

ビジネスのための生成AIと それを支えるプラットフォーム

2024年11月15日

日本アイ・ビー・エム株式会社

テクノロジー事業本部

Data & AI テクニカルスペシャリスト

井上忠宣



本日の内容

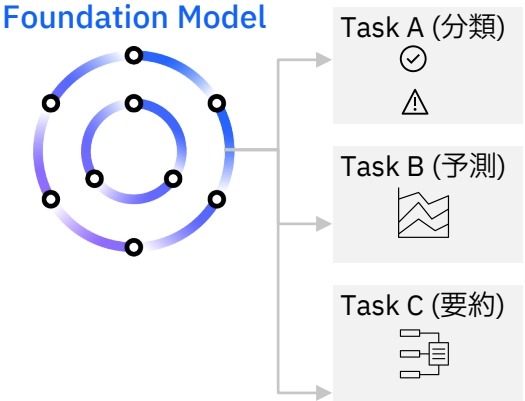
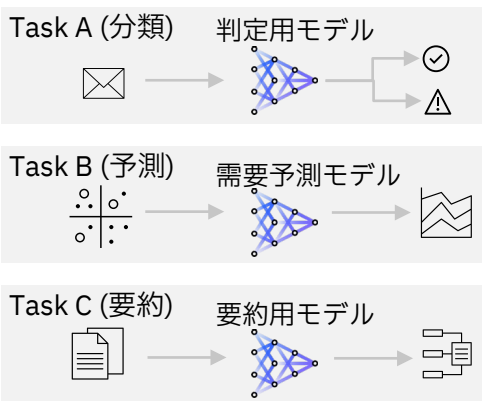
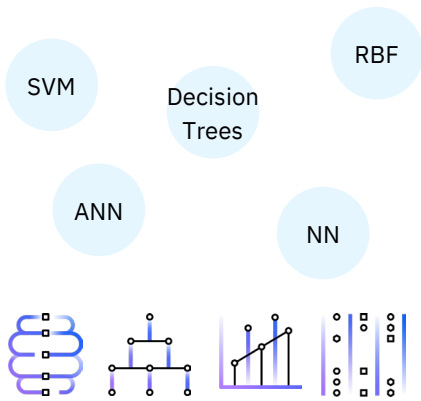
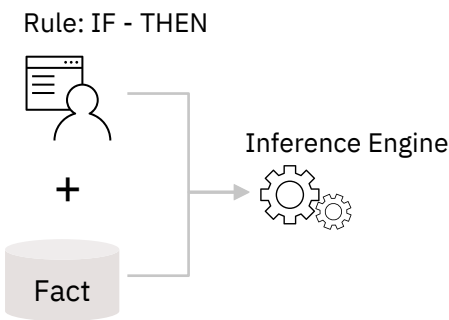
- 1 生成AIを前提としたビジネス設計へ
- 2 ビジネスのための生成AI
- 3 ビジネスのための生成AIを支えるプラットフォーム
- 4 生成AI活用によるビジネス変革
- 5 まとめ

本日の内容

- 1 生成AIを前提としたビジネス設計へ
- 2 ビジネスのための生成AI
- 3 ビジネスのための生成AIを支えるプラットフォーム
- 4 生成AI活用によるビジネス変革
- 5 まとめ

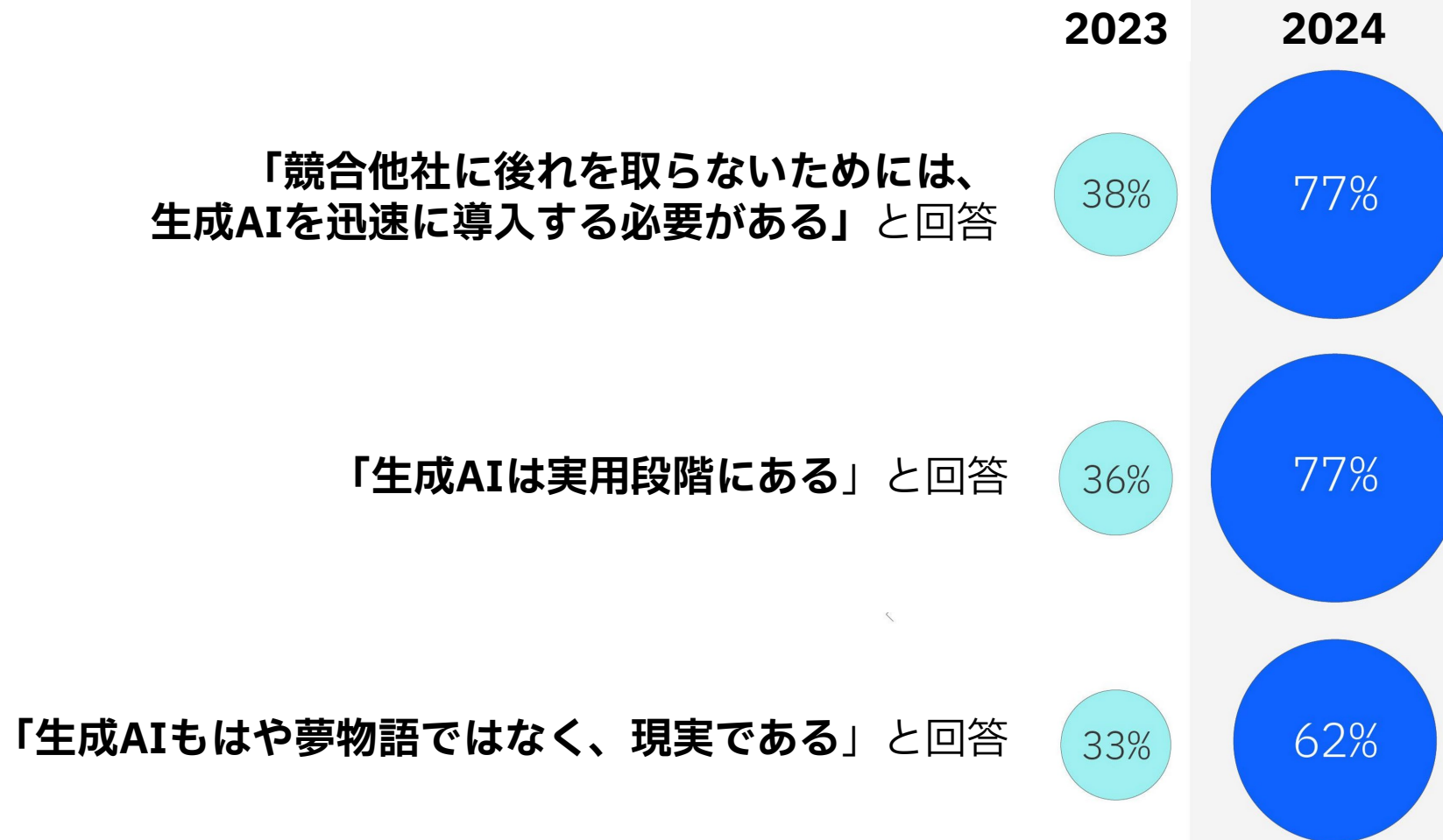
基盤モデルの登場によりAIは新たな時代に突入

	エキスパートシステム	機械学習	ディープラーニング	基盤モデル
メリット	ルールベースでの自動化	ロジック・ルールが(人手でなく)、データから作れる	- 予測精度の向上 - 人手による特徴量設計が不要	1つのFoundation Modelを複数のタスクに適用できる - 人手によるラベル付けが大幅に減る
課題	人手でのロジック・ルール作り	予測精度 (場合により)	- ブラックボックス問題 - 大規模計算	- AI倫理全般 - 超大規模計算
時期	1980年代	1980年代 ~ 2010	2012 ~	2017 ~



経営層の生成AIへの期待 (1/2)

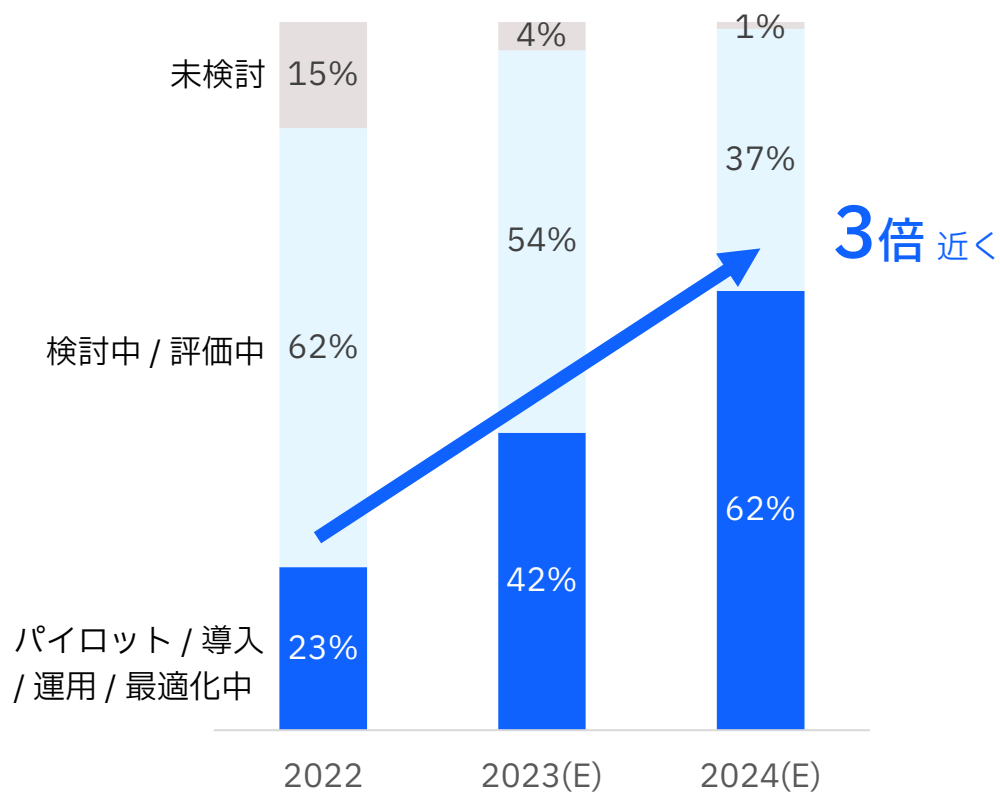
生成AIを一過性のトレンドと捉えていた経営層も考えを修正、
自社の競争力のため実用化を加速させる必要性を感じている



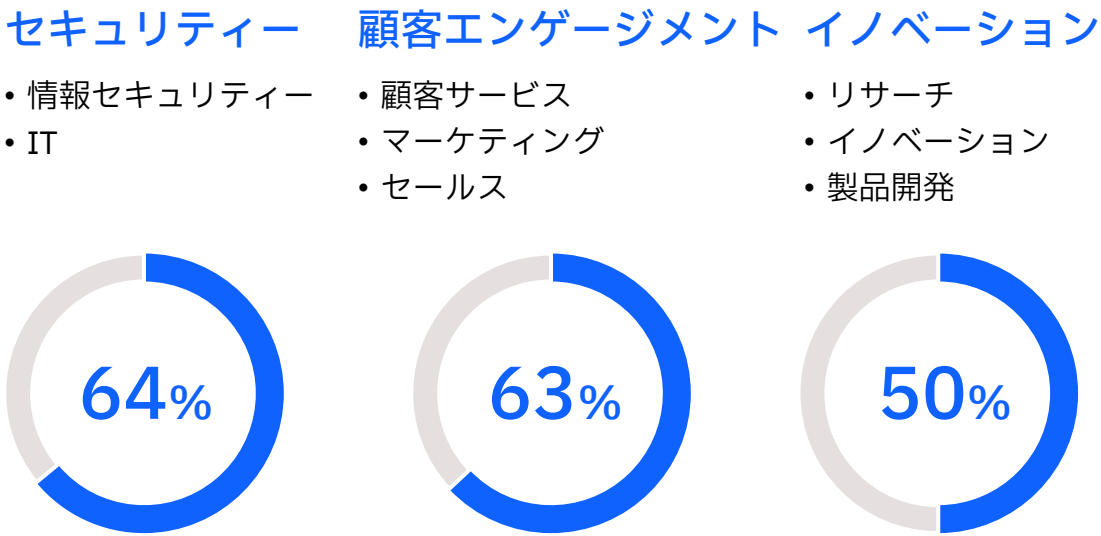
経営層の生成AIへの期待 (2/2)

世界の600人の経営層が、2023年時点で2022年からの3年で、生成AIの導入は3倍近くになると予想

生成AI適用予定

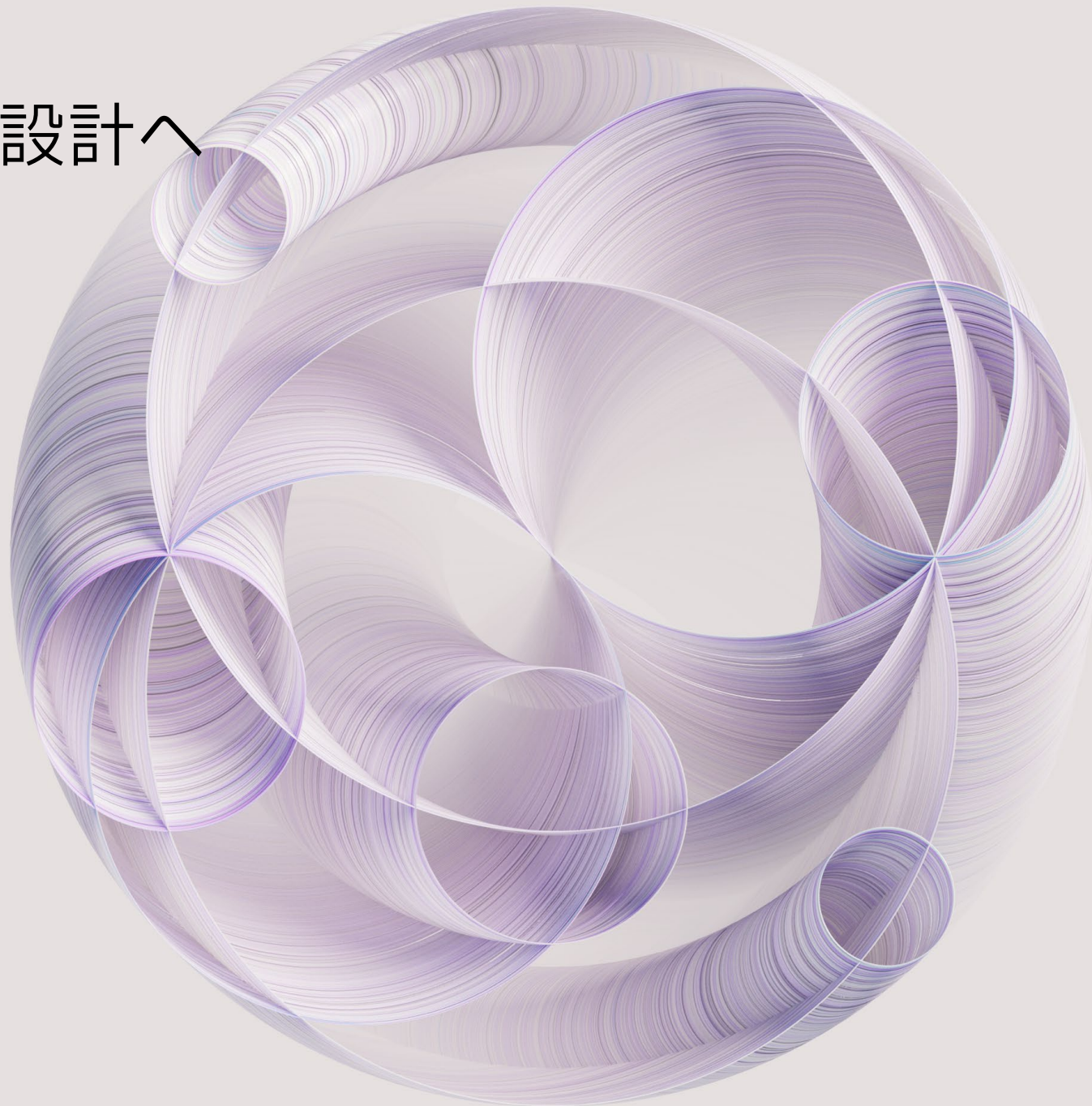


生成AI適用の期待領域



AIファースト
生成AIを前提としたビジネス設計へ

+AI → AI+

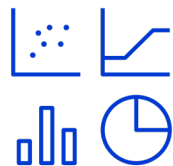


本日の内容

- 1 生成AIを前提としたビジネス設計へ
- 2 ビジネスのための生成AI
- 3 ビジネスのための生成AIを支えるプラットフォーム
- 4 生成AI活用によるビジネス変革
- 5 まとめ

ビジネスのための生成AI

個人の生産性向上の範囲を超え、生成AIを企業の競争力を高める原動力にするためには、3つの視点を踏まえたAI活用の推進が重要となる。



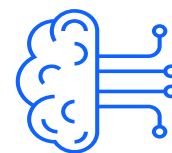
ビジネス文脈の理解

ビジネスの文脈やルールを理解したモデルの活用



プロセスへの組み込み

既存のビジネスプロセスにAIを統合



信頼性の確保

説明性, 公平性, 堅牢性, 透明性をもったAI

生成AI利用の3つのアーキテクチャー・パターン

生成AIを活用する際、大きく3つのアーキテクチャー・パターンがあります。

① 組み込み型

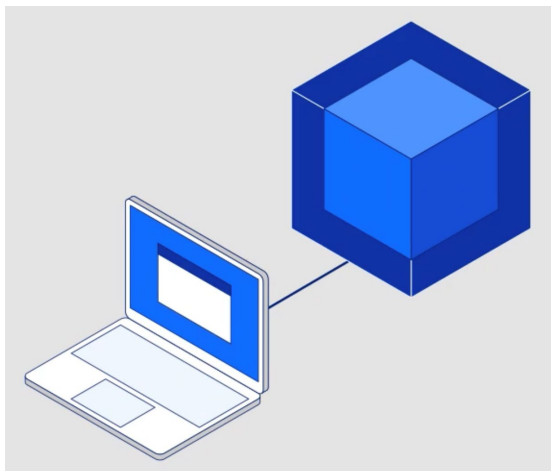
- ・ 製品やサービスに生成AIが含まれているもの

② API型

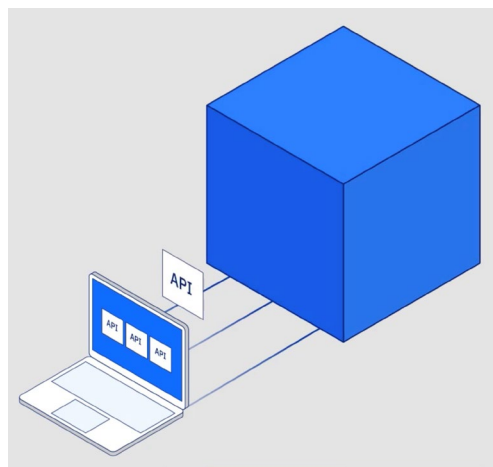
- ・ 外部の生成AIをAPIとして呼び出して利用するもの

③ 自社構築AI型

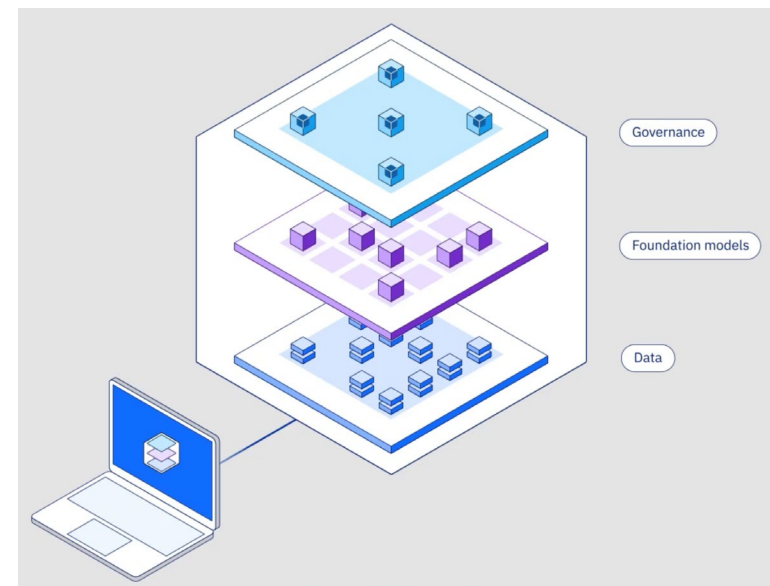
- ・ 所有するデータでAIを独自に拡張し利用するもの



① 組み込み型



② API型



③ 自社構築AI型

これからの企業はAI の価値創造者になる

AIを利用するだけではなく
自社のデータを使い
信頼できる独自のAIを構築することで
新たなビジネスを創出する



生成AI利用の3つのアーキテクチャー・パターン

生成AIを活用する際、大きく3つのアーキテクチャー・パターンがあります。

① 組み込み型

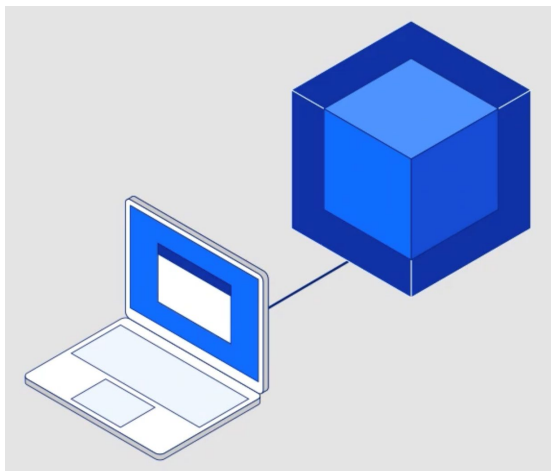
- ・ 製品やサービスに生成AIが含まれているもの

② API型

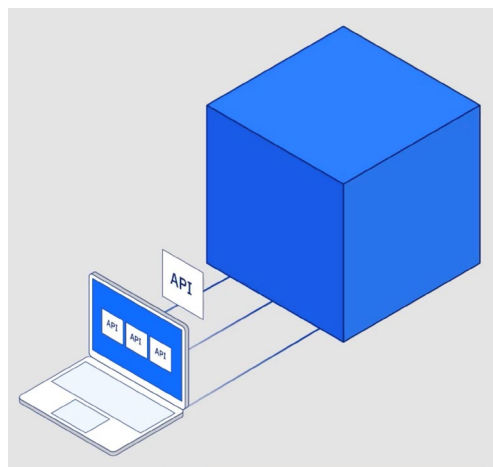
- ・ 外部の生成AIをAPIとして呼び出して利用するもの

③ 自社構築AI型

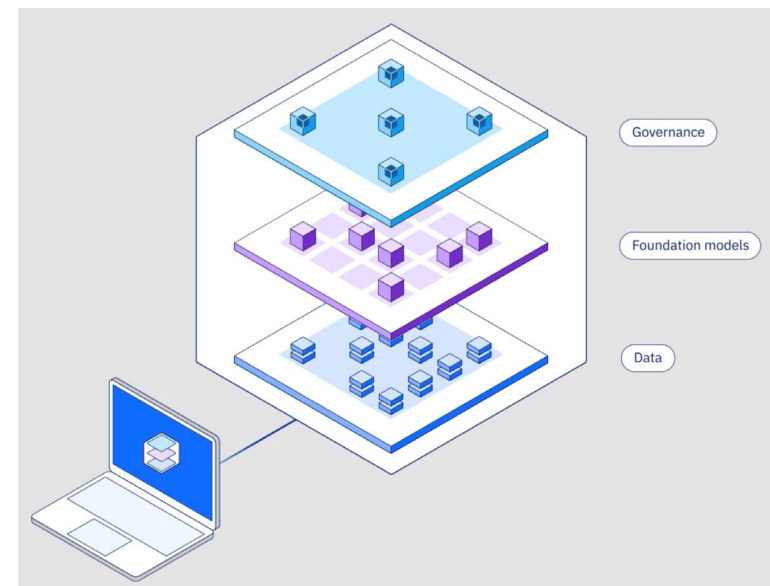
- ・ 所有するデータでAIを独自に拡張し利用するもの



① 組み込み型



② API型



③ 自社構築AI型

生成AI活用のステップ

企業が自社固有・業界固有のデータを活用し、生成AIを業務へ適用させるためには、**業務特性とROI**を踏まえて、生成AI活用のステップを段階的に登っていく必要がある。

開発にかかる 工数			アプリケーション	プロンプト 生成	チュー ニング	基盤 モデル
	Lv.1 既成モデル活用	サービス化された生成AIをそのまま活用し、主に個人の生産性向上を図る				
	Lv.2 プロンプト エンジニアリング	- 既存のモデルをパラメータ等を微調整する程度でそのまま活用 - 生成AIサービスをラップするアプリケーションを開発し、企業固有のセキュリティ要件に対応（アクセス制御、オプトアウト等）		ベンダー管理		
	Lv.3 モデルのカスタム ・チューニング	- 投げかけるプロンプトの工夫や自動化により、業務への適応度を向上させる(例: few-shot prompting, chain of thought) - 知識ベースとの組み合わせにより、自社固有データを検索し、文脈情報として追加したプロンプトにより生成させる（RAG）				
	Lv.4 独自モデルの構築	自社・業界固有のデータを用いて、企業独自のモデルをゼロから構築				

本日の内容

- 1 生成AIを前提としたビジネス設計へ
- 2 ビジネスのための生成AI
- 3 ビジネスのための生成AIを支えるプラットフォーム
- 4 生成AI活用によるビジネス変革
- 5 まとめ

ビジネスのための生成AIを支えるプラットフォーム

3つの視点を踏まえたAI活用の推進が重要

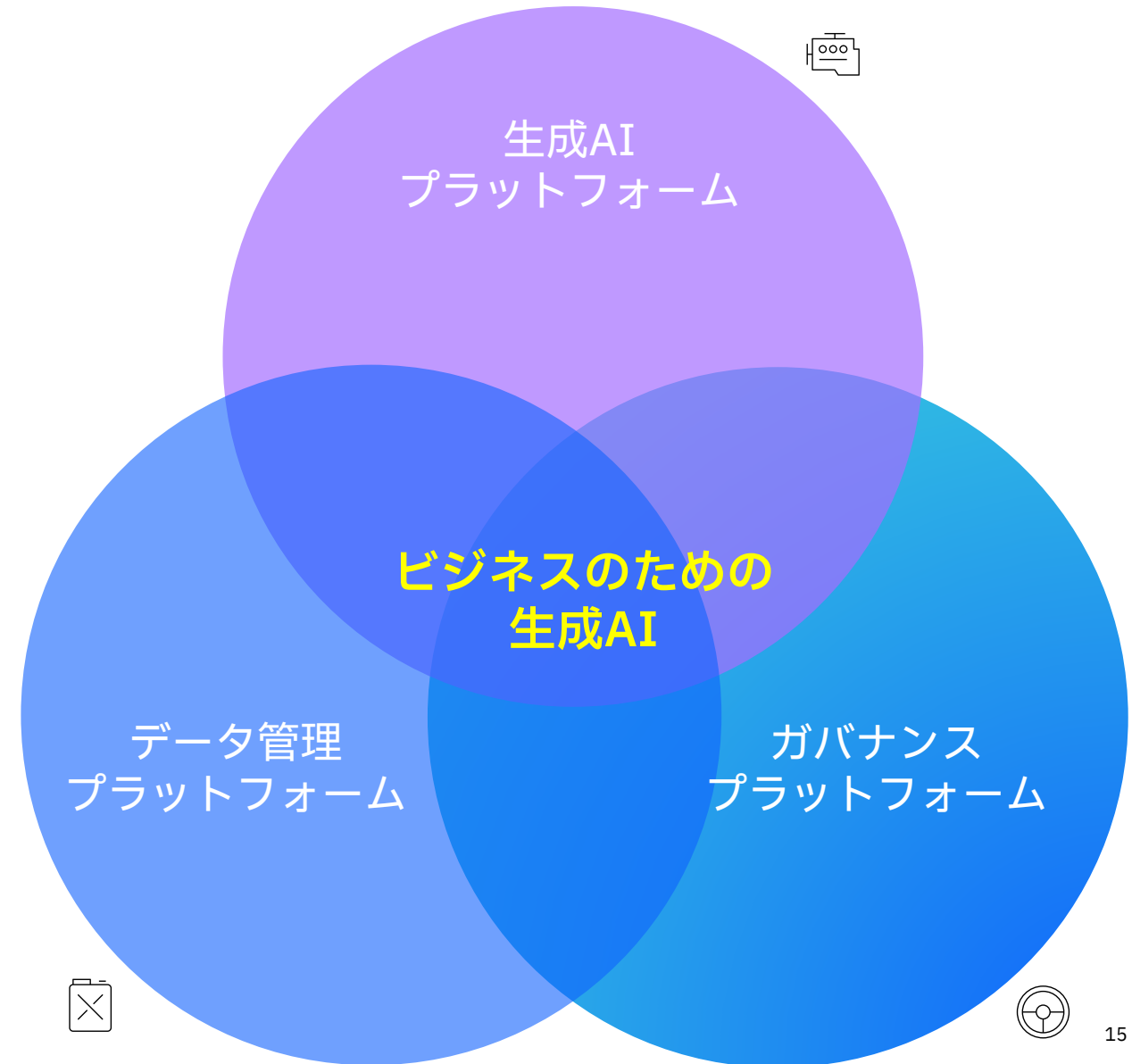
- ① ビジネス文脈の理解
- ② プロセスへの組み込み
- ③ 信頼性の確保



ビジネスのための生成AIを支える3つのプラットフォーム

- ① 生成AIの活用
- ② 自社データの管理
- ③ 信頼性確保のためのガバナンス

(もしなかったら…)



ビジネスのための生成AIを支えるプラットフォームの例

IBMの場合 = watsonx

watsonx.ai

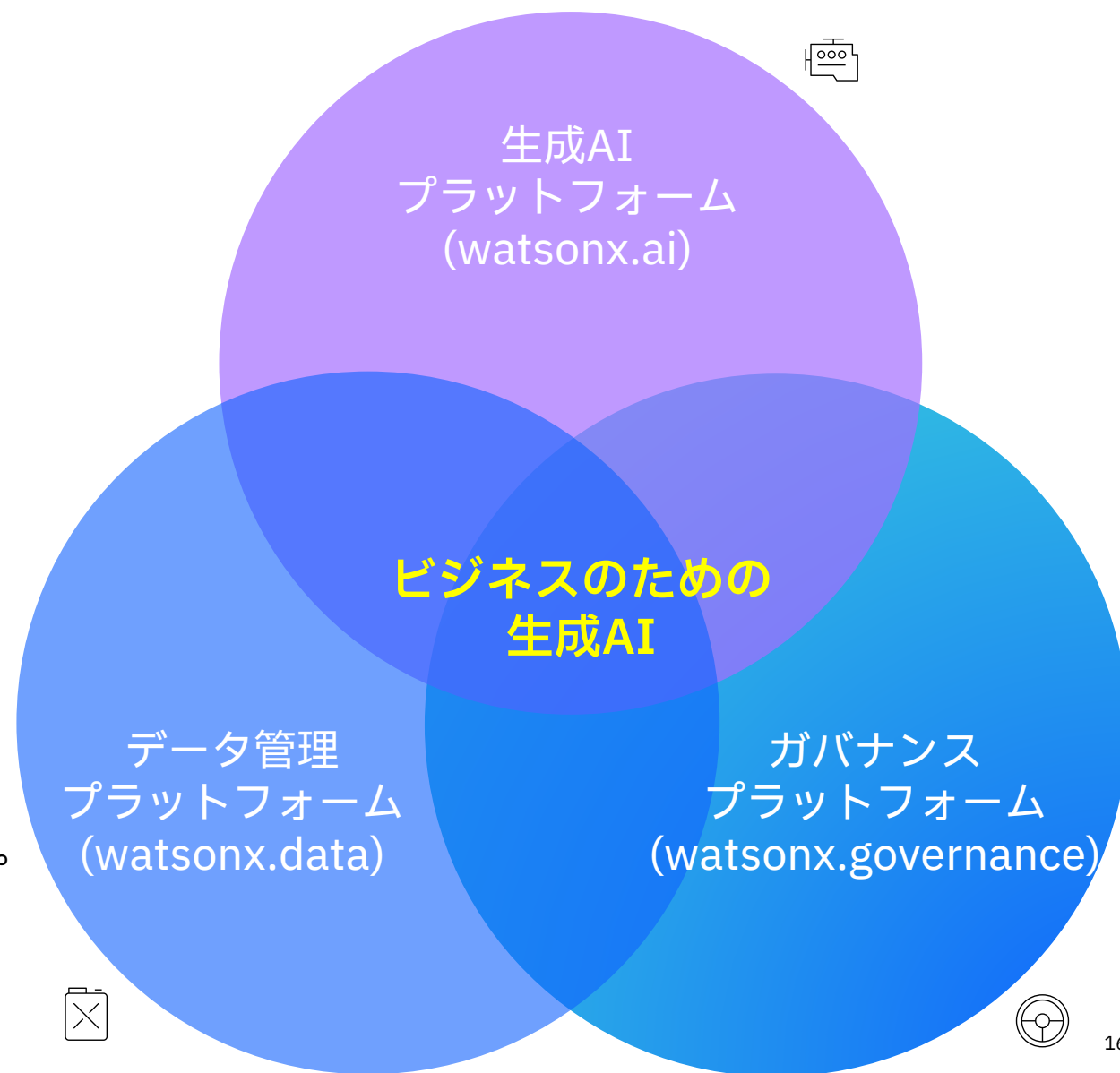
企業独自の**競争力と差別化**を保持するために、基盤モデルをはじめ様々な**AIモデルを活用・構築**することができる

watsonx.data

AIを**ビジネスのあらゆる領域**で活用するために、企業の散在する**データを一元管理**し、さらに活用できるよう加工する仕組みを提供する

watsonx.governance

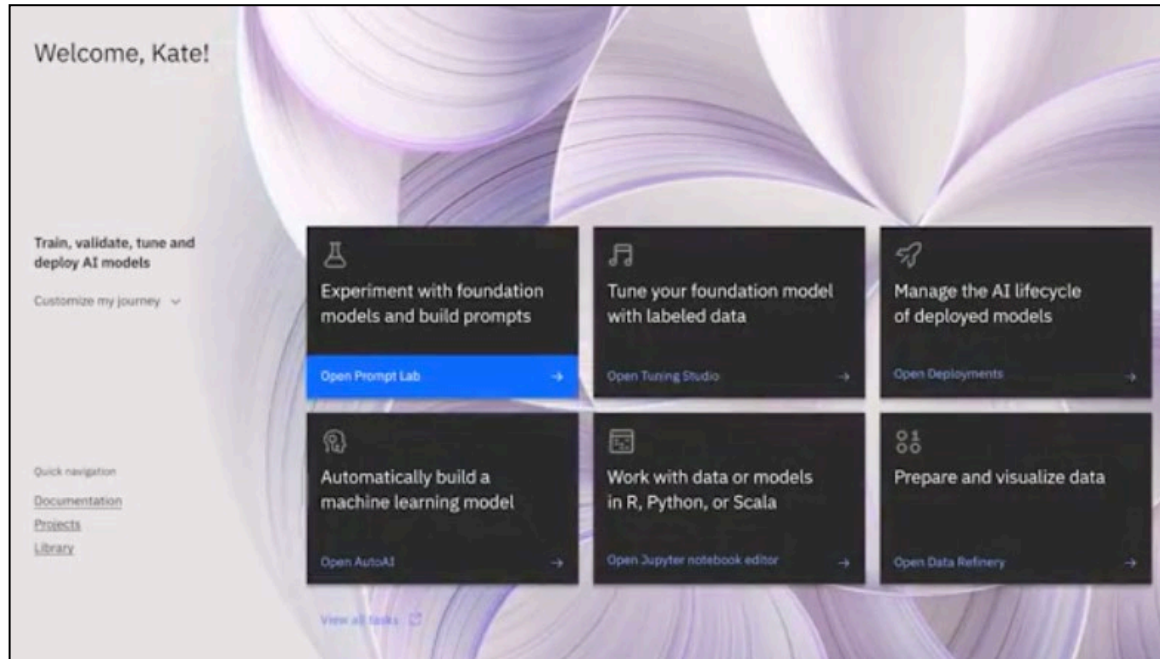
信頼できるAIの構築を支援する、データとAI双方のガバナンスを包含したツールキット。責任ある、透明性・説明可能性を確保したAIのワークフローを実現する



①生成AIプラットフォームの例

watsonx.ai

AI開発者向けエンタープライズ・スタジオ



生成 AI、基盤モデル、および機械学習機能のトレーニング、検証、チューニング、導入を行う

watsonx.ai のコンポーネント:

- **基盤モデル・ライブラリー:**
様々なモデルを提供
- **プロンプト・ラボ:**
基盤モデルを試し、さまざまなユースケースやタスクのためのプロンプトを構築
- **チューニング・スタジオ:**
ラベル付きデータで基盤モデルをチューニング
- **データサイエンスと MLOps:**
モデルのトレーニング、開発、ビジュアル・モデリング、および合成データ生成で機械学習モデルを自動で構築

①生成AIプラットフォームの例

watsonx.ai

基盤モデル・ライブラリー

マルチ基盤モデル

ひとつのモデルですべてのユースケースについて最適なわけではありません。

デプロイ済みのモデルを選択

➤ すぐに基盤モデルの活用が可能

- IBM 独自開発モデル **Granite**
- オープンソースモデル
- サード・パーティーモデル
- BYOM

基盤モデルの選択

モデルを選択するには、モデルが実行するタスクなどの特性を確認します。モデルのベンチマークを 0 ~ 100 のスコアで比較する。スコアは高い方がいい。

すべてのモデル

モデルのベンチマーク

▼

Q

モデルまたはタスクの検索

チャット

×

フィルターのリセット

自分のモデルを持ち込みたい？

 granite-13b-chat-v2 The Granite model series is a family of IBM-trained, dense decoder-only models, which are particularly well-suited for generative tasks. プロバイダー: IBM タイプ: InstructLab	 granite-13b-instruct-v2 The Granite model series is a family of IBM-trained, dense decoder-only models, which are particularly well-suited for generative tasks. プロバイダー: IBM タイプ: 標準提供モデル	 granite-20b-code-instruct The Granite model series is a family of IBM-trained, dense decoder-only models, which are particularly well-suited for generative tasks. プロバイダー: IBM タイプ: 標準提供モデル	 granite-20b-multilingual The Granite model series is a family of IBM-trained, dense decoder-only models, which are particularly well-suited for generative tasks. プロバイダー: IBM タイプ: InstructLab	 granite-3-2b-instruct The Granite model series is a family of IBM-trained, dense decoder-only models, which are particularly well-suited for generative tasks. プロバイダー: IBM タイプ: 標準提供モデル
 granite-3-8b-instruct The Granite model series is a family of IBM-trained, dense decoder-only models, which are particularly well-suited for generative tasks. プロバイダー: IBM タイプ: 標準提供モデル	 granite-34b-code-instruct The Granite model series is a family of IBM-trained, dense decoder-only models, which are particularly well-suited for generative tasks. プロバイダー: IBM タイプ: 標準提供モデル	 granite-3b-code-instruct The Granite model series is a family of IBM-trained, dense decoder-only models, which are particularly well-suited for generative tasks. プロバイダー: IBM タイプ: 標準提供モデル	 granite-7b-lab The Granite model series is a family of IBM-trained, dense decoder-only models, which are particularly well-suited for generative tasks. プロバイダー: IBM タイプ: InstructLab	 granite-8b-code-instruct The Granite model series is a family of IBM-trained, dense decoder-only models, which are particularly well-suited for generative tasks. プロバイダー: IBM タイプ: 標準提供モデル
 granite-guardian-3-2b The Granite model series is a family of IBM-trained, dense decoder-only models, which are particularly well-suited for generative tasks. プロバイダー: IBM タイプ: 標準提供モデル	 granite-guardian-3-8b The Granite model series is a family of IBM-trained, dense decoder-only models, which are particularly well-suited for generative tasks. プロバイダー: IBM タイプ: 標準提供モデル	 llama-2-13b-chat Llama-2-13b-chat is an auto-regressive language model that uses an optimized transformer architecture. プロバイダー: IBM タイプ: 標準提供モデル	 llama-3-1-70b-instruct Llama-3-1-70b-instruct is an auto-regressive language model that uses an optimized transformer architecture. プロバイダー: IBM タイプ: InstructLab	 llama-3-1-8b-instruct Llama-3-1-8b-instruct is an auto-regressive language model that uses an optimized transformer architecture. プロバイダー: IBM タイプ: 標準提供モデル

Graniteのオープンソース化とInstructLab

基盤モデルそのものだけでなく、開発過程も含めたオープン化、コミュニティの醸成を目指す

Graniteのオープンソース化

Granite

IBMが提供するオープンモデル・ファミリー

Hugging Faceで提供
huggingface.co/ibm-granite

 Language

 Code

 Time series

2024年6月リリース予定

 Geospatial*

ENGLISH
Granite-7B-Base
INSTRUCTION-TUNED
Granite-7B-Instruct

BASE
Granite-34B-Code-Base
Granite-20B-Code-Base
Granite-8B-Code-Base
Granite-3B-Code-Base

INSTRUCTION-TUNED
Granite-34B-Code-Instruct
Granite-20B-Code-Instruct
Granite-8B-Code-Instruct
Granite-3B-Code-Instruct

Granite-TimeSeries-TTM
Granite-TimeSeries-PatchTSMixer
Granite-TimeSeries-PatchTST

EARTH OBSERVATION
Granite-EarthObservation-HLS-Biomass
Granite-EarthObservation-HLS-CanopyHeight
Granite-EarthObservation-HLS-Landslide

WEATHER AND CLIMATE
Granite-WeatherClimate-Precip-Downscaling
Granite-WeatherClimate-WindForecasting

* Derived from IBM-NASA/Prithvi models

18モデル(2024/6時点)のオープンソース化を発表。

LLM開発の民主化を実現するInstructLab発表



オープンモデルの事後学習にて質問・回答を効率的に学習するためのオープンソースの仕組み

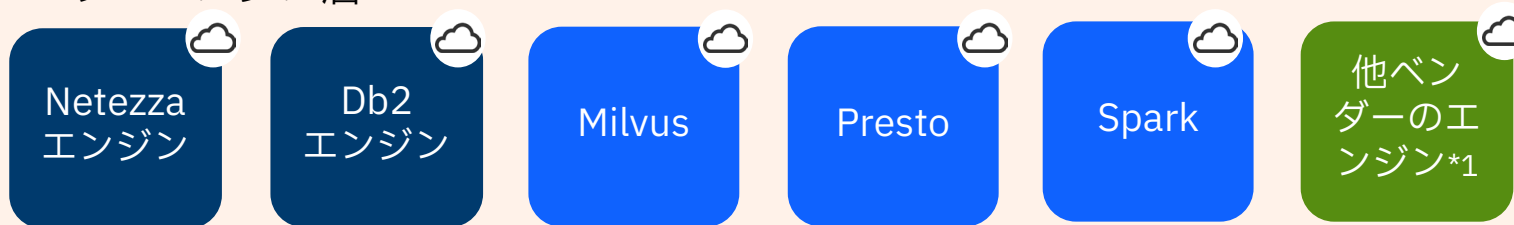
これまで追加学習のハードルとなってきた学習データの収集・管理が容易となり、誰でもLLM開発への貢献、ビジネスニーズに応じた追加学習が可能に。

② データ管理プラットフォームの例

watsonx.data: IBMの第2世代レイクハウス

クエリーエンジンや稼働環境の選択肢を増やし、ガバナンス機能も提供。
単一のデータレイクで集中管理を行い、ワークロードに適したクエリーエンジンの選択でコストを最適化。

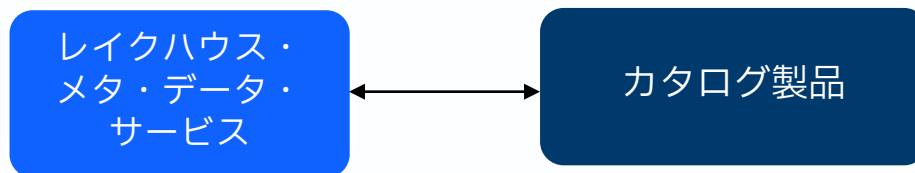
クエリーエンジン層



ワークロードに応じたコストの最適化

- 高性能な処理性能が求められるBIクエリにはDWHエンジンを使用
- 一定の時間内に収まればよいバッチ処理にはPresto等を使用

カタログ層



ガバナンス向上とデータ民主化

- 機密データのマスキングやアクセス制御
- データ分析者はメタデータや自然言語処理によって、データを即座に発見

データ形式・ストレージ層



ストレージのコスト削減・管理容易性

- オブジェクト・ストレージにデータを集めて利用することでストレージコスト削減

*1 icebergをサポートしていることが前提

② データ管理プラットフォームの例

watsonx.data: ベクトル・データベース

推奨システムにベクトル・データベースを導入することで、高速かつ効率的な類似検索を実現し、ユーザーにより適切に推奨できます。



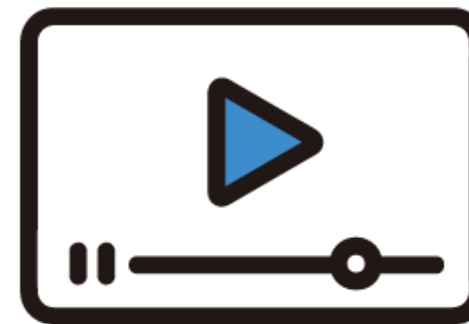
ECサイト

↓
購入単価・購入率の向上



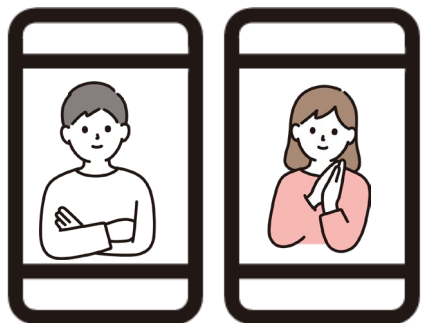
旅行メディア

↓
嗜好に沿った提案による予約率向上



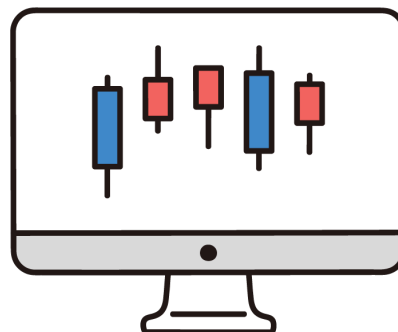
動画配信サービス(メディアエンタメ)

↓
解約の防止



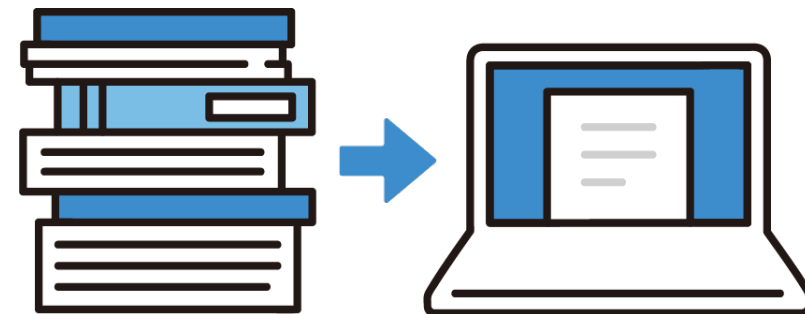
人材マッチング

↓
マッチング率の向上



金融

↓
金融商品の提案



社内文書検索

↓
知識継承・ナレッジ取得

③ガバナンスプラットフォームの例

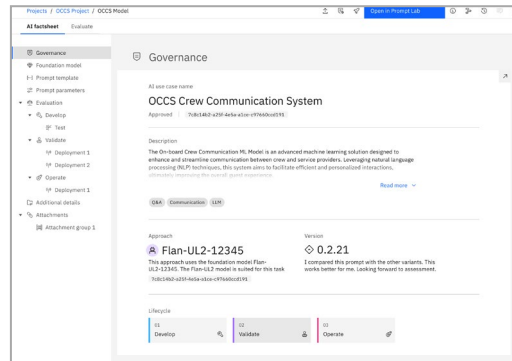
watsonx.governance

- 責任ある、透明で説明可能な AI を構築するための、統一されたAIガバナンス・プラットフォーム

ライフサイクル・ガバナンス

AI モデルの：

- 管理
- 監視
- 制御



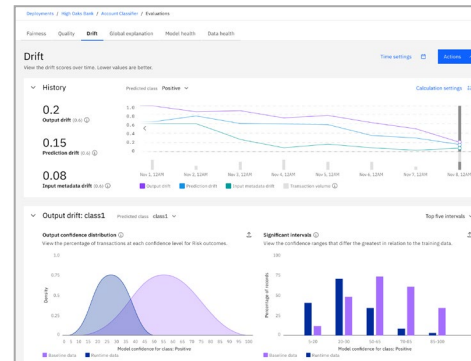
包括的

各段階でメタデータを取り込み、
AI ライフサイクル全体を制御

リスク管理

挙動の監視

- 公平性
- バイアス
- ドリフト（性能の劣化）
- LLMメトリクス

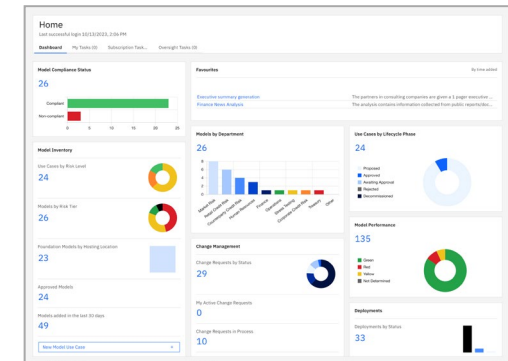


オープン

サード・パーティーのツールで構築・デプロイされたモデルのガバナンスをサポート

コンプライアンス

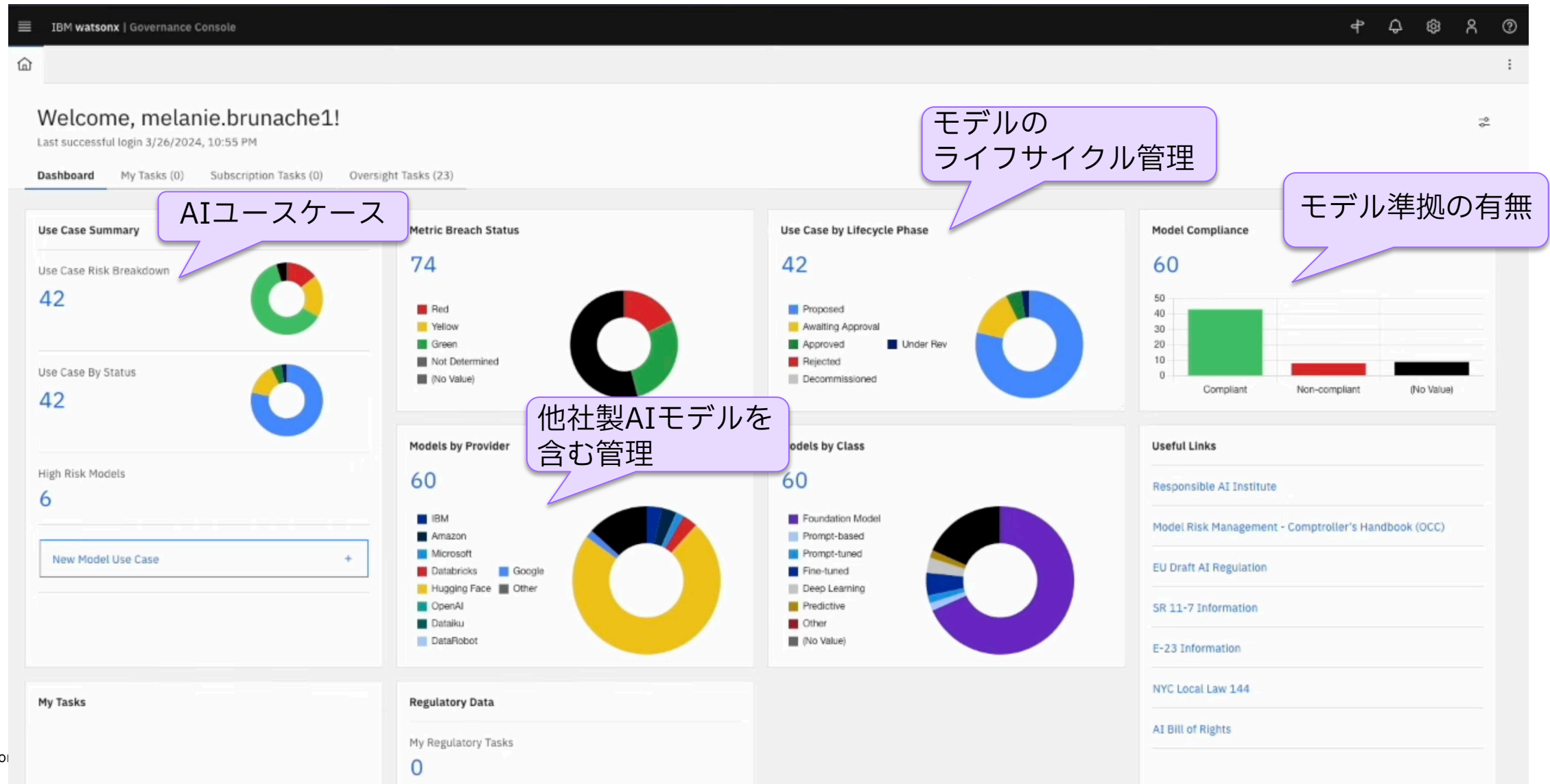
今後、世界中で施行される
安全性と透明性に関する
規制や政策に対応



メタデータの自動記録 およびPython
ノートブックを通じたデータ変換/
リネージュの取得

③ガバナンスプラットフォームの例

watsonx.governance ダッシュボードによる可視化



本日の内容

- 1 生成AIを前提としたビジネス設計へ
- 2 ビジネスのための生成AI
- 3 ビジネスのための生成AIを支えるプラットフォーム
- 4 生成AI活用によるビジネス変革
- 5 まとめ

生成AI活用によるビジネス変革

顧客接点、顧客体験	人事・財務・SCM	IT開発・運用	コア業務オペレーション
カスタマーサービス お客様が自ら簡単な手段で 問い合わせの解決策を発見	人事 採用、配置、人材育成におけ る手作業の削減と自動化	アプリケーション開発 コード生成、リバースエンジニ アリングの効率化	脅威管理 インシデント対応スピード の向上
マーケティング パーソナライゼーションの 向上、コンテンツサプライ チェーンの改善	サプライチェーン Source to Payプロセスの自 動化、リソースニーズの削減 、サイクルタイムの改善	ITの自動化 自動化処理開発の生産性向上、 デプロイメント課題の特定、障 害検知・対応の高度化	資産管理 持続可能な成果を確保しつつ、 資産パフォーマンスを最適化
コンテンツ生成 例) デジタルでのスポーツ 視聴体験の向上	計画と分析 意思決定最適化、ワークフロ ーの自動化による高付加価値 作業への集中	AIOps コスト効率を意識した継続的な ITシステムパフォーマンス管理	製品開発 例) AIで分子表現から構造を 推測することによる創薬の迅 速化
ナレッジワーカー 意思決定の改善、生産性 向上による高付加価値作 業への集中	規制コンプライアンス 要件、リスクに基づくコンプ ライアンス準拠の支援、規制 変化へのプロアクティブな対 応	データエンジニアリング 生成AIを活用したデータ統合ア プローチの再設計	環境インテリジェンス 悪天候や気候の影響を積極的 に計画・管理するためのイン テリジェンスの提供

IT業界の場合

環境変化

- 生産年齢人口減少
- IT需要増加

- 人材の流動化
- 有識者の高齢化

- 社会ニーズの多様化
- 技術進化

IT業務への変革要請

エンジニア人材不足
→ 工数依存からの脱却

自社システムを熟知していない
エンジニアが増加
→ 有識者依存からの脱却

新たな社会ニーズ・技術進化への
追従が必要
→ 研鑽余力の確保

必要なビジネス変革

スピード向上と工数削減・適正化

有識者の知見を組織内で蓄積・活用

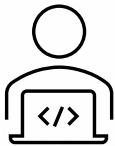
新たな領域での価値創造

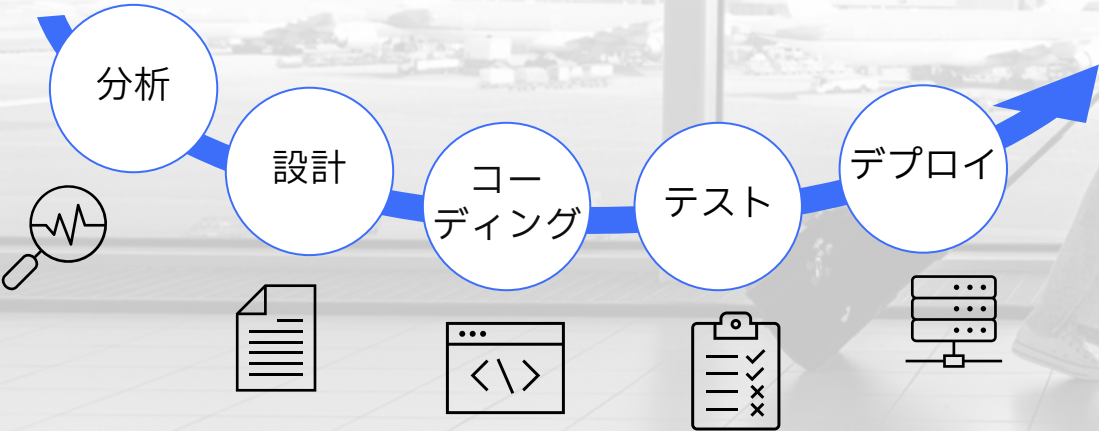
生成AI



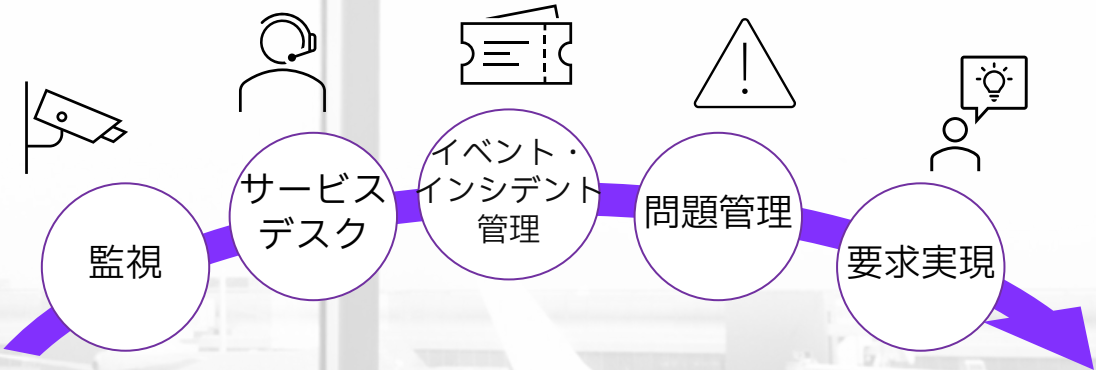
生成AIを活用したIT開発・運用の変革

システム構築 モダナイゼーション

 **30%**の生産性向上



IT運用の自動化



 **30%**の生産性向上

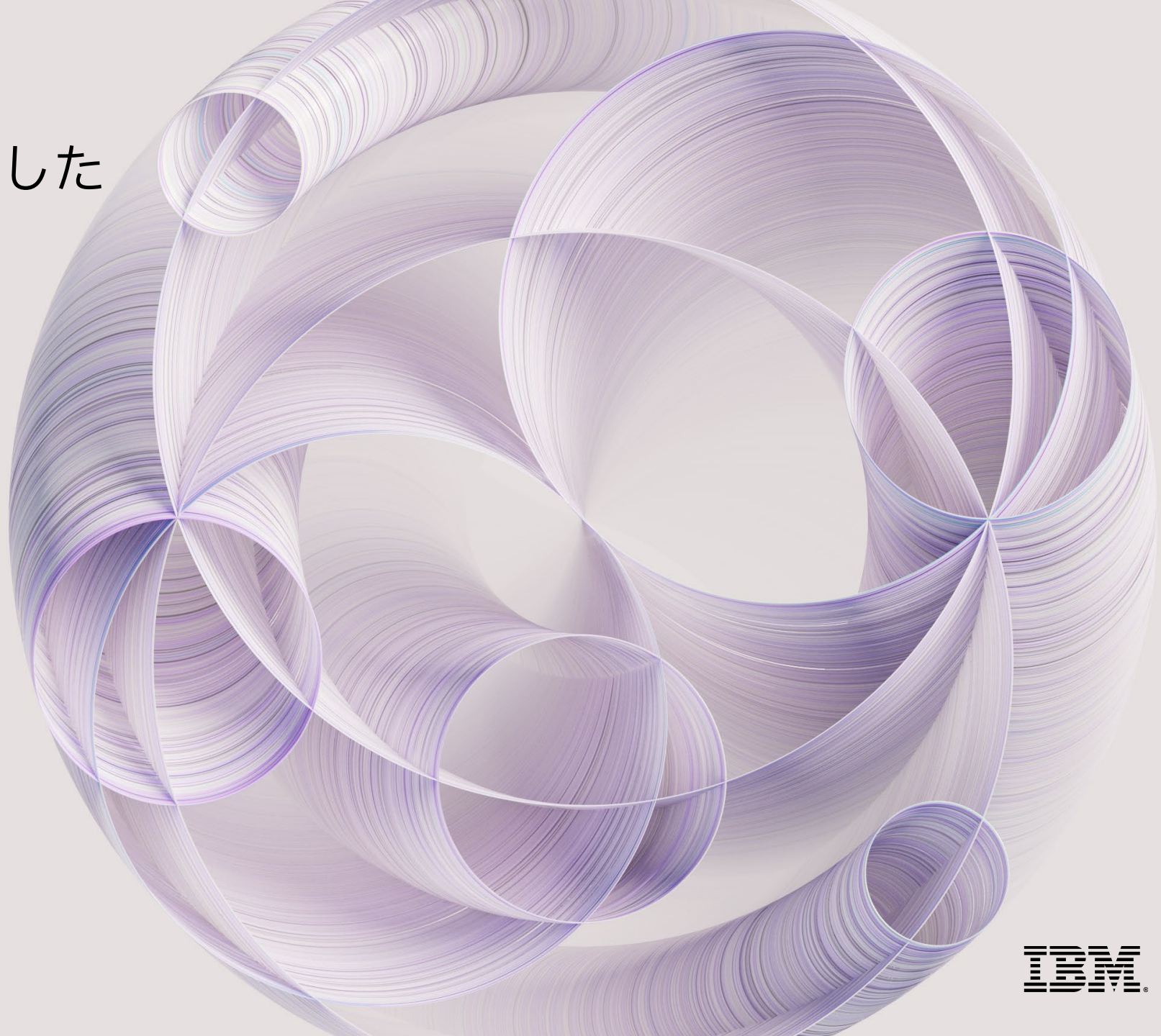
本日の内容

- 1 生成AIを前提としたビジネス設計へ
- 2 ビジネスのための生成AI
- 3 ビジネスのための生成AIを支えるプラットフォーム
- 4 生成AI活用によるビジネス変革
- 5 まとめ

まとめ

- AIファースト: 生成AIを前提としたビジネス設計へ
- ビジネスのための生成AIには、3つの視点を踏まえたAI活用の推進が重要
 - ① ビジネス文脈の理解
 - ② プロセスへの組み込み
 - ③ 信頼性の確保
- これからの企業はAI の価値創造者になる
 - 自社データを使い信頼できる独自のAIを構築することで、新たなビジネスを創出する
- ビジネスのための生成AIを支える3つのプラットフォーム
 - ① 生成AIの活用 
 - ② 自社データの管理 
 - ③ 信頼性確保のためのガバナンス 
- 多岐にわたる生成AI活用によるビジネス変革

ご清聴ありがとうございました



ワークショップ、セッション、および資料は、IBMまたはセッション発表者によって準備され、それぞれ独自の見解を反映したものです。それらは情報提供の目的のみで提供されており、いかなる参加者に対しても法律的またはその他の指導や助言を意図したのではなく、またIBM製品やサービスがお客様に適用ある特定の法令に適合することを保証するものでもありません。本講演資料に含まれている情報については、完全性と正確性を期するよう努めておりますが、「現状のまま」提供され、明示または黙示にかかわらず、商業性、特定の目的への適合性、非侵害性を含め、いかなる保証も伴わないものとします。本講演資料またはその他の資料の使用によって、あるいはその他の関連によって、いかなる損害が生じた場合も、IBMは責任を負わないものとします。本講演資料で言及されるIBM製品、プログラム、またはサービスは、IBMがビジネスを行っているすべての国・地域でご提供可能なわけではありません。本講演資料で言及される将来の展望（製品リリース日付や製品機能を含む）は、市場機会またはその他の要因に基づいてIBM独自の決定権をもっていつでも変更できるものとし、将来の製品または機能が使用可能になること、もしくは特定の結果を確約することを意図するものではありません。本講演資料は、言及される IBM製品またはサービスに適用ある契約条件を変更するものでも、追加の表明または保証を意図するものでもありません。

本講演資料に含まれている内容は、参加者の活動によって特定の結果が生じると述べる、または暗示することを意図したものでも、またそのような結果を生むものでもありません。パフォーマンスは、管理された環境において標準的なIBMベンチマークを使用した測定と予測に基づいています。ユーザーが経験する実際のスループットやパフォーマンスは、ユーザーのジョブ・ストリームにおけるマルチプログラミングの量、入出力構成、ストレージ構成、および処理されるワークロードなどの考慮事項を含む、数多くの要因に応じて変化します。したがって、個々のユーザーがここで述べられているものと同様の結果を得られると確約するものではありません。記述されているすべてのお客様事例は、それらのお客様がどのようにIBM製品を使用したか、またそれらのお客様が達成した結果の実例として示されたものです。実際の環境コストおよびパフォーマンス特性は、お客様ごとに異なる場合があります。

IBM、IBM ロゴ、ibm.com、Granite、IBM watsonx、watsonx、watsonx.ai、watsonx.data、watsonx.governance、watsonx Orchestrate、watsonx Assistant、watsonx Code Assistantは、世界の多くの国で登録されたInternational Business Machines Corporationの商標です。他の製品名およびサービス名等は、それぞれIBMまたは各社の商標である場合があります。現時点での IBM の商標リストについては、www.ibm.com/legal/copytrade.shtmlをご覧ください。

JavaおよびすべてのJava関連の商標およびロゴは Oracleやその関連会社の商標または登録商標です。