

AIでデータの価値を引き 出すための7つのステップ

効果的なデータ管理のための包括的なガイド

データで競争力を構築

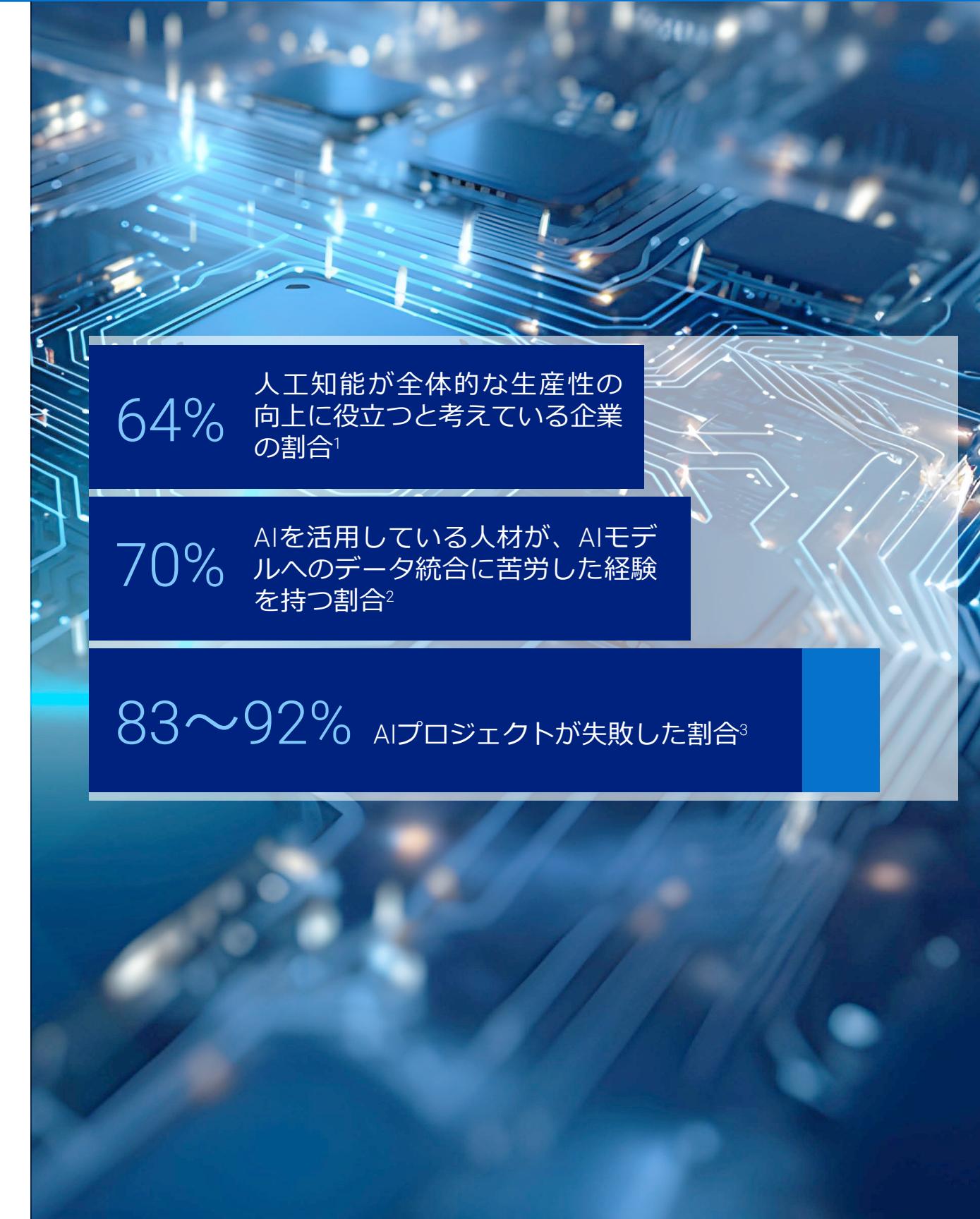
ビジネス環境が急速に進化している今日、データは重要なリソースであり、イノベーションと競争優位性の基盤です。堅牢で正確かつ包括的なデータセットを賢く使用して、ビジネス上の意思決定を迅速に下すリーダーは、差別化を図ることができます。これが最も顕著に現れるのが、[生成AI](#)の導入においてです。

生成AIは比類のない進歩と効率性を約束しており、完璧なデータはその創造的で反復的かつ柔軟な出力に必須ではありません。しかし、ある時点で、自分の「データハウス」が整っているかどうかを評価する必要があります。つまり、単にデータにアクセスするだけでなく、

強力なデータ管理の実践が必要です。これは、規模が大きくなるにつれて必要性がさらに増していきます。

強力なデータ管理は必要だが、簡単ではない

[AIのワークフロー](#)の導入は、データサイエンティストなどが適切なデータセットを特定し、そのクリーンさと包括性を確保するという複雑で要求の厳しいプロセスの集大成です。この取り組みは、技術的な課題を超えて、組織全体のプロセスと堅牢なデータ管理フレームワークを必要とします。



1 Forbes Advisor : [『How Businesses Are Using Artificial Intelligence in 2024』](#)

2 McKinseyレポート : [『The state of AI in early 2024』](#)

3 Fortune : [『Want your company's AI project to succeed? Don't hand it to the data scientists, says this CEO』](#)

データ管理工程

このような複雑な状況を乗り切るために、Dellはデータ管理工程の包括的な概要を7つの重要なステップにまとめました。これらのステップは、Dellのエキスパート データ サイエンティストがさまざまな組織を対象に実施する広範なワークショップとコンサルティングの集大成であり、拡張可能で効果的なAIモデルを作成するうえでの共通の課題と成功戦略を示しています。Dellは、今後数か月にわたり、各ステップをさらに深く掘り下げて、お客様が組織特有の障害を克服して、AI導入を実現できるよう支援していきます。

このe-bookでは、効果的なデータ管理の手法を理解して実装するための明確で実用的なフレームワークについて説明します。これらの原則を取り入れることで、生成AIの実験から規模を拡大し、データを未加工のリソースから戦略的資産へと変換して、AI主導の未来におけるイノベーションと持続的な競争優位性への道を切り開くことができます。

こちらのインフォグラフィックをダウンロードすると、これらの7つのステップを簡単に参照できます。



ステップ1

ビジネスニーズの特定



ステップ2

関連データ検出の迅速化



ステップ3

データの調査とアクセスのシンプル化



ステップ4

分析、ML実験、モデリングの最適化



ステップ5

データと分析の製品化の拡大



ステップ6

データ管理とガバナンスの自動化



ステップ7

ビジネス上の成果の評価

ステップ1

ビジネスニーズの特定



このステップでは、データ管理工程の方向性を定めて、戦略的なビジネスニーズに合わせて取り組みを調整します。明確な目標とメトリックがなければ、有意義な価値を得ることはできません。

- まず、運用の目標と、それによって得られる価値を理解するところから始めます。
- 望ましい成果と成功の測定方法について、組織全体で調整します。
- 創造される価値の明確なビジョンを確立します。

達成する**目標**とその**測定方法**を知ることで、その後のすべてのデータ管理作業が目的に沿ったものとなり、明確で達成可能な目標へと方向性が定められます。



ステップ2

関連データ検出の迅速化



明確なロードマップを使用して、関連データの検出を迅速化します。利用可能なすべてのデータが必要なわけではないため、データサイエンティストは、問題の解決に関連するデータを迅速に特定する必要があります。ソースのデータセットをカタログ化し、メタデータを作成することで、データとその価値の明確な関連性を確立します。この焦点を絞ったアプローチにより、データへの取り組みが効率的かつ効果的なものになります。

関連データを速やかに特定することで、**時間**と**リソース**を節約し、データサイエンティストは戦略目標に直接関連する影響度の高いデータに集中できるようになります。

68%

企業が利用できるデータのうち、未活用のデータの割合⁴

4 Seagate Technology: 『Rethink Data: Put More of Your Data to Work—From Edge to Cloud』

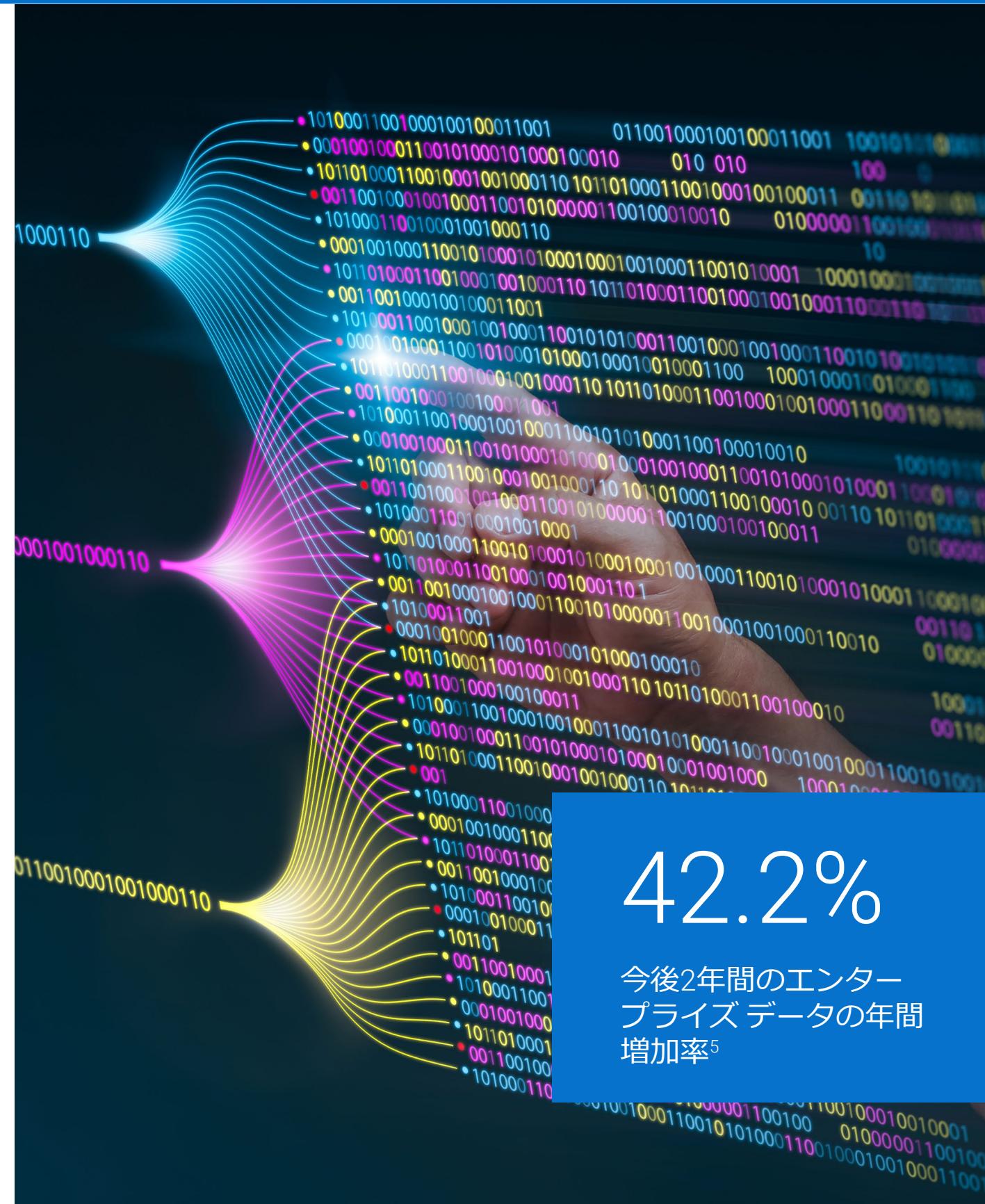
ステップ3

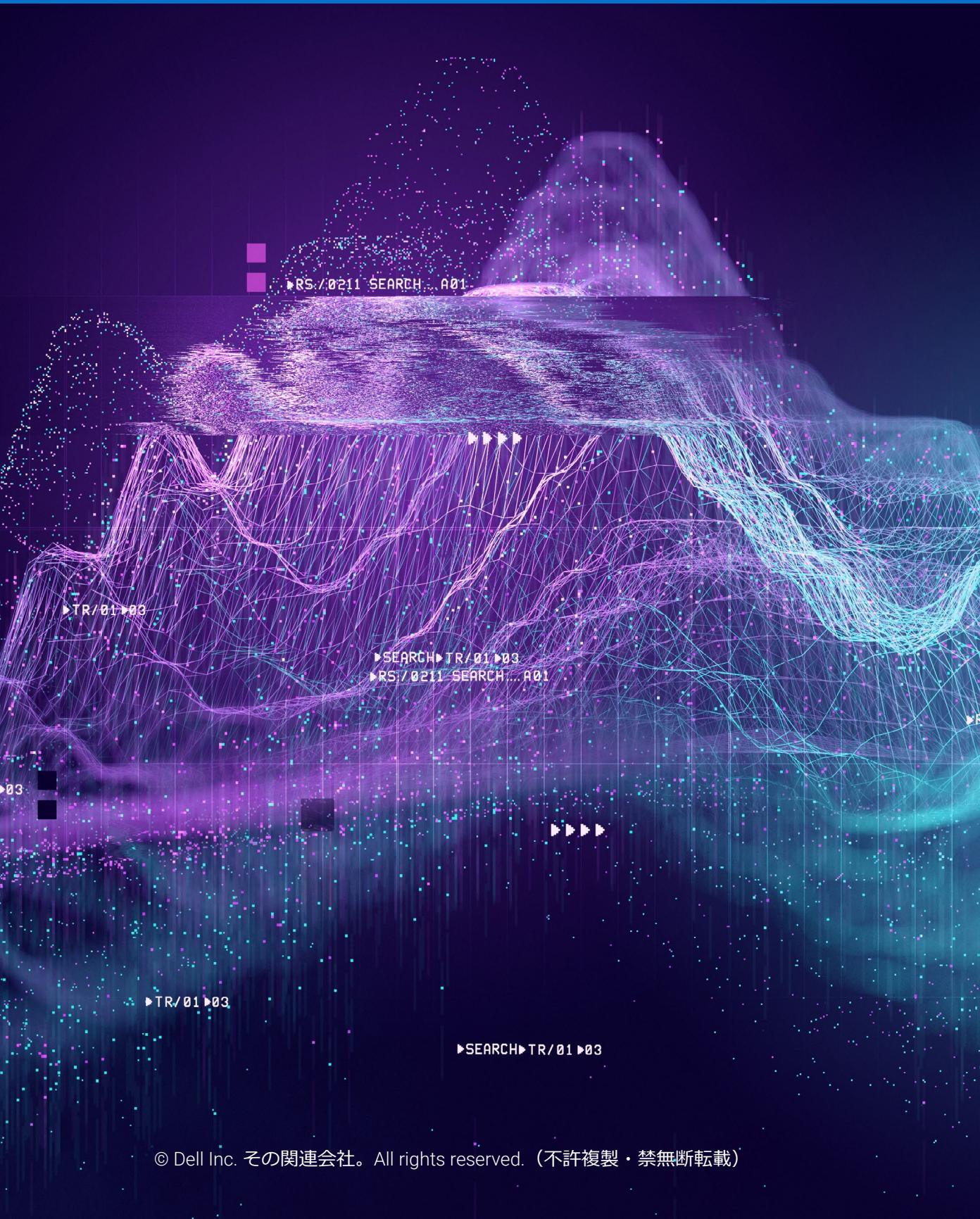
データの調査とアクセスのシンプル化

 分析のためにデータを一元化された場所に移動すると、リアルタイムのユース ケースの価値が低下します。この非効率を避けるために、データ サイエンティストがデータが存在する場所でアクセスできるようにします。プロセスを標準化し、自動検出を実装して、構造化データと非構造化データをアクセスしやすい場所に配置して、リアルタイムの検出と利用を促進します。このアプローチにより、時間を節約するだけでなく、データ運用の俊敏性も向上し、**インサイトの収集が迅速化して意思決定の応答性が向上します。**

データが爆発的に増加している中、AIのワークフローの規模を適切に調整するには、効果的なデータ管理が必要です。Dell Data Lakehouseなどのツールのご検討をおすすめします。このツールは、単一のフェデレーテッド クエリ エンジンを利用して、場所やデータ ソースを問わずにあらゆるエンタープライズ データの検出、クエリ、処理を実行する完全なソリューション スタックです。

5 Seagate Technology: 『Rethink Data: Put More of Your Data to Work—From Edge to Cloud』





ステップ4

分析、ML実験、モデリングの最適化



問題を解決できる変数を特定するために、継続的な実験とモデリングを奨励します。データ品質やプライバシーの課題に直面したときに特に役立つ合成データの作成を検討し、AIの導入を開始する際のプロセスの迅速化に役立ちます。増強と微調整のみを必要とする事前訓練済みの基礎モデルを活用することは、優れた出発点となります。

プロジェクトを拡大したら、データへの容易なアクセスに重点を置きます。これにより、チームは反復的なテストを通じて分析を最適化できます。複数の反復とアルゴリズムをサポートするプラットフォームは、重要なデータ変数を明らかにし、生成AIの有効性を高めて、データの価値を引き出すことができます。このようなテスト、学習、精緻化の反復的なプロセスにより、**モデルの堅牢性とインサイトの実用性**が確保され、継続的な改善とイノベーションが推進されます。

ステップ5

データと分析の製品化の拡大



データサイエンス プロジェクトから、信頼性と再現性に優れたデータサイエンス製品に移行します。これらの製品は独立して動作し、定期的に改善のためのレビューが行われます。効率的なデータ取得プロセスが、インサイトを得るまでの時間を短縮し、インテリジェンスが組み込まれたデータパイプラインがリアルタイムにデータを分析および使用し、事前定義された目標を達成できるようになります。データ製品を拡大することで、一貫性と信頼性を確保し、アドホックなプロジェクトを持続可能で価値を生み出す資産に変えることができます。



ステップ6

データ管理とガバナンスの自動化



製品が再現可能になるにつれて、組織のデータ管理とガバナンスのプロセスを自動化します。これにより、システムが自己監視して、問題になる前に異常を検出することができます。統合されたデータ ガバナンス基準、ポリシー、セキュリティ対策に従ってワークフローを確実に進行させるために、包括的な可観測性を取り入れます。自動化によって、手作業が低減するだけでなく、精度とコンプライアンスも向上し、大規模にデータを管理するための堅牢なフレームワークが得られます。

ステップ7

ビジネス上の成果の評価



ビジネス上の成果からデータを評価し、それをプロセスにフィードバックすることで、継続的な学習と適応の文化を育てます。この継続的な評価は、実際の結果と予測された結果を比較し、成果を改善し、結果を最適化するための改良を可能にします。人間の介入を最小限に抑えて機能を活用することで、持続的なパフォーマンスと価値を確保します。成果を継続的に分析することで、フィードバックループを作成し、継続的な改善を促進し、ビジネス目標の進化に合わせてデータ イニシアティブを調整します。



Dell AI Factory with NVIDIAでデータを活用

現在のデータセンターとIT運用モデルは、生成AIのスピードと規模を活用するにはおそらく適していません。必要なのは、[AI](#)固有の要件に対応するために設計された新しいタイプのデータセンターです。

AIファクトリーが必要

産業革命を加速させた物理的な工場と同様に、AIファクトリーはAI革命の原動力となります。ただし、AIファクトリーは、物理的な製品の代わりにインテリジェンスを生み出します。これらは、汎用性と柔軟性に優れ、クラウド、データセンター、ワークステーション、エッジ ロケーションなど、データが存在するあらゆる場所で運用できます。

2.4倍

大規模なMLモデルのトレーニングへの年間支出増加率（主にGPUやサーバーなどのハードウェア）⁶

⁶ Epoch AI: [『How Much Does it Cost to Train Frontier AI Models?』](#)





ユース ケースを理解し、適切なデータ管理ソリューションを適用することは、AIの取り組みを最適化するための基盤です。

Dell AI Factory with NVIDIAは、お客様のAIイノベーションを加速するために必要なビルディング ブロックをまとめて提供します。

- データ重視
- AIで最適化されたインフラストラクチャ
- AIソフトウェアとモデル
- オープンなエコシステム
- エキスパート サービス
- ベスト プラクティス

お客様のビジネス要件に基づいたテクノロジーにより、適切なソリューションを迅速に提供し、総所有コストを削減します。DellとNVIDIAは、これらのソリューションを検証し、リファレンス設計とベンチマー킹を提供するために、34万時間以上を費やしてきました。

NVIDIAとDellは共同で、AIの導入を加速するフルスタックソリューションである**Dell AI Factory with NVIDIA**を提供しています。

NVIDIA AI Enterpriseを使用すると、次のことが可能になります。



アクセス、管理、導入、サポート

さまざまなインフラストラクチャソリューションにわたる生成AIモデル、アプリケーション、ソリューション、アクセラレーター



NVIDIA NIMSの活用

すべてのデータパイプライン作業を最大限に活用するために、堅牢なデータパイプラインを構築



移行が簡単

データパイプラインの不可侵性を維持しつつ、1つのモデルから次のモデルに移行し、投資を保護して、利用可能な最適なモデルを常に採用

継続的なデータ管理で 成功を維持

データ管理工程は、1回限りのタスクではありません。継続的なプロセスであり、一連の手法とツールを活用してデータから継続的に価値を引き出すことができるよう支援します。そして、生成AIイニシアティブを成功させるうえで必要不可欠です。Dellは、NVIDIAと協力して、世界で最も幅広いAIソリューション ポートフォリオを提供しています。



エンドツーエンドソリューション

NVIDIA AIを搭載したエンドツーエンドソリューションが、デスクトップからデータセンター、クラウドまで、セキュリティ、アクセス性、拡張性を提供します。



AIへの取り組み

Dell AI Factory with NVIDIAは、AIの採用と実装に対するDellの取り組みを体現しており、お客様がAIイニシアティブを加速してデータを最大限に活用し、目標を達成できるよう支援します。



データナビゲーションのシンプル化

DellとNVIDIAのソリューションを組み合わせることで、データ管理工程を簡単に進めることができます。AI主導のデータワークフロー領域では、その工程が最終目的地と同じくらい重要です。



データ管理は1回限りのタスクではありません。

データ管理工程は、継続的なプロセスであり、一連の手法とツールを活用してデータから継続的に価値を引き出すことができるよう支援します。そして、生成AIイニシアティブを成功させるうえで必要不可欠です。

「万能な」アプローチは存在しません。

お客様には、独自のデータ基盤があり、ハードウェアとソフトウェア、データの場所とタイプ、独自のユース ケースから成る独自のエコシステムがあります。このe-bookで概説するステップは、お客様の組織固有のニーズに対応するための包括的なガイドとして役立ちます。

お客様は1人ではありません。

[Dell Services](#)は、ライフサイクルのすべての段階で、お客様の組織に合わせて高度な専門知識を提供し、タイムトゥバリューを短縮します。成功をもたらす戦略の調整、データの検証、データ プラットフォームの迅速な実装、安全の確保、運用の最適化などを、Dellの信頼できる専門家がサポートします。

データ管理工程のどの段階においても、Dellがエンタープライズ データを効果的に活用してAIプロジェクトを推進し、タイムトゥバリューを短縮するお手伝いをします。

詳細については、[Dell.comをご覧ください](#)

用語集

生成AI

生成AIとは、新しいコンテンツやデータの作成に重点を置いた人工知能のサブセットを指します。このタイプのAIは、敵対的生成ネットワーク(GAN)やトランスポーマーなどのアルゴリズムやモデルを使用して、テキスト、画像、音楽、その他のタイプのメディアを生成します。生成AIは、トレーニング データのスタイルとコンテンツを模倣したオリジナルの出力を生成する他に類のない能力があり、クリエイティブなタスク、コンテンツ作成、複雑なシナリオのシミュレーションに非常に価値があります。

AI

人工知能(AI)は、通常は人間の知能を必要とするタスクを実行できるシステムを作成することに焦点を当てた、コンピュータサイエンスの広範な分野です。タスクには、学習、推論、問題解決、知覚、言語理解、意思決定などがあります。AIには、機械学習、自然言語処理、ロボティクスなど、さまざまな手法やアプローチが含まれます。AIの目的は、新しい状況に適応し、時間の経過とともにパフォーマンスを向上させ、幅広いアプリケーションにわたって人間の活動を支援または自動化できるインテリジェントなエージェントを作成することにあります。

AIのワークフロー

AIのワークフローとは、AIシステムによって処理される特定のタスクまたはプロセスを指します。これらのワークフローはアプリケーションによって大きく異なり、データ処理、モデルのトレーニング、推論、分析を含む場合があります。AIワークフローの特徴は、モデルのトレーニング、大規模なデータ セットの処理、リアルタイムの意思決定に大量の計算を必要とすることです。AIワークフローの管理は、これらの要求の厳しいタスクを効率的に処理するために、ハードウェア、ソフトウェア、アルゴリズムを最適化することを含み、AIシステムが効果的大規模に動作することを保証します。

参考文献と追加資料

[インフォグラフィック](#) : データ管理工程

[Dell Professional Services for Generative AI](#)

[Dellのデータ管理ソリューション](#)

[Dell AI Factory with NVIDIA](#)

[生成AIのセキュリティに関する5つの重要な考慮事項](#)