストレージサービスであるAmazon S3に保存されています。
- モデルのトレーニングと推論は、AWSのセキュアなコンピューティングサービスであるAWS Lambda上で実行されます。
- ユーザー認証にはAWS Identity and Access Management(IAM)が使用されており、アクセスは最小限の権限で制御されます。

また、Amazon Bedrockは継続的にセキュリティの検証と改善を行っています。コンプライアンス認定も多数取得しており、グローバルな大企業からも信頼されています。

このように、AWSの安全なインフラとサービス、そして継続的なセキュリティ対策により、Amazon Bedrockは安全に利用できる生成系AIサービスです。

回答の参考ドキュメント

[基盤モデル API サービス - Amazon Bedrock - AWS](#)

[基盤モデル API サービス - Amazon Bedrock - AWS](#)

根拠 (URL) も  
含めて回答



# RAG 活用例：大量の研究・試験データに基づいて回答する 対話型 AI チャットボット



The screenshot displays the Amazon HealthLake ChatBot interface. At the top, the title "Amazon HealthLake ChatBot" is shown. The chat history includes two user questions and two bot responses. The first question asks about deaths related to carbamazepine, and the bot responds that one death has occurred. The second question asks about common side effects, and the bot lists dizziness, drowsiness, anxiety, nausea, and vomiting. At the bottom, there is a search bar with the text "Search FDA labels for the most common adverse reactions seen with carbamazepine", and three buttons: "Send", "Clear Chat", and "View Kendra Evidence".

Amazon HealthLake ChatBot

カルバマゼピン（抗てんかん剤）を使用した患者はこれまで何人亡くなったか

カルバマゼピン専用薬を使用した患者は、1人亡くなっています

カルバマゼピン専用薬治療で見られる最も一般的な副作用について米国食品医薬局のFDAラベルを確認して

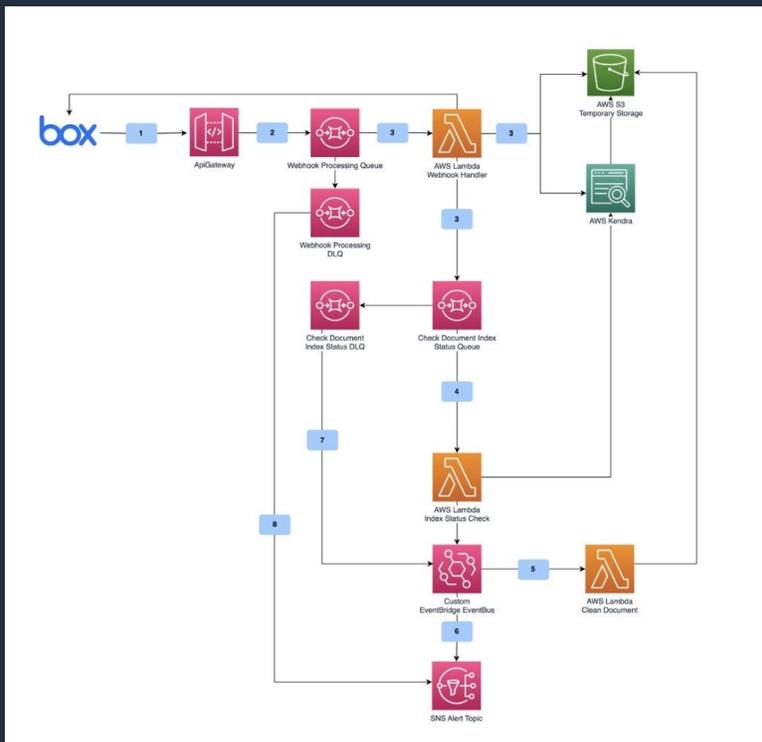
特に治療の初期段階でカルバマゼピンの経口治療で見られた最も一般的な副作用は、めまい、眠気、不安定、吐き気、嘔吐でした

Search FDA labels for the most common adverse reactions seen with carbamazepine

Send Clear Chat View Kendra Evidence

# 事例：アストラゼネカ株式会社 様

## AWS を活用したエンタープライズ検索と規制対応の改善



規制が厳しく、消費者保護に重点が置かれていることで知られる日本では、製薬会社は信頼を維持し、患者の健康を確保するために、一連の倫理ガイドラインを遵守しなければなりません。アストラゼネカ株式会社 様(AZKK)はAmazon Kendra を使用したエンタープライズ検索における利用者体験の向上に取り組みました。

検索拡張生成 (RAG) ベースのアプリケーションで大規模言語モデル (LLM) を搭載した対話型AIを使用し、既存の Amazon Kendra ベースのソリューションを Amazon Bedrock や Amazon Sagemaker JumpStart などのネイティブ AWS サービスで強化し、それぞれのユースケースに最適なさまざまな LLM にアクセスできるようにしています。

“このプロジェクトは年間12,000時間の生産性の改善を実現しており、社内のAIエンジニアの才能と、AWSチームとの緊密なパートナーシップによってもたらされるインパクトを示しています。”

**Harsh Gandhi**

Head of Information and Digital, IT, アストラゼネカ株式会社



Amazon Bedrock



# 生成系 AI を始めるには

# Step1. まずは実際に触ってみるところから： AWS アカウントさえあればすぐにお試し可能

Chat playground [Info](#)  Streaming

AI Anthropic Claude V2 v2 Start new chat

Using on-demand provisioned throughput. [Switch provisioned throughput](#)

AWSってなんですか？小学生でもわかるように教えてください。 Maximum length: 853

AWSはAmazon Web Servicesの略です。アマゾンが提供しているクラウドサービスのことです。

クラウドサービスとは、インターネットを通じてコンピューターの処理能力やストレージ、ソフトウェアなどを提供するサービスのことです。

AWSを使うと、自分のコンピューターを準備する必要がなく、インターネットさえあれば必要な時に必要なだけコンピューターの能力を借りることができます。

例えば、自分のサイトやアプリを作りたい時に、サーバーを自分で用意する必要はなく、AWSでサーバーを借りることができます。大量のデータを処理したい時も、一時的に強力なコンピューターを借りることができるので便利です。

AWSを使えば、自分でコンピューターの管理をする必要がなくなるので、アプリやサイトを簡単に作るできるようになります。アマゾンが安全な運用をしてくれるので、小学生でも気軽に利用できます。

Add instructions Maximum length: 853 [Update inference configurations](#)

Human:

デモ画面：チャットボット

Image playground [Info](#)

S. Stability AI Stable Diffusion XL v0.8 - preview

Using on-demand provisioned throughput. [Switch provisioned throughput](#)

Prompt  
Researchers in the life science field who conduct experiments in the Laboratory

Response



Inference configuration [Info](#)

Reset

Prompt strength 10

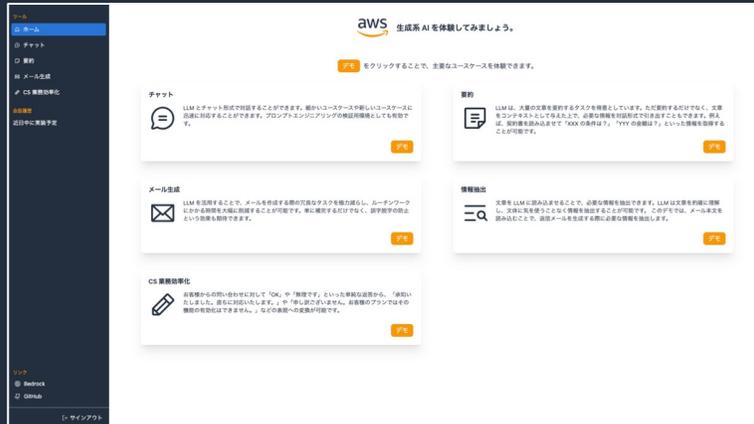
Generation step 50

Seed 0

デモ画面：画像生成

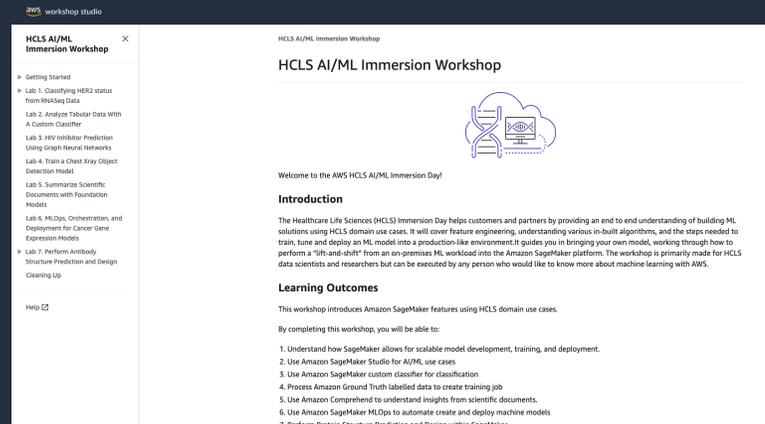
# Step2. 生成系 AI のユースケースを理解する

デモでユースケースを知る

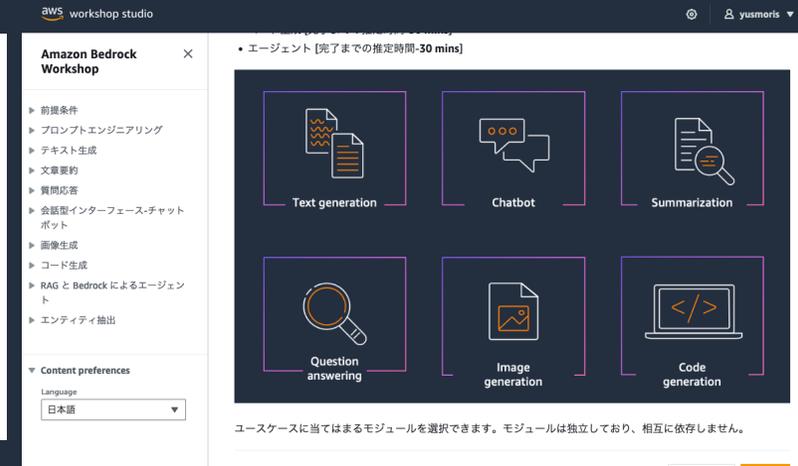


[Generative AI Use Cases JP](#)

手を動かしてより理解を深めるハンズオン



[HCLS AI/ML Immersion Workshop](#)



[Amazon Bedrock Workshop](#)

...and more

# Step3. AI/機械学習 on AWS の実践を加速するプログラム

## 生成系 AI 特化の支援



### AWS Generative AI Innovation Center

ヒアリングからPoCに至るまで  
皆様の業務における  
生成系 AI の具体的なユース  
ケースの実現を支援。

## ML ナレッジの育成



### ML Enablement Workshop

機械学習の使いどころを学び、  
実行可能な計画を立てられるよ  
うになることが目的の無償ワー  
クショップ。

## ML ユースケース評価



### ML Discovery Workshop

AWS のエキスパートとの議論を  
通じビジネス価値・実現性の詳  
細な評価を行う。

まずはお気軽に AWS にお問い合わせください。

# 本セッションのまとめ

- 創薬研究領域での AWS 活用によってクラウドのスケーラビリティや俊敏性を生かした機械学習が実現
- 生成系 AI は創薬において生産性向上・ドラッグデザインの高度化をもたらす
- AWS では生成系 AI を簡単に活用いただけるサービスをご提供
  - Amazon Bedrock : 複数の基盤モデルから最適なものを API 経由で利用

# Thank you

