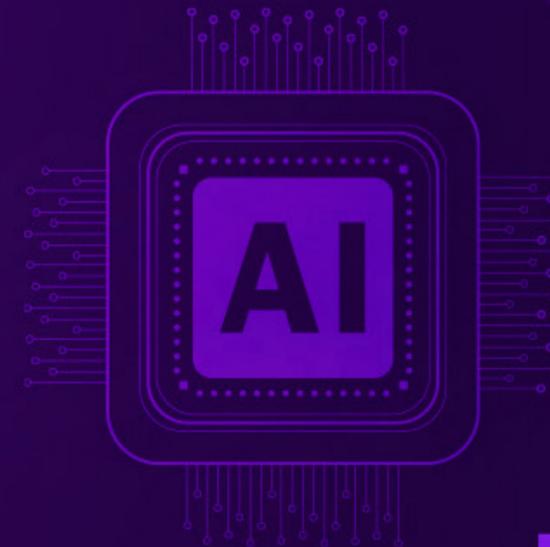


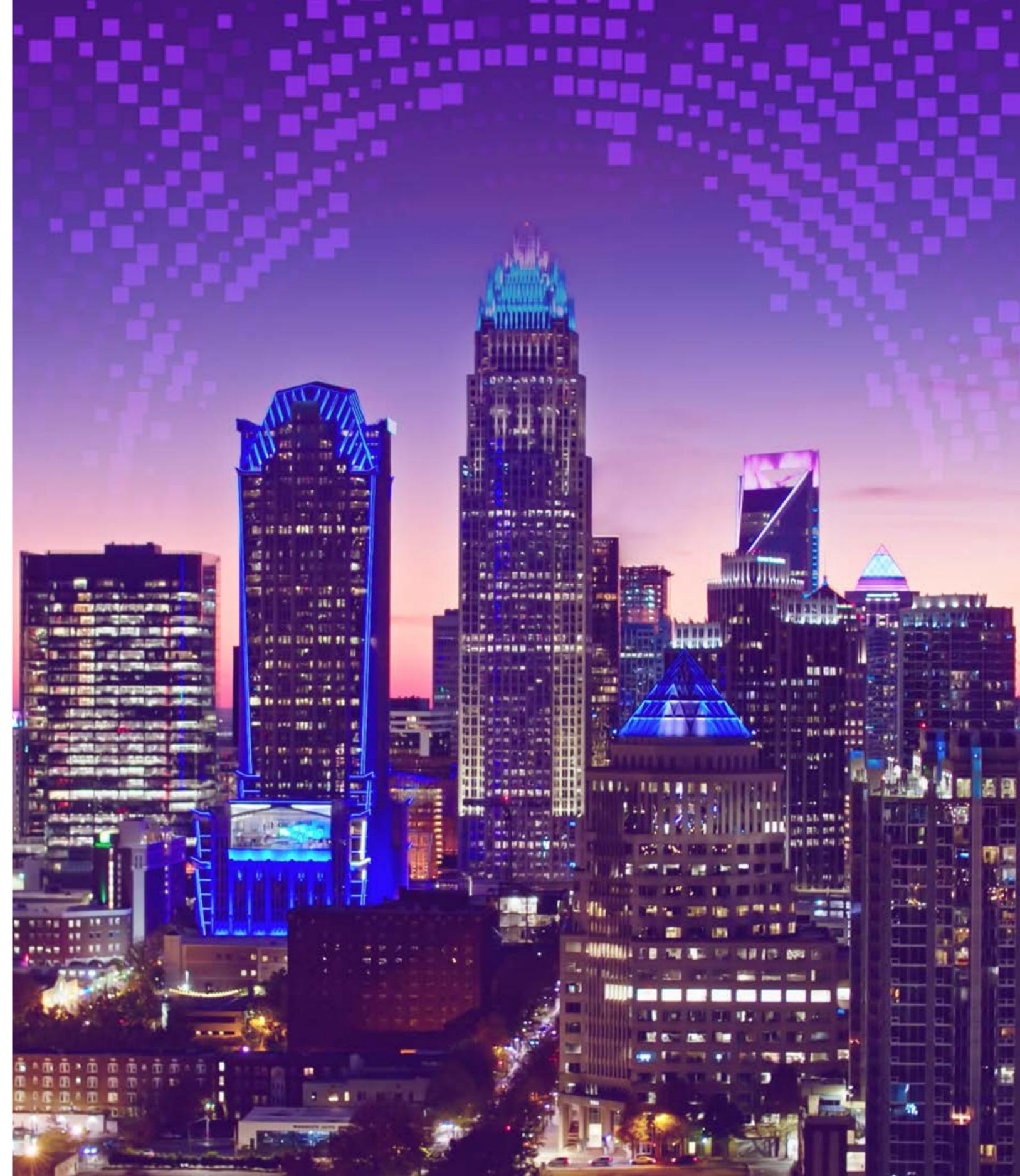
precisely

# 信頼できる AI 初級編: データを AI 対応にするためのヒント



# データの AI 対応が完了している企業は4%しかありません。

そのため、この eBook では、ビジネスにおける AI 実装を成功させるためには、データ完全性が必要であることを強調します。AI の効果を実現するには信頼性のあるデータが重要であることをまず明らかにし、データ完全性の恩恵を受けることができる AI の利用法をご説明し、Precisely と AWS の協力によって AI に関する望ましい結果を得ることが可能になることをご提唱すると共に、AI の可能性をフルに発揮させるために、組織はデータ完全性を最優先とするべきとご提案します。



# AI の取り組みを成功させるためには、 信頼できるデータが必要です

現代では過去にないレベルで競争が激化しているため、企業が人工知能（AI）の力を先を競って利用しようとしていることは容易にご理解いただけるでしょう。その用途は、生産性と効率の向上、お客様に対する個別化されたサービスの提供、さらに競争上の優位を得るような新しいアイデアの創出などがあります。

しかし、適切な準備なしに AI に飛びつくことにはリスクもあります。2023 年には、[AI が書いた要約に誤った参照先が含まれていた](#)り、著名なコンサルティング会社が[存在しないスキャンダル](#)に巻き込まれたり、その他多くの AI の失敗がありました。

どの場合でも、誤った結果をもたらしたモデルは、目的に対して不適切なデータを学習に使用しており、その結果、歪んで誤った出力を生み出しました。このことは、AI の基盤として信頼できるデータが必要であることを強調するものとなっています。

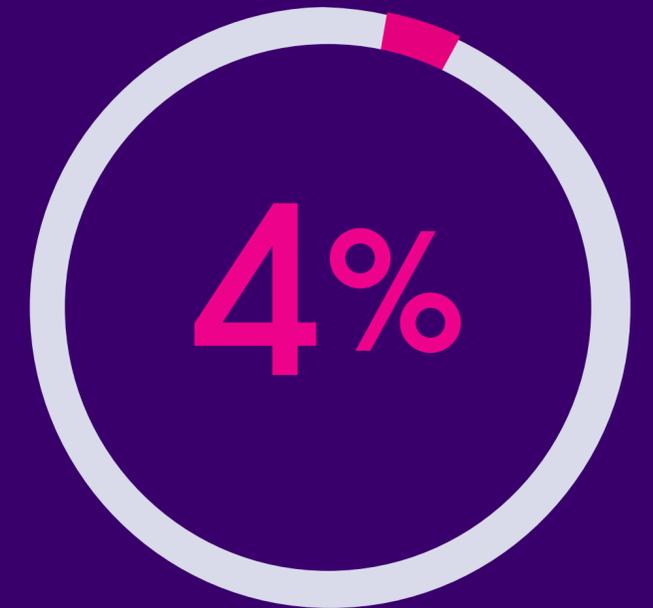
そして、信頼できるデータにはデータ完全性が必要です。言い換えれば、最大の正確性、整合性、そしてコンテキストを持ったデータです。こう考えてみましょう：AI 出力は、そこに入力されたデータ以上の強さを持つことはできません。

しかし、データ完全性にはしばしば障害が伴います。たとえば、データを十分な速度で統合すること、データの責任ある利用について理解しガバナンスを確保すること、データの品質を監視して改善すること、コンテキストをより高めるためにエンリッチメントを行うこと、さらに、セキュリティとプライバシーを保障すること、などです。

AI アプリケーションが、信頼性のある、AI に対応したデータを使って学習できることの利点は無限にあります。それでも、多くの企業にとって現在の現実を考えると目が覚めます。

2023 年の Gartner IT シンポジウム [リサーチ](#)でのスーパー グループ インタビューによると、データの AI 対応ができていると解答したのはわずか4%でした。

以上のすべてのことを前提として、成功と信頼性のある結果を確実に得るために必要な、AI の利用方法とデータ完全性の基本を見ていきましょう。



**目が覚める現実：**  
データが既に AI 対応  
になっていると回答  
した人は、わずか4%

# 稼働中の AI:データ完全性で成功した6つの利用方法

AI アプリケーションは、以前は自動化と予知分析の専用とされてきました。生成 AI を使えば、新しく、想像力に溢れる利用方法を推進して、競争的優位を大きく強化できるような、コンテンツ、アイデア、そしてデータなどを創造することができます。Fortune/Deloitte による 2023 年10月の研究によれば、CEO の79%は、「技術革新を加速することが、生成 AI 実装の最大の利用方法である」と回答しました。

自然言語に基づいた直感的な企業データへのアクセスが実現されることによって、以前には実現が難しかった利用方法が現実となり、豊穡な新しい可能性がようやく開かれようとしています。



# 1 利用方法: AI による推奨

## 利点

より高速で、より個別化された推奨

## 方法

AI による推奨システムとは、AI と、たとえば過去の選好、振る舞い、データ入力状況などの膨大なユーザー データを利用して、製品、コンテンツ、サービスなどについて個別化された推奨を行う、非常に高度な技術です。

## データ完全性が重要な理由

データが完全性、特に、データ品質の重要な要素を備えている場合にのみ、正確で関連性の高い推奨を行うことができます。データ完全性は、モデルの効果的な学習にも役立ちます。結果として、推奨の性能が上がり、ユーザー満足度が向上して、エンゲージメントレベルが向上します。

# 2 利用方法: AI に基づくワークフロー

## 利点

生産性向上、効率の向上、およびコストの低減

## 方法

ワークフローは、リアルタイムでデータを処理する AI モデルによって自動化できます。営業マーケティング キャンペーン、プロジェクト管理ワークフロー、プログラミング補助など、さまざまに応用することが可能で、その利点は明らかです。90,000 社以上の開発会社を対象とした Stack Overflow 調査によると、ほぼ70%が AI ツールを利用しているか、利用を計画しています。

## データ完全性が重要な理由

統合システムは複数のデータ ソースを利用するのが普通で、そのため、複雑な取引システムを含む関連するすべてのソースから必須のデータをまとめなくてはなりません。データ品質とガバナンスを実践することによってデータの信頼性は向上し、エラーは提言され、正確な結果を確実に得られるようになります。その結果、ワークフローも効率的になります。円滑な運営と高品質な結果を得るためには、データ完全性が非常に需要です。

# 3

利用方法:

## 機械学習 (ML) アプリケーション

### 利点

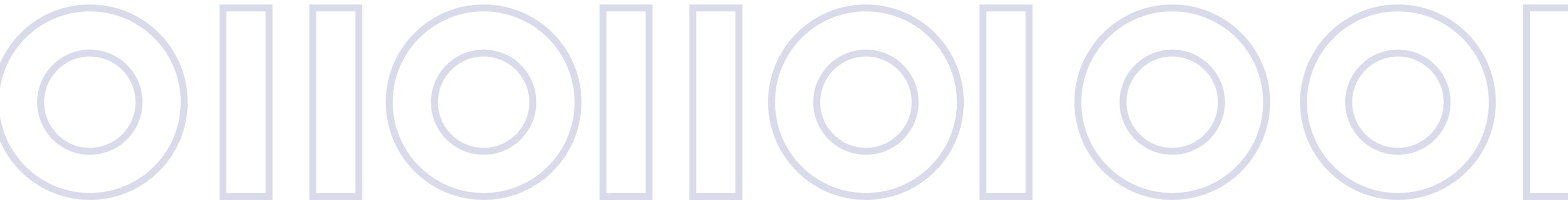
正確性の向上によってビジネスプロセスを加速します。

### 方法

ML アプリケーションは、コンピューターがデータを使って学習し、迅速に価格見積を作成するなど自律的に予測や意思決定を行って、顧客満足度を向上させることができます。しかしそれも、学習に完全性が高いデータを使用した場合にのみ実現可能です。

### データ完全性が重要な理由

データの準備にかかる時間を劇的に削減するためには、データを統合し、高い品質とガバナンスを維持して、エンリッチメントと空間分析を通じてコンテキストを付加する必要があります。それらのステップがすべて合わさってフレッシュなデータが得られるようになり、そうになると ML モデルの出力が信頼できるものになります。



# 4

利用方法:

## 基礎モデル (FM) 学習

### 利点

自然言語処理の機能があると、基礎モデル (FM) で、コンテンツやプログラムを生成したり、テキストを要約したり、感情を分析したり、質問に答えたり、その他さまざまなことを行うことができますようになります

### 方法

基礎モデル (FM) とは、巨大なデータセットで事前に学習を行った ML (機械学習) モデルで、一般的なパターンや特徴をとらえることができるように設計されています。

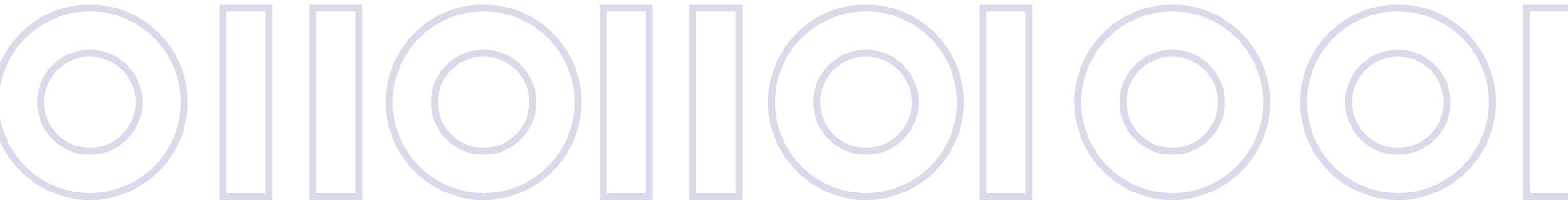
しかし、FMに関する大きな課題として、学習バイアスの可能性の問題があります。たとえば、マイノリティーが所有するか、その自宅を基盤に行うビジネスに対する貸付申込を処理する、グローバルな銀行

で使用される FM は、データに社会的な偏見、ステレオタイプ、不均衡など、固有のバイアスが含まれている場合にはリスクを負います。

FM が文章を生成したり回答を作成したりする時に、既存の社会的、性的、もしくは人種的なバイアスを不注意にコピーして増幅し、結果として差別的な出力を行い、さまざまな分野での不公平を悪化させてしまう可能性もあります。

### データ完全性が重要な理由

バイアスを避けるためには、複数のシステムから抽出した関連データセットの中から注意深く精選され、品質を確保できるよう前処理され、すべてのグループを正確に表すサードパーティーの属性データでエンリッチメントされたデータによって、生成 AI に学習させる必要があります。



# 5 利用方法: チャットボット

## 利点

ユーザーのエンゲージメントを向上させる、効率的で個人化されたアシストを行います。

## 方法

大規模言語モデル（LLM）上に構築されたチャットボットは、ユーザーの問い合わせに対して自然で、コンテキストに富んだ反応を返すことができます。LLM とは、人間に似せた形で言語を理解して生成できるよう、膨大な量のテキストデータで学習させた、一種の FM です。

チャットボットは、一連の会話に基づいて動的に返答を生成する能力を持っている点で他のシステムとは異なり、多くの産業分野や、カスタマーサポートのような利用方法において、ユーザー エンゲージメントを強化できます。

カスタマーサービスにおける生成 AI のインパクトは、既に感じられています。全米経済研究所（National Bureau of Economic Research）による 5,179人のカスタマーサポート担当者を対象とした調査では、AI ツールを使用した場合は平均で14%生産性が向上したとの回答がありました。その数字は、非熟練の担当者では34%にまで上昇します。

## データ完全性が重要な理由

チャットボットの回答の品質を上げるには、LLM に高品質で完全なデータを使って学習させる必要があります。FM は、回答を生成するにあたってデータをそのまま利用します。問題点を修正することはしません。生成 AI モデルに対して、すべての関連するデータを提供し、AI モデルが正確で、整合性があり、コンテキストを持っているようにするには、データ完全性が重要です。



# 6

利用方法:

## RAG を使用した AI アシスタント

### 利点

AI アシスタントに対する、コンテキスト理解の強化

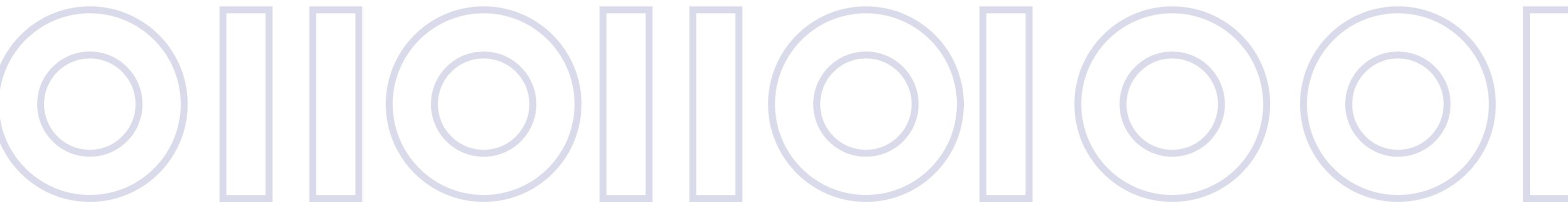
### 方法

AI アシスタントは検索拡張生成（RAG）を使用して構築できます。RAG とは、FM と既存の知識を組み合わせることによって、モデルの性能を向上して出力する結果をより関連性が高く役立つものにするものです。RAG を使用したシステムは質問に答えるようなシナリオに優れていて、莫大なナレッジベースから文脈に合った詳細な回答を提供します。

### データ完全性が重要な理由

データがモデルの反応の原動力ですので、そのデータの完全性が本質的に重要になります。不正確なデータは間違った回答につながります。同様に、不完全もしくは品質の低いデータも、もっともらしいウソの回答を生み出します。

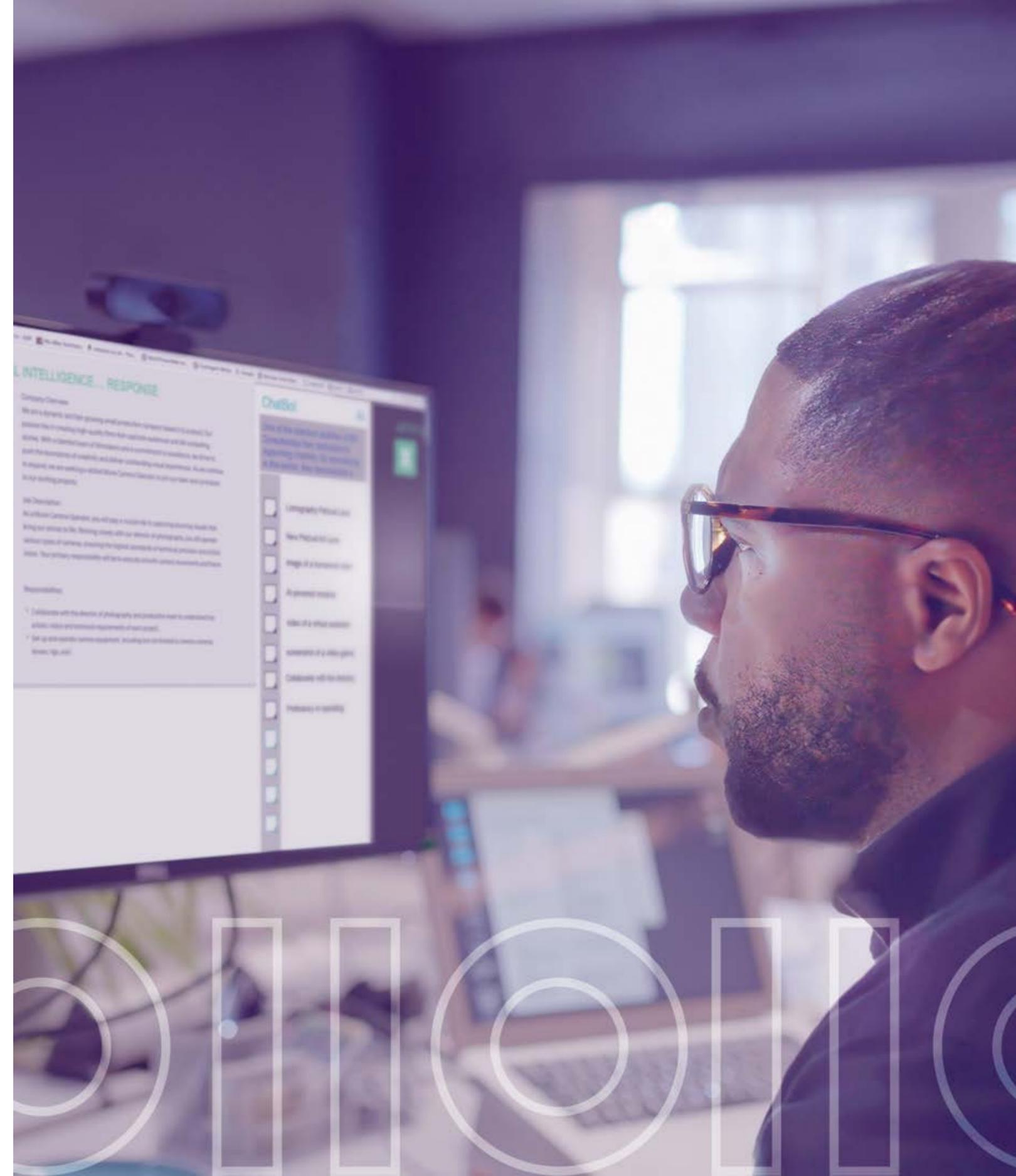
不動産や保険などの産業における、その領域独自のシナリオに関しては、サードパーティー データによるエンリッチメントを行うことでコンテキストを追加し、AI アシスタントの回答の独自領域対応性や関連性を強化することができます。たとえば、物件属性データを、不動産資産の説明文を生成する RAG ベースの生成 AI アプリケーションに統合するようなことも可能です。



## 最大の AI に関する課題をデータ完全性で解決する

不完全なデータ、コンプライアンス不全、AI 出力のコンテキスト不足…これらの、および似たような課題に、お心当たりはあるでしょうか?朗報として、もしお心当たりがあるようであれば、それらは解決可能です。

上で見てきたような AI の豊かな恩恵を受けるために、堅牢で個別の条件に合ったデータ完全性を実現できるように、積極的に投資を行うアプローチを取ることができます。よくある課題を、より詳細に見てみましょう。

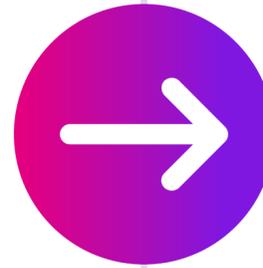


## 課題

# 視野が狭く、バイアスがかかった結果

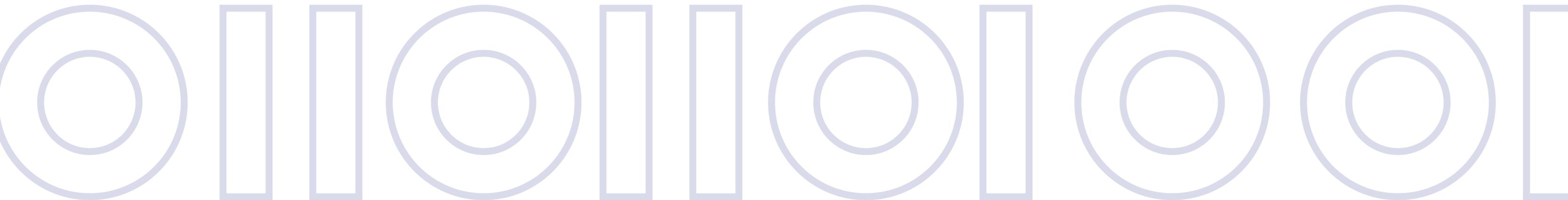
データセットが不完全で、データ インフラ スタックが分割されている場合、質問に対する AI の理解は限定的となり、回答結果もバイアスがかかった信頼性に欠けるものとなります。

企業データはさまざまなレガシー システムで生成されて保存されるので、AI が実行される（クラウドの）プラットフォーム上で、関連する不可欠なデータがすべて利用できるわけではありません。データの一部にしかアクセスできない場合、結果にはバイアスがかかる可能性があります。



## ソリューション データ統合

分離状態にあるデータを統合して、AI アプリケーションが実行される場所で利用可能とすることが、バイアスのない、信頼性のある結果を得るためには重要です。そうすることで、総合的に、完全に、適切なタイミングで、AI が必要とする場所ですべての関連するデータを確実に利用できるようになります。



## 課題

# 信頼できない結果

AI の予測や推奨が不正確だと、信頼性が失われてこれらの技術の普及を妨げる結果につながる可能性があります。これは、データ品質が低いことによって起きた結果の一つです。

AI が利用するためには、正確で、整合性があり、目的に沿ったデータが必要であり、そのためにデータ品質の分野で再検討が促されています。信頼性のあるデータでなければ、先進的な AI モデルも意味がありません。期待値は大きく、データ品質の障害を克服することは優先事項となっています。

さらに、各国政府は、データ プライバシー保護規制の遵守をより厳しく求めるようになっており、そのために企業もデータ主権性関連法令の遵守を迫られています。



## ソリューション

# データ品質とガバナンス

AAI に関する取り組みのためには、データ品質に関して、正確で、整合性があり、目的に合致しているデータを確実に使えるような、新しいアプローチが必要となります。それは時に、データ品質とビジネスルール、自動化された検証とクレンジング、さらにデータ可観測性とデータガバナンス ソリューションとの統合の形式をとります。

予防的データ品質ツールは、データ パイプラインを監視し、先進の ML（機械学習）技術を使用することによって、異常と外れ値を迅速に検出します。さらに、AI を利用して、問題が下流のシステムに及ぶ前に、確実に対処されるようなルールを作り出すよう提案します。

データ ガバナンスは、AI アプリケーションで使用されるデータがどのように処理されるか、明快な理解を提供します：

- 収集: データ タイプとは何でしょうか？
- 保存: 場所はどこでしょうか？
- 利用: 誰がアクセスできるでしょうか？

このように理解を統一することによって、データを保護するためのポリシーと手順を適用すること容易になります。

センシティブなデータを防護し、評判を保護し、AI アプリケーションから最大の利益を得るためには、データ セキュリティやコンプライアンスに対して予防的なアプローチが必要です。堅牢なデータ ガバナンスを実装し、最新の規制についていけることが必須です。

## 課題

# コンテキスト的な関連性の欠如

既存の現実世界シナリオのニュアンスと相互依存性に対するコンテキスト的な理解なしには、AI の推定や提言は、全体の構図のごく一部だけに基づくものとなってしまいます。そうすると結果として、不完全で、不正確で、コンテキスト的な関連性の低い結果を出力することとなり、下流に危険な影響を及ぼす可能性があります

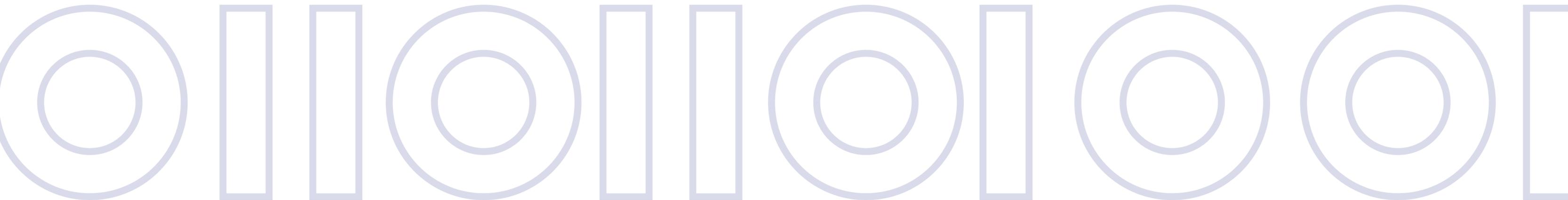


## ソリューション

# 空間分析とデータ エンリッチメント

AI 出力の正確性とコンテキスト性を向上させるためには、信頼できるサードパーティのデータと空間的な洞察によって、データ エンリッチメントを行わなくてはなりません。たとえば、住所の詳細と、物件ポートフォリオに伴う環境的なリスク要因を組み合わせ、予測モデルの精度と自然災害に対する洞察を向上させたいとします。

そのプロセスには、データを現実世界シナリオに正確にマッピングできるような、空間分析とデータ エンリッチメントを通じて行う必要があります。正確かつ関連性の高いデータで学習を行った AI および ML を利用することで、より大きく、より良い結果が得られます。



# Precisely & AWS : AI 向けデータ完全性に関する3つの考慮すべき点

多くの経営者が、新しい技術を自社のビジネスモデルに統合することは、最優先の戦略であると考えています。そして、現在の焦点は、インフラを AI を通じてどのように変革するか、ということです。

ただし、不確実性に基づく AI アプリケーションでは、技術的な入力方法によって異なる結果が出るため、幻想に終わる傾向があります。総合的なデータ完全性こそが、信頼の構築を成功させます。

AI アプリケーションを、正確で、整合性があり、コンテキストがあるデータで学習させることが、成功を加速させるような、信頼性のある結果を得るためには重要なことです。そして、AI に対するデータ完全性については主に3つの考慮すべき点に至ります。以下に、Precisely と Amazon Web Services (AWS) を合わせて利用することで、望ましい結果を得る方法と、そのために知っておくべきことをご説明します。



1

# より完全なデータセットがあれば、AI の十全な可能性の実現を支援できます

## データ完全性で実現できること

オンプレミス、クラウド、そしてハイブリッド環境にあるすべての関連する重要なデータで AI を学習させることによって、バイアスを最小化し、精度と信頼性を改善し、理解を増進することができます。そこには、メインフレーム コンピューターに保存された複雑なデータも含まれます。

## 実現する方法

分散化されたデータの垣根を取り壊して、新鮮なデータを AWS サービスに迅速に転送しましょう。Precisely のモダン データ パイプラインならそれができます。機械学習と推論に最適化された AWS インスタンスを持った、コンピューターの最も広範囲かつ最も奥深い機能と、スケーラブルなクラウド ストレージをご覧ください。大量のデータセットから価値を引き出すクラウド リソースを社内に提供することで、AI をビジネスに役立てることができます。



## 2

# AI アプリケーションに信頼できるデータを与えることで、信頼できる結果が強化されます

## データ完全性で実現できること

AI の結果は、確実に信頼できるものでなくてはなりません。正確な予測、提言、および効果的なプロセス自動化のためには、クラウドネイティブなソリューションで得られた、完全性のあるデータでモデルに学習させる必要があります。

信頼できる AI 出力を得るためには、データは厳密な品質基準を満たしてはなりません：正確で、完全で、有効な構造を持ち、標準化され、重複がないものである必要があります。完全性が高いデータは、さらに、タイムリーで、堅牢なフレームワークを使ったガバナンスに支えられ、変更や異常値に対して監視されていなくてはなりません。

## 実現する方法

パイプライン上のデータシステムに透明性があれば、データの品質を向上させて監視し、さらにデータと AI モデルにガバナンスを提供できます。このプロセスは Precisely Data Integrity Suite を使って合理化できます。それは、Amazon Redshift のような運用および分析システム全体にわたって、AWS サービス内でネイティブに動くデータ完全性プロセスです。モデル学習、実装、推論、および監視について完全性のあるデータを使用することで、AI の出力に対する信頼を構築することが容易になります。

Amazon SageMaker および Amazon Bedrock 内の ML および生成 AI モデルが学習と微調整を行う際に、信頼できるデータを使用することが不可欠です。自動化されたデータ完全性プログラムがあれば、データに対する理解を改善し、品質を向上するために必要なステップを自動化できます。Precisely Data Integrity Suite のサービスは、この目標を実現するために必要なツールをご提供します。

モデル学習に必要な大量のデータを取り扱う時には、伸縮自在かつコスト効率が高い、スケーラビリティを持ったクラウド ネイティブなアプローチが必要です。そして、データ品質の継続性を保証するためには、パイプライン全体の可視性を提供できるような、監視と予防的な異常値検知が必要です。たとえば、警告が自動化されていれば、悪いデータが Amazon Bedrock もしくはその他のサービスの学習用に送られることを防止できます。



3

データにコンテキストを追加することで、より関連性が高く、ニュアンスに合った回答が得られます

## データ完全性で実現できること

信頼できるサードパーティーのデータと空間洞察をデータに付加してエンリッチメントを行うことで、AI アプリケーションの精度とコンテキスト的な関連性の向上を加速しましょう。

## 実現する方法

AI アプリケーションの開発とその普及を加速するためには、データにコンテキストを付加して、AI がニュアンスを把握し、整合性を維持し、コンテキスト的に関連性が高い回答ができるようにしましょう。

これは、Precisely がご提供するサードパーティー データと空間洞察を、Amazon SageMaker と Amazon Bedrock 環境に導入することで実現できます。このような、精選されて信頼できるデータセットがあれば、場所、人々、物件、ビジネス、および環境リスク要因に関して既知のことや、AI が学習したこと、さらに詳細を付加することができます。



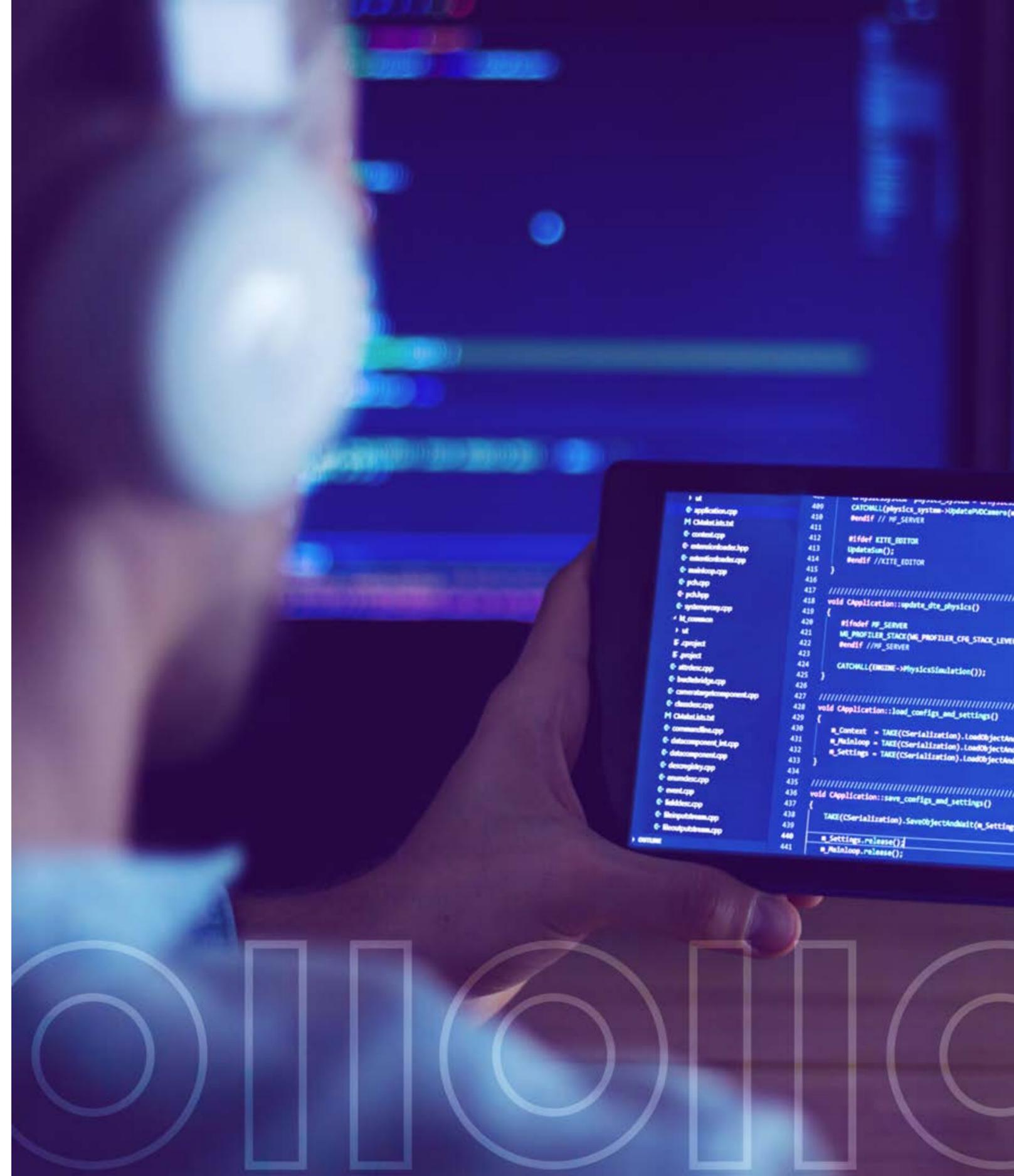
AWS は、AI を取り入れる企業のあらゆるステージにおいて、総合的な AI および ML（機械学習）サービス、インフラ、および実装リソースをご提供します。完全性が高いデータとこれらのサービスを利用して AI アプリケーションを構築することによって、AI の普及を加速し、アプリケーションを正確で関連性が高い状態に保つことができます。

Precisely Data Integrity Suiteと、Amazon Web Service (AWS) のサービスを組み合わせることで、データを AI 対応として、AI ベースのソリューションの可能性を最大限引き出しましょう。

Precisely とAWS は、共同して不可欠なデータと、その最適な品質を保証することによって、信頼できる AI 出力を得られるよう支援します。最適な品質は、堅牢なフレームワークによるガバナンス、劣化を検出する監視、空間洞察とサードパーティデータセットから得られる重要なコンテキストによるエンリッチメントに支えられています。

これらを考慮することで、確実に高品質なデータを得られるようにすると共に、結果として提供される AI アプリケーションが完全な信頼性を持つことができるようになります。データ完全性の向上に注力することで、組織内での AI の開発と普及を加速することができます。

- 1. Judge sanctions lawyers for brief written by A.I. with fake citations**  
<https://www.cnbc.com/2023/06/22/judge-sanctions-lawyers-whose-ai-written-filing-contained-fake-citations.html>
- 2. Academics apologise for AI blunder implicating Big Four**  
<https://www.accountingweb.co.uk/tech/tech-pulse/academics-apologise-for-ai-blunder-implicating-big-four>
- 3. Fall 2023 Fortune/Deloitte CEO Survey Insights**  
<https://www2.deloitte.com/us/en/pages/chief-executive-officer/articles/ceo-survey.html>
- 4. Developer sentiment around AI/ML**  
<https://stackoverflow.co/labs/developer-sentiment-ai-ml>
- 5. Generative AI at Work**  
<https://www.nber.org/papers/w31161>



## まとめ

生成 AI アプリケーションの興隆に伴って、データ完全性を優先事項にすることはこれまで以上に不可欠となっています。

性能が高く、信頼できて、高品質の出力ができる AI を構築する取り組みにおいては、データは完全で正確で、信頼できて整合性があり、コンテキストを持っていなくてはなりません。そういったデータがあれば、途中で立ちはだかる障害を克服する手助けとなり、信頼と普及を醸成することも容易になり、AI の開発と革新も加速されます。

AI の最大の利用方法は何ですか?その可能性をフルに引き出したとしたら、何が達成できるでしょうか?個別の目標がどうであれ、信頼できる AI は信頼できるデータから始まることは心に留めておいてください。今、データ完全性への過程を始めることで、将来の AI がより有望になります

さあ、始めましょう

## AI リソースのためのデータ完全性

### 製品

[Precisely Data Integrity Suite](#)

### ウェビナー

[信頼できる AI: 開発ライフサイクル全体を通じてデータ完全性を重視する](#)

### ブログ

[Precisely と AWS を利用した信頼できる生成 AI](#)

[データ エンリッチメントを利用して AI を加速する](#)

[Precisely Softwareにおける AI の力](#)

[自動化の先には何があるのか](#)

[Precisely Data Integrity Suite で重要なデータの問題に取り組む](#)



## Precisely について

Preciselyはデータ完全性のグローバルリーダーとして、100か国以上、12,000社のお客様にデータの正確性と一貫性を提供しており、ご利用のお客様にはFortune100企業の99社が含まれています。Precisely のデータ統合、データ品質、データ ガバナンス、ロケーション インテリジェンス、データ エンリッチメント製品は、ビジネスのより優れた意思決定を促進し、より優れた成果を創出します。

詳しくは [www.precisely.com](http://www.precisely.com) をご覧ください。

©2024 Precisely. 本書で使用されている会社名およびプロダクト名は、各社の商標である可能性があります。