



生成 AI ワークロード評価

AWS 規範ガイドンス



AWS 規範ガイドンス: 生成 AI ワークロード評価

Copyright © 2026 Amazon Web Services, Inc. and/or its affiliates. All rights reserved.

Amazon の商標およびトレードドレスは Amazon 以外の製品およびサービスに使用することはできません。また、お客様に誤解を与える可能性がある形式で、または Amazon の信用を損なう形式で使用することもできません。Amazon が所有していないその他のすべての商標は Amazon との提携、関連、支援関係の有無にかかわらず、それら該当する所有者の資産です。

Table of Contents

序章	1
このガイドの目的	1
対象者と利点	2
スコープ	2
目標とするビジネス成果	4
評価に関する考慮事項と前提条件	7
明確なユースケースから始める	7
ビジネスの連携を確保する	8
ガバナンスと監視を実装する	8
データおよび技術的な前提条件に対処する	8
コンピューティングリソースの要件を検討する	8
プライバシーとセキュリティへの影響に対処する	9
ステークホルダーを早期に関与させる	9
反復と学習	9
生成 AI ワークロード評価アンケート	10
準備状況	10
ユースケース	12
アーキテクチャ	16
[Storage (ストレージ)]	16
規制とコンプライアンス	18
Integration	18
テスト	20
デプロイと自動化	22
データ戦略	24
評価インサイトを実用的な結果に変換する	27
次のステップ	29
よくある質問	30
主な目的は何ですか?	30
この評価は誰が使用すべきですか?	30
主なコンポーネントは何ですか?	30
これはアーキテクチャの定義にどのように役立ちますか?	30
どのような利点がありますか?	30
これを正常に実装するにはどうすればよいですか?	31
課題は何ですか?	31

規制とコンプライアンスの要件は何ですか？	31
ステークホルダーの役割は何ですか？	31
成功を測定するにはどうすればよいですか？	31
組織の規模によってアプローチはどのように異なりますか？	32
リソース	33
ドキュメント履歴	34
.....	xxxv

生成 AI ワークロード評価

Tabby Ward と Deepak Dixit, Amazon Web Services (AWS)

2024 年 11 月 ([ドキュメント履歴](#))

生成 AI ワークロード評価は、生成 AI ワークロードを作成または更新する組織の準備状況の評価し、改善するための戦略的方法です。この評価は重要です。生成 AI をビジネスオペレーションに組み込むと、モノの仕組みが大きく変化し、新しい効率と機能を提供できるからです。ただし、生成 AI を正常に採用するには、現在のシステムを徹底的に理解し、将来の明確な計画を立てることが重要です。

生成 AI ワークロードとは、テキスト、画像、コード、その他のデータ型などの新しいコンテンツを作成できる人工知能モデルの使用を伴う計算タスクを指します。これらのワークロードには、通常、大量のコンピューティング能力、GPU の特殊なハードウェア、トレーニングと推論のための大規模なデータセットが必要です。生成 AI ワークロードをオペレーションに統合するには、いくつかの課題があります。

- インフラストラクチャ要件: 生成 AI モデルに必要な重要な計算リソースと特殊なハードウェアのプロビジョニング。
- データ管理: 大規模なデータセットを処理しながら、データの品質、プライバシー、コンプライアンスを確保します。
- スキルギャップ: AI テクノロジーとモデルのデプロイに関する専門知識の欠如。
- 倫理的考慮事項: AI 生成コンテンツのバイアス、公平性、透明性に対処します。
- 統合の複雑さ: 生成 AI を既存のワークフローとレガシーシステムにシームレスに組み込む。
- コスト管理: 実装と運用の高いコストと潜在的な利点のバランスを取ります。

これらの課題を克服するには、慎重な計画、インフラストラクチャと人材への投資、実装への戦略的アプローチが必要です。

このガイドの目的

生成 AI は、多くの業界で急速に重要なコンポーネントとなっています。変革の機会を提供しますが、統合、コンプライアンス、スケーラビリティに関しても課題があります。多くの組織は、技術基盤の弱さ、変化への抵抗、データ品質の問題により、AI のフル活用に苦労しています。生成 AI ワークロード評価は、モダナイゼーションの要件を特定し、実装の範囲を定義し、レガシーシステムと

思考に挑戦することで、これらの課題に対処します。また、最小限の実行可能な製品 (MVPs) を決定し、ターゲットソリューションアーキテクチャを開発して、AI 導入への構造化された戦略的アプローチを確保するのに役立ちます。

このガイドは、組織が生成 AI テクノロジーを採用する複雑さを克服するのに役立つ構造化されたアプローチとして機能します。このガイドは、最初から要件を明確に定義するのではなく、以下を支援します。

- 組織内の生成 AI の潜在的なユースケースを特定します。
- 生成 AI の導入に向けた組織の準備状況を評価します。
- ユースケース目標とストレッチ目標の定義と改善。
- 生成 AI 実装の範囲と要件を決定します。
- ターゲットソリューションアーキテクチャの開発。

対象者と利点

この評価は、生成 AI ワークロードのモダナイゼーションの技術的側面を評価したいソリューションアーキテクト、エンタープライズアーキテクト、アプリケーションアーキテクト向けに特別に設計されています。また、チームの全体的な準備状況、リソースの割り当て、有効化要件を計測したいプログラマージャーや人事マネージャーにとっても役立ちます。業界のベストプラクティスは、AI 導入の準備を整えるための包括的な評価の重要性を強調しています。これには、アーキテクチャ、ストレージ、コンプライアンス、統合、テスト、デプロイ、自動化の評価が含まれます。

スコープ

以下のトピックは、生成 AI ワークロード評価方法の対象です。

- 現在の生成 AI テクノロジーとモデル (大規模言語モデル、イメージ生成モデルなど)
- 生成技術を使用する狭い AI アプリケーション
- 生成 AI と既存のシステムやワークフローの統合
- 生成 AI モデルのトレーニングと微調整のためのデータ戦略
- 現在の生成 AI アプリケーションの倫理的考慮事項と責任ある AI プラクティス
- 本番環境における生成 AI のテストとデプロイ戦略
- 生成 AI 実装のセキュリティとプライバシーに関する考慮事項
- 生成 AI ワークロードのパフォーマンスの最適化とスケーラビリティ

- さまざまな業界における生成 AI のユースケースとアプリケーション
- 生成 AI 出力と品質保証プロセスの評価

以下のトピックは対象外です。

- 人工知能 (AGI) と人工知能 (ASI) のシナリオ
- 現在の生成モデルを超える AI の投機的将来の進歩
- AI の量子コンピューティングアプリケーション
- ニューロモーフィックコンピューティングとブレインコンピュータインターフェイス
- AI システムの意識と自己認識
- 現在の生成 AI アプリケーションを超える高度な AI の長期的な社会的影響
- 架空の将来の AI テクノロジーの規制フレームワーク
- 機械における知性と意識の本質に関する哲学的議論
- AI の極端なエッジケースまたは高度に投機的なユースケース
- 独自の AI モデルまたはアーキテクチャの詳細な技術仕様

目標とするビジネス成果

生成 AI ワークロード評価は、生成 AI ワークロードのモダナイズを成功させるために不可欠な、いくつかのターゲットを絞った結果を提供することを目指しています。これらの成果により、組織は AI テクノロジーを効果的かつ効率的に統合する準備が整います。

各ターゲット結果について、生成 AI ワークロード評価は以下に焦点を当てます。

- **相互依存関係:** 結果とモダナイゼーションプロセスの他の側面との間の相互依存関係を特定して明確にします。これには、モダナイゼーションへの包括的なアプローチを確保するために、ある成果が他の成果にどのように影響または影響を受けるかを理解することが含まれます。
- **ステークホルダーの連携:** さまざまなステークホルダーを各成果に合わせるための戦略を概説します。これには、賛同とサポートを促進するために、各成果の価値と影響をさまざまな組織レベルや部門に伝えることが含まれます。
- **優先順位付け:** 複数のユースケースや結果が特定された場合は、ビジネスへの影響、リソース要件、戦略的調整などの要因に基づいて優先順位を付けるためのフレームワークを提供します。
- **継続的な改善:** 成果ごとに、継続的な評価と改善のためのメカニズムを確立します。これにより、モダナイゼーションの取り組みは、変化する技術環境やビジネスニーズに適応し、対応し続けることができます。

以下は、目標とする各成果の詳細な説明です。

ターゲット アーキテクチャ

- **定義:** 評価は、生成 AI ワークロードの明確でスケーラブルなターゲットアーキテクチャを定義するのに役立ちます。
- **コンポーネント:** これには、適切なクラウドサービスの選択、データパイプラインの設計、システムの相互運用性の確保が含まれます。
- **メリット:** 明確に定義されたアーキテクチャは、スケーラビリティ、信頼性、パフォーマンスの最適化をサポートし、モダナイゼーションの強力な基盤を提供します。

お客様の準備状況

- **評価:** 組織のインフラストラクチャ、プロセス、文化の現在の状態を評価し、生成 AI モダナイゼーションの導入準備状況を決定します。

- 基準: これには、技術的能力、データ品質、変化を受け入れる組織の意欲の評価が含まれます。
- 成果: ギャップと改善が必要な領域を特定すると、組織が最新のソリューションとテクノロジーにスムーズに移行する準備が整います。

ユースケースの目標とストレッチの目標

- ユースケースの目標は、特定のビジネス上の問題や機会に焦点を当てて、ターゲットソリューション実装の明確な目標を確立します。

生成 AI モダナイゼーションにおけるユースケースの目標は、生成 AI ソリューションを実装することで組織が達成することを目指す、具体的で測定可能な目標を指します。これらの目標は通常、より広範なビジネス目標と整合しており、組織内の特定の課題や機会に対処することに重点を置いています。ユースケースの目標の例には、次のようなものがあります。

- 生成 AI を活用したチャットボットを使用することで、カスタマーサービスの応答時間を 50% 短縮します。
- 生成 AI 支援コード分析により、コードレビューの効率を 30% 向上させます。
- 生成 AI パターン認識を使用して不正検出の精度を 25% 向上させます。
- ストレッチ目標は、生成 AI モダナイゼーションが組織内で達成できるものの境界を押し広げる野心的なターゲットを定義します。
- 影響: 達成可能な目標と意欲的な目標の両方を設定することは、生成 AI モダナイゼーションイニシアチブを戦略的ビジネス目標と整合させ、イノベーションを促進するのに役立ちます。

労力の見積もり

- 目的: 正確な労力の見積もりは、リソース計画を支援し、プロジェクトが時間どおりに予算内で配信されるようにします。
- 範囲: 生成 AI モダナイゼーション計画の実装に必要なリソース、時間、予算を見積もります。
- 要因: 技術的な複雑さ、統合の課題、潜在的なリスクを考慮します。

有効化の二ーズ

- トレーニングと開発: 生成 AI モダナイゼーションの導入を成功させるために必要なスキルと知識を特定します。
- リソース: トレーニングプログラム、ワークショップ、その他の有効化アクティビティの必要性を判断します。

- 成果: スタッフが必要なスキルを備えていることを確認することで、生成 AI モダナイゼーションイニシアチブの有効性が向上し、長期的な成功をサポートします。

実装計画

- ロードマップ: 生成 AI モダナイゼーションを達成するために必要なステップを概説する詳細な計画を立てます。
- マイルストーン: 進行状況を追跡するための主要なマイルストーンと成果物を定義します。
- メリット: 明確な実装計画は方向性と説明責任を提供し、生成 AI モダナイゼーションへの構造化されたアプローチを促進します。

評価に関する考慮事項と前提条件

明確なユースケースから始める

生成 AI が対処できる特定のビジネス上の問題や機会を特定します。戦略的ビジネス目標に沿ったユースケースに焦点を当て、測定可能な利点を提供します。組織内で一般的に直面する課題をターゲットとするユースケースに優先順位を付け、ソリューションアーキテクチャが複数のシナリオのパターンとして機能するようにします。

潜在的な生成 AI アプリケーションを一般的に理解して評価プロセスを開始することは有益ですが、必須ではありません。このガイドに含まれている[アンケート](#)は、明確に定義されたユースケースを持つ組織から幅広いアイデアのみを持つ組織まで、さまざまなレベルの準備に対応しています。評価プロセスは以下に役立ちます。

- これらの最初のユースケースのアイデアを絞り込み、明確にします。
- 新しい潜在的なユースケースを特定します。
- ユースケースごとに具体的で測定可能な目標を作成します。
- 各ユースケースの実現可能性と潜在的な影響を評価します。

架空の例を考えてみましょう。ある金融サービス企業が生成 AI のモダナイゼーションを検討することを決定しました。まず、カスタマーサービスと不正検出プロセスを改善するという幅広いアイデアから始めます。

- 初期評価: このアンケートは、生成 AI 導入のための現在のシステム、データ品質、組織の準備状況を評価するのに役立ちます。
- ユースケースの絞り込み: 評価プロセスを通じて、最初のアイデアを 2 つの特定のユースケースに絞り込みます。
 - 顧客の問い合わせのための生成 AI を活用したチャットボットの実装
 - 生成 AI を使用したリアルタイムのトランザクション不正検出
- 目標設定: ユースケースごとに、特定の目標を定義します。
 - カスタマーサービスの応答時間を 6 か月以内に 40% 短縮
 - 不正検出の精度を 20% 向上させ、誤検出を 15% 削減
- ストレッチ目標: これらの野心的なターゲットも設定します。
 - AI 支援対応で 80% の顧客満足度を達成する

- 新しい不正パターンを識別する予測不正検出モデルを開発する
- MVP 定義: このアンケートは、即時価値を提供する重要な機能に焦点を当て、ユースケースごとに MVP を決定するのに役立ちます。
- ターゲットアーキテクチャ: 最後に、1 つまたは両方のユースケースをサポートするターゲットアーキテクチャを開発し、スケーラビリティと既存のシステムとの統合を確保します。

ビジネスの連携を確保する

生成 AI イニシアチブを全体的なビジネス戦略と目標に合わせます。ユースケースごとに、生成 AI がビジネスの成長、効率性、イノベーションにどのように貢献するかを示す明確な価値提案を作成します。生成 AI 実装が主要業績評価指標 (KPIs)。

ガバナンスと監視を実装する

部門横断的な運営委員会を作成し、生成 AI イニシアチブを監督します。責任ある AI の使用に関するポリシーとガイドラインを策定し、倫理的考慮事項と潜在的なバイアスに対処します。生成 AI プロジェクトのレビュープロセスを確立して、組織の標準と規制要件への準拠を確保します。

データおよび技術的な前提条件に対処する

データ品質を評価して改善し、データガバナンスプラクティスを実装して、生成 AI モデルへの信頼性の高い入力を確保します。生成 AI のニーズに固有のデータ収集、ストレージ、管理に対応するデータ戦略を策定します。データインフラストラクチャを評価して強化し、生成 AI ワークロードに必要なデータの量と速度をサポートします。

コンピューティングリソースの要件を検討する

現在の IT インフラストラクチャを評価し、生成 AI ワークロードの計算能力のギャップを特定します。クラウドサービスやオンプレミスの高性能コンピューティングクラスターなどのオプションを考慮して、スケーラブルなコンピューティングリソースを計画します。リソース割り当てを最適化して、トレーニングワークロードと推論ワークロードの両方のパフォーマンスとコスト効率のバランスを取ります。

プライバシーとセキュリティへの影響に対処する

生成 AI のトレーニングと運用で使用する機密データを保護するための堅牢なセキュリティ対策を実装します。個人情報进行处理するときは、一般データ保護規則 (GDPR) やカリフォルニア消費者プライバシー法 (CCPA) などのデータ保護規制を確実に順守してください。生成 AI 機能の不正アクセスや不正使用を防ぐため、安全なモデルデプロイとモニタリングのプロトコルを開発します。

ステークホルダーを早期に関与させる

リーダーシップの賛同とサポートを得るために、最初から主要なステークホルダーに関与させます。モダナイゼーションイニシアチブの利点と潜在的な影響、特に生成 AI ワークロードについて明確に伝える。ステークホルダーが生成 AI テクノロジーとその影響を理解するのに役立つトレーニングとリソースを提供します。

反復と学習

ターゲットソリューションを絞り込むことができる増分アプローチを採用します。フィードバックループを使用して、ワークロードのアーキテクチャとプロセスを継続的に改善します。生成 AI 実装のパフォーマンスと影響を定期的に評価し、実際の結果と進化するビジネスニーズに基づいて、必要に応じて戦略を調整します。

生成 AI ワークロード評価アンケート

以下のセクションでは、組織の生成 AI ワークロードのモダナイゼーションのさまざまな側面を評価するために使用できる質問を提供します。この包括的なアンケートでは、ユースケース、アーキテクチャ、ストレージ、コンプライアンス、統合、テスト、デプロイ、データ戦略など、主要な分野にわたる質問により、生成 AI ワークロードを採用および実装する組織の準備状況を評価します。このアンケートは、技術インフラストラクチャから規制上の考慮事項まで、生成 AI 実装の重要な側面に対処することで、AI モダナイゼーションジャーニーにおける強み、ギャップ、機会を特定するのに役立ちます。

セクション:

- [準備状況](#)
- [ユースケース](#)
- [アーキテクチャ](#)
- [\[Storage \(ストレージ\)\]](#)
- [規制とコンプライアンス](#)
- [Integration](#)
- [テスト](#)
- [デプロイと自動化](#)
- [データ戦略](#)

Microsoft Excel 形式でアンケートをダウンロードし、それを使用して情報を記録することもできます。

 [ダウンロードアンケート](#)

準備状況

質問	レスポンスの例
これらのワークロードに利用できる AWS アカウントはありますか？	はいまたはいいえ。

質問	レスポンスの例
との既存のエンタープライズ契約はありますか AWS ?	はいまたはいいえ。
生成 AI ワークロードを処理するための現在のクラウドインフラストラクチャはどの程度スケーラブルですか？	当社のクラウドインフラストラクチャはスケーラビリティが高く、大規模な生成 AI ワークロードを効率的に処理するように設計されたコンピューティングリソースと分散ストレージシステムの自動スケーリング機能を備えています。
大規模な前処理と特徴量エンジニアリングのためのデータパイプライン機能はありますか？	データパイプラインは、Apache Spark などの分散処理フレームワークを使用して大規模なデータ処理と特徴量エンジニアリングを行い、バッチデータ処理とストリーミングデータ処理の両方をサポートします。
アカウントのプロビジョニングおよび管理機能はありますか？	はいまたはいいえ。
生成 AI テクノロジーを導入する組織の AI リテラシーと準備状況をどのように説明しますか？	当社の組織は AI 教育プログラムに多額の投資を行い、ほとんどの技術スタッフが基本的な AI/ML トレーニングを完了しています。組織には、生成 AI などの新しいテクノロジーを受け入れるイノベーションの文化があります。
組織内にどのような AI/ML の専門知識が存在し、どのように分散されていますか？	経験豊富なデータサイエンティストと ML エンジニアによる専用の AI センターオブエクセレンスがあります。さまざまなビジネスユニットのドメインエキスパートをスキルアップして AI リテラルにし、生成 AI のユースケースを特定します。
クラウドプログラムの目的、利点、コストを明確に説明する高レベルのビジネスケースはありますか？	はいまたはいいえ。

質問	レスポンスの例
ソリューションを本番稼働に移行するためのタイムラインを教えてください。	週、月など。
主要なステークホルダー (CFO、CIT/CTO、COO など) が資金コミットメントを行いましたか？	はいまたはいいえ。
生成 AI イニシアチブでデータ保護規制を確実に遵守するにはどうすればよいですか？	AI チームと密接に連携する専任のコンプライアンスチームがあります。当社は、プライバシーへの影響評価を定期的の実施し、設計原則によるデータ保護を実装し、すべての生成 AI プロジェクトの詳細なデータ処理記録を維持します。
新しい生成 AI テクノロジーと統合する既存のシステムはどの程度成熟していますか？	当社の IT アーキテクチャは、新しい生成 AI テクノロジーの柔軟な統合を可能にするマイクロサービスと APIs に基づいています。これらのシステムは、相互運用性を確保するために、一般的なデータ形式とプロトコルで標準化されています。
ML モデルの運用にはどのような経験がありますか。また、これは生成 AI システムにどのように当てはまりますか。	自動モデルデプロイパイプライン、モニタリングシステム、A/B テストフレームワークなど、MLOps プラクティスを確立しました。これらのプラクティスは、大規模な生成 AI モデルの固有の要件を処理するように適応されています。

ユースケース

質問	レスポンスの例
ユースケースの主な目標または成功基準は何ですか？	カスタマーサポートの応答時間を改善するには、セールスコンバージョンを増やし、製品のレコメンデーションを強化します。また、ユー

質問	レスポンスの例
	ザー満足度、タスク完了率、応答品質などを向上させるため。
このユースケースは、組織の戦略的目標とどのように一致していますか？	これは、カスタマーサービスの応答時間を短縮することで顧客満足度を向上させるという戦略的目標と一致しています。
ユースケースで予想されるデータまたはリクエストの量を教えてください。	500 トランザクション/秒 (TPS)。
生成 AI ワークロードをサポートするには、どのようなタイプのデータソースが必要ですか？	内部構造化データベース (顧客レコード、販売データなど)、ドキュメント、E メール、ソーシャルメディアからの非構造化テキストデータ、音声および画像認識タスク用のオーディオおよびビデオファイル、IoT デバイスおよびセンサーからのリアルタイムストリーミングデータ、エンリッチメント用のパブリックデータセットおよび APIs。
これらのソースのデータを更新または更新する必要がある頻度はどのくらいですか？	トランザクションデータベース: ほぼリアルタイムの更新、ドキュメントリポジトリ: 毎日のバッチ更新、ソーシャルメディアフィード: 時間ごとの更新、IoT センサーデータ: 継続的なリアルタイムストリーミング、パブリックデータセット: 毎月または四半期ごとの更新。
生成 AI モデルでは、入力としてどのようなデータ形式が必要ですか？	構造化データ: CSV、JSON、SQL データベーステーブル、テキストデータ: プレーンテキスト、PDF、HTML、画像データ: JPEG、PNG、TIFF、オーディオデータ: WAV と MP3、ビデオデータ: MP4 と AVI。

質問	レスポンスの例
生成 AI ワークロードに関するデータ品質に関する主な懸念事項は何ですか？	完全性: 重要なフィールドが欠落していないことを確認する、精度: データの正確性を検証してエラーを排除する、一貫性: ソース間で統一された形式と値を維持する、適時性: リアルタイム推論のためにデータが最新であることを確認する、関連性: データが特定の生成 AI タスクと一致することを確認する。
主要なパフォーマンス要件 (応答時間、スループット、精度など) は何ですか？	95% の精度、< 500 ミリ秒の応答時間、1000 リクエスト/秒を処理する機能。高精度 (95% 以上)、中精度 (80 ~ 90%)、ベストエフォートなど。
このユースケースの成功を測定するための他の KPIs はありますか？	主要な KPIs には、エラー率の低下、トランザクションあたりの時間の節約、顧客満足度スコアが含まれます。
どの程度のモデル精度が必要で、コストとどのようにバランスが取れていますか？	中程度のコストで高精度 (>90%)、低コストで中程度の精度 (70 ~ 80%) など。
生成 AI ソリューションの主なユースケースまたはシナリオは何ですか？	カスタマーサービスチャットボット、コンテンツ生成、製品のレコメンデーションなど。
生成 AI システムのターゲットユーザーまたはペルソナは何ですか？	カスタマーサービスエージェント、マーケティングチーム、従業員、エンドユーザーなど。
リクエストまたはユーザーの予想されるボリュームはどれくらいですか？	1 日あたり 1,000 リクエスト、月間アクティブユーザー数 10,000。
特定のユースケースの制約や要件はありますか？	リアルタイムレスポンス、多言語サポート、データプライバシーなど。
生成 AI ソリューションを開発および維持するための予算が割り当てられていますか？	初期開発コストは 200,000 USD と推定され、年間メンテナンスコストは 50,000 USD です。
このユースケースで予測される投資収益率 (ROI) とペイバック期間を教えてください。	3 年間の予想 ROI は 150% で、ペイバック期間は 18 か月です。

質問	レスポンスの例
考慮すべき隠れたコストや潜在的な削減額がありますか？	潜在的な削減には、残業コストの削減が含まれます。隠れたコストには、スタッフ向けの追加のトレーニングが含まれる場合があります。
この生成 AI ソリューションのスケーラビリティと将来の拡張の可能性は何ですか？	このソリューションは、当社のオペレーションに合わせて拡張するように設計されており、将来的には他の部門にも拡張される可能性があります。
生成 AI モデルの公平性を確保し、バイアスを軽減するにはどうすればよいですか？	多様なデータ収集、定期的なバイアス監査、バイアス緩和手法の実装を通じて、バイアスを軽減する予定です。
倫理的懸念や意図しない結果に対処するために、どのようなプロセスを実施していますか？	当社では、確立された AI インシデント対応計画、定期的な倫理リスク評価、従業員向けの匿名報告システム、外部の専門家とのコラボレーション、フィードバックに基づくデプロイされたモデルの継続的なモニタリングと調整を通じて、倫理的懸念を管理します。
組織内のさまざまなプロジェクトや部門で、生成 AI ワークロード評価の優先順位付けと順序付けをどのようにアプローチしますか？	すべての部門で大まかな調査を実施して潜在的な生成 AI ユースケースを特定し、ビジネスへの影響、技術的実現可能性、倫理的考慮事項の 3 つの主要な基準に基づいて評価します。潜在的な影響が大きいプロジェクト、技術的な障壁が低いプロジェクト、および倫理的懸念が最小限に抑えられているプロジェクトが優先されます。

アーキテクチャ

質問	レスポンスの例
どのような生成 AI モデルまたはアーキテクチャが検討されていますか？	トランスフォーマー、畳み込みニューラルネットワーク (CNN)、再帰型ニューラルネットワーク (RNN)、決定木など。
データおよび計算の予想される規模または量を教えてください。	数百万のユーザー、ペタバイトのデータなど。
トレーニングと推論のハードウェア要件 (CPUs や GPUs など) は何ですか？	ハイエンド GPUs、CPU クラスタ、クラウドインスタンスなど。
生成 AI モデルは、時間の経過とともにどのように更新または再トレーニングされますか？	継続的な学習、定期的な再トレーニング、手動更新などを通じて。
データ前処理と特徴量エンジニアリングの要件は何ですか？	テキストのクリーニング、イメージの拡張、機能の選択など。
生成 AI システムは、エッジケース、外れ値、または信頼性の低い入力をどのように処理しますか？	人的監視へのフォールバック、明確化のリクエストなどを通じて。
生成 AI アプリケーションのレイテンシー要件は何ですか？	リアルタイム、ほぼリアルタイム、バッチ処理など。

[Storage (ストレージ)]

質問	レスポンスの例
トレーニングデータはどこに保存されますか？	クラウドストレージ (Amazon S3、ファイルストレージ、ブロックストレージ、オブジェクトストレージなど)、オンプレミスストレージなど。

質問	レスポンスの例
トレーニングデータとモデルアーティファクト (容量、耐久性、可用性など) のストレージ要件は何ですか？	ペタバイト規模のストレージ、高耐久性 (99.999999999% の耐久性)、高可用性など。
トレーニングデータとモデルアーティファクトのデータ保持とバックアップの要件は何ですか？	x 年間のデータ保持、日次バックアップ、オフサイトバックアップなど。
AI トレーニングデータセット (CSV、JSON、Parquet、HDF5 など) の保存に主に使用されるファイル形式はどれですか？	構造化データ用の Parquet ファイル、大きな多次元配列用の HDF5、画像やテキストなどの非構造化データ。TFRecord などの特殊な形式を使用して、トレーニング中のデータのロードを最適化します。
トレーニングデータセットは、個々のファイル、データベース、または特殊な AI データ形式を使用してどのように編成されていますか？	小～中規模のデータセットは、柔軟性のためにオブジェクトストレージに個別の Parquet ファイルとして保存されます。大規模なデータセットは、スケールを処理するために分散データベース (Cassandra) に保存されます。
生成 AI トレーニングデータ専用のデータ圧縮またはエンコーディング手法を使用していますか？	表形式データには、Parquet で利用可能なディクショナリエンコーディングとビットパッキング手法を使用します。イメージの場合、モデルに最適化された品質設定で、損失のある JPEG 圧縮を使用します。
トレーニングデータセットのさまざまな反復のバージョンングとストレージをどのように処理しますか？ これはストレージの全体的なニーズにどのような影響を与えますか？	ML プラットフォームと統合されたデータバージョンングシステム (DVC) を使用します。

規制とコンプライアンス

質問	レスポンスの例
生成 AI ソリューション (GDPR、HIPAA、PCI-DSS など) に関連する規制やコンプライアンス要件は何ですか？	個人データを処理するための GDPR、医療データの HIPAA、支払いデータの PCI-DSS など。
組織では、どのような倫理的生成 AI ガイドラインやフレームワークを採用していますか？	独自の責任ある AI ガイドラインを実装しました。すべての生成 AI プロジェクトは、承認とデプロイの前に倫理審査を受けます。
生成 AI システムのセキュリティ要件は何ですか？	データ暗号化、安全なネットワーク通信、定期的なセキュリティ監査。
データプライバシーと保護の要件は何ですか？	データの匿名化、暗号化、アクセスコントロールなど。
ソリューションが機密データまたは機密データを処理するための要件は何ですか？	厳格なアクセスコントロール、データマスキング、データレジデンシー要件など。
ユーザー認証と認可はどのように処理されますか？	OAuth、API キー、シングルサインオン (SSO)、ロールベースのアクセスコントロール (RBAC) を使用します。
ソリューションは本番環境でどのようにモニタリングおよび管理されますか？	Prometheus や Datadog などのモニタリングツールを使用して、ELK スタック、アラートシステムなどのログ記録ツールを使用します。

Integration

質問	レスポンスの例
生成 AI ソリューションを既存のシステムまたはデータソースと統合するための要件は何ですか？	REST APIs、メッセージキュー、データベースコネクタなど。

質問	レスポンスの例
生成 AI ソリューションのデータはどのように取り込まれ、前処理されますか？	バッチ処理、ストリーミングデータ、データ変換、特徴量エンジニアリングを使用します。
生成 AI ソリューションの出力はどのように消費されるか、ダウンストリームシステムと統合されますか？	API エンドポイント、メッセージキュー、データベースの更新などを通じて。
生成 AI ソリューションに使用できるイベント駆動型統合パターンはどれですか？	メッセージキュー (Amazon SQS、Apache Kafka、RabbitMQ など)、pub/sub システム、ウェブフック、イベントストリーミングプラットフォーム。
生成 AI ソリューションを他のシステムに接続するために使用できる API ベースの統合アプローチはどれですか？	RESTful APIs、GraphQL APIs、SOAP APIs (レガシーシステム用)。
生成 AI ソリューション統合に使用できるマイクロサービスアーキテクチャコンポーネントはどれですか？	サービス間通信、API ゲートウェイ、コンテナオーケストレーション (Kubernetes など) のサービスメッシュ。
生成 AI ソリューションにハイブリッド統合を実装するにはどうすればよいですか？	リアルタイム更新用のイベント駆動型パターン、履歴データのバッチ処理、外部システム統合用の APIs を組み合わせることができます。
生成 AI ソリューションの出力をダウンストリームシステムと統合するにはどうすればよいですか？	API エンドポイント、メッセージキュー、データベースの更新、ウェブフック、ファイルエクスポートを通じて。
生成 AI ソリューションを統合するには、どのセキュリティ対策を検討する必要がありますか？	認証メカニズム (OAuth や JWT など)、暗号化 (転送中および保管中)、API レート制限、アクセスコントロールリスト (ACLs)。
LlamaIndex や LangChain などのオープンソースフレームワークを既存のデータパイプラインや生成 AI ワークフローに統合するにはどうすればよいですか？	LangChain を使用して、複雑な生成 AI アプリケーション、特にエージェントとメモリの管理機能を構築する予定です。今後 6 か月以内に LangChain を使用する生成 AI プロジェクトの 60% を目標としています。

質問	レスポンスの例
選択したオープンソースフレームワークと既存のデータインフラストラクチャとの互換性を確保するにはどうすればよいですか？	スムーズな互換性を確保するために、専用の統合チームを作成しています。第 3 四半期までに、現在のデータレイク構造内で効率的なデータインデックス作成と取得のために LlamaIndex を使用する完全に統合されたパイプラインを持つことが目標です。
ラピッドプロトタイピングと実験のために、LangChain などのフレームワークのモジュラーコンポーネントをどのように活用する予定ですか？	開発者が LangChain のコンポーネントを使用して迅速にプロトタイプを作成できるサンドボックス環境をセットアップしています。
これらの急速に進化するオープンソースフレームワークの更新や新機能に対応するための戦略は何ですか？	LangChain と LlamaIndex の GitHub リポジトリとコミュニティフォーラムをモニタリングするチームを割り当てました。パフォーマンスの向上と新機能に焦点を当てて、主要な更新を四半期ごとに評価して統合する予定です。

テスト

質問	レスポンスの例
テスト要件 (ユニットテスト、統合テスト、end-to-endテストなど) は何ですか？	個々のコンポーネントのユニットテスト、外部システムとの統合テスト、重要なシナリオの end-to-endテストなど。
生成 AI トレーニングのために、さまざまなソース間でデータの品質と一貫性を確保するにはどうすればよいですか？	自動データプロファイリングツール、定期的なデータ監査、一元化されたデータカタログを通じてデータ品質を維持します。ソース間の一貫性を確保し、データリネージを維持するために、データガバナンスポリシーを実装しました。

質問	レスポンスの例
生成 AI モデルはどのように評価および検証されますか？	ホールドアウトデータセット、人間による評価、A/B テストなどを使用します。
生成 AI モデルのパフォーマンスと精度を評価するための基準は何ですか？	精度、再現率、F1 スコア、多重度、人間による評価など。
エッジケースとコーナーケースはどのように識別され、処理されますか？	包括的なテストスイート、人間による評価、攻撃者によるテストなどを使用します。
生成 AI モデルの潜在的なバイアスをどのようにテストしますか？	人口統計パリティ分析、等価機会テスト、敵対的偏見排除手法、反事実テストなどを使用します。
モデルの出力の公平性を測定するために使用されるメトリクスはどれですか？	異なる影響率、均等化されたオッズ、属性パリティ、個々の公平性メトリクスなど。
バイアス検出のためにテストデータセットで多様な表現を確保するにはどうすればよいですか？	属性グループ間の層別サンプリング、多様性の専門家とのコラボレーション、合成データを使用してギャップを埋めるなど。
デプロイ後のモデルの公平性を継続的にモニタリングするには、どのプロセスを実装しますか？	定期的な公平性監査、自動バイアス検出システム、ユーザーフィードバック分析、更新されたデータセットを使用した定期的な再トレーニングなど。
生成 AI モデルの交差バイアスをどのように対処しますか？	交差公平性分析、サブグループテスト、交差性に関するドメインエキスパートとのコラボレーションなどを使用します。
さまざまな言語や文化的コンテキストでモデルのパフォーマンスをどのようにテストしますか？	多言語テストセット、文化専門家とのコラボレーション、ローカライズされた公平性メトリクス、異文化比較研究などを活用します。

デプロイと自動化

質問	レスポンスの例
スケーリングとロードバランシングの要件は何ですか？	インテリジェントなリクエストルーティング、自動スケーリングシステム。モデルキャッシュ、遅延ロード、分散ストレージシステムなどの手法を採用することで、高速コールドスタートを最適化し、バースト的で予測不可能なトラフィックパターンを処理するシステムを設計します。
新しいバージョンを更新およびロールアウトするための要件は何ですか？	ブルー/グリーンデプロイ、Canary リリース、ローリング更新など。
ディザスタリカバリとビジネス継続性の要件は何ですか？	バックアップと復元の手順、フェイルオーバーメカニズム、高可用性設定など。
生成 AI モデルのトレーニング、デプロイ、管理を自動化するための要件は何ですか？	自動トレーニングパイプライン、継続的デプロイ、自動スケーリングなど。
新しいデータが使用可能になると、生成 AI モデルはどのように更新および再トレーニングされますか？	定期的な再トレーニング、増分学習、転移学習などを通じて。
モニタリングと管理を自動化するための要件は何ですか？	自動アラート、自動スケーリング、自己修復など。
生成 AI ワークロードに推奨されるデプロイ環境は何ですか？	モデルトレーニングに AWS を使用し、推論にオンプレミスインフラストラクチャを使用してデータレジデンシー要件を満たすハイブリッドアプローチ。
生成 AI のデプロイに希望する特定のクラウドプラットフォームはありますか？	AWS のサービス、特にモデルの開発とデプロイには Amazon SageMaker AI、基盤モデルには Amazon Bedrock。
生成 AI ワークロードにはどのようなコンテナ化テクノロジーを検討していますか？	Kubernetes とオーケストレーションされた Docker コンテナを標準化して、ハイブリッド

質問	レスポンスの例
	環境全体で移植性とスケーラビリティを確保したいと考えています。
生成 AI パイプラインに CI/CD 用の推奨ツールはありますか？	バージョン管理および CI/CD パイプライン用の GitLab。Jenkins と統合され、テストとデプロイを自動化します。
生成 AI ワークフローを管理するために、どのようなオーケストレーションツールを検討していますか？	ワークフローオーケストレーション、特にデータの前処理パイプラインとモデルトレーニングパイプライン用の Apache Airflow。
生成 AI ワークロードをサポートするために、オンプレミスインフラストラクチャに特定の要件はありますか？	オンプレミスの推論ワークロードをサポートするために、GPU アクセラレーションサーバーと高速ネットワーキングに投資しています。
異なる環境間でモデルのバージョンニングとデプロイを管理するにはどうすればよいですか？	モデルの追跡とバージョンニングに MLflow を使用し、Kubernetes インフラストラクチャと統合して環境間でシームレスにデプロイする予定です。
生成 AI デプロイでは、どのようなモニタリングおよびオプザバビリティツールを検討していますか？	メトリクス収集用の Prometheus、視覚化用の Grafana、モデル固有のモニタリング用の追加のカスタムログ記録ソリューション。
ハイブリッドデプロイモデルでデータの移動と同期にどのように対処していますか？	を使用して AWS、オンプレミスストレージと間の AWS DataSync 効率的なデータ転送を行います。自動同期ジョブは、トレーニングサイクルに基づいてスケジュールされます。
さまざまな環境にわたる生成 AI デプロイには、どのようなセキュリティ対策を実装していますか？	オンプレミスの Active Directory と統合されたクラウドリソースに IAM を使用して、end-to-endの暗号化とネットワークセグメンテーションを実装し、データフローを保護します。

データ戦略

質問	レスポンスの例
生成 AI ワークロードにとって重要な特定のデータ型と、現在アクセス可能なデータ型の割合	顧客通話ログと製品レビューデータは重要です。現在、これらのデータ型の 85% は生成 AI プロジェクトでアクセスできます。
データの品質をどのように確認および測定しますか？	完全性、正確性、一貫性、適時性などのデータ品質メトリクスを実装しました。自動ツールを使用してこれらのメトリクスを定期的に評価し、データのクレンジングとエンリッチメントに特化したチームを設けています。
データの何パーセントが生成 AI 使用の品質基準を満たしていますか？	現在、データの 78% が品質基準を満たしています。データクリーニングプロセスの改善により、今後 12 か月以内に 95% を目指しています。
生成 AI におけるデータ使用量に関する信頼ステークホルダーの間でどのように構築する予定ですか？	AI 掲示板を実装し、AI の決定を明確に説明し、透明性と公平性を確保するために四半期ごとの AI 監査を実施しています。
データソースと系統に関するドキュメントはどの程度包括的ですか？	当社は、オリジン、更新頻度、使用状況など、すべてのデータソースのメタデータを含む詳細なデータカタログを維持します。データリネージュツールを使用して、システム全体でのデータフローと変換を追跡します。
AI モデルのバイアスを防ぐために、データセットの多様性をどのように確保しますか？	当社は、さまざまな属性から積極的にデータを調達し、データセットに表現バイアスがないか定期的に監査しています。また、合成データ生成手法を使用して、過小評価されているカテゴリのバランスを取ります。
重要な生成 AI モデルのデータ更新レートはどれくらいですか。また、この頻度はどのように決定しますか。	重要なモデルは毎週更新されます。この頻度は A/B テストのパフォーマンスメトリクスによっ

質問	レスポンスの例
	て決定され、更新間のパフォーマンス低下は 2% 以下を目指します。
重要なデータセットのバージョンをいくつ、どのくらいの期間維持していますか？	各重要なデータセットの最後の 5 つのバージョンを維持し、各バージョンの保存期間は 18 か月です。
生成 AI イニシアチブに関与し、データにアクセスできる部門横断的なチームはいくつありますか？	3 つの部門横断的なチームがあります。各チームには、データサイエンティスト、ドメインエキスパート、倫理学者、ビジネスアナリストが含まれます。
どのようなデータガバナンスポリシーとプラクティスを実施していますか？	データポリシーを監督する部門横断的なデータガバナンス委員会があります。ガバナンスフレームワークへの準拠を確保するために、ロールベースのアクセスコントロール、データ分類スキーム、および定期的な監査を実装しました。
データプライバシーを確保し、適切な同意を取得し、機密性を維持するためにどのような対策を講じていますか？	GDPR と CCPA に沿った包括的なデータプライバシーフレームワークを実装しました。これには、データ使用に対する明示的な同意の取得、データ匿名化手法の実装、定期的なプライバシー影響評価が含まれます。
前四半期にバイアスについて監査された AI トレーニングデータセットの割合はどれくらいですか？	AI トレーニングデータセットの 70% が前四半期にバイアスについて監査されました。自動バイアス検出ツールを実装して、四半期ごとの監査を 100% 達成しています。
現在のデータ処理能力はどれくらいですか。また、将来の生成 AI ワークロードにはどの程度のプロジェクトが必要ですか。	現在の容量は 10 TB/日です。1 年以内に 30 TB/日が必要と予測され、この需要に合わせてインフラストラクチャをスケーリングしています。

質問	レスポンスの例
データプライバシーと生成 AI モデルのデータニーズのバランスを取る戦略は何ですか？	高度な匿名化技術と合成データ生成を実装しています。当社の目標は、AI の使用可能なデータを 40% 増やし、来年のプライバシーリスクを 60% 削減することです。
機械学習 (ML) データセットの何パーセントが正確にラベル付けされ、目標精度率はどれくらいですか？	現在、ML データセットの 85% が正確にラベル付けされています。ヒューマンラベリング手法と自動ラベリング手法の両方を採用することで、次の四半期に 95% の精度を目指しています。

評価インサイトを実用的な結果に変換する

このセクションでは、アンケートの回答を分析し、それらのインサイトを使用して生成 AI モダナイゼーションイニシアチブのターゲットアーキテクチャやその他の主要な成果物を形成するためのフレームワークを提供します。このフレームワークは、データ収集と実装のギャップを埋め、評価がモダナイゼーション戦略に直接通知し、推進できるようにします。

ターゲットアーキテクチャの定義：

- アンケートの回答を使用して、クラウドサービスの選択とデータパイプラインの設計を通知します。
- このガイドで説明されているように、アーキテクチャ設計がスケーラビリティと相互運用性をサポートしていることを確認します。

顧客準備状況の評価：

- 現在のインフラストラクチャ、プロセス、組織文化に関連するアンケートの回答を分析します。
- ギャップを特定し、それに対処するための計画を作成します。MVP の成功に不可欠なギャップに優先順位を付けます。

ユースケースとストレッチ目標：

- アンケートの回答から特定のビジネス上の問題を抽出して、明確なユースケースの目標を定義します。
- 生成 AI モダナイゼーションに関する組織の長期的なビジョンに沿ったストレッチ目標を設定します。

労力の見積もり：

- アンケートデータを使用して、MVP とフル実装の両方のリソース、時間、予算を見積もります。
- MVP で始まる段階的なアプローチを作成し、後続のフェーズの概要を説明します。

有効化のニーズ：

- アンケートの回答に基づいて、スキルギャップとトレーニングのニーズを特定します。
- 即時の MVP ニーズと長期的な生成 AI 導入の両方をサポートするトレーニング計画を立てます。

実装計画：

- MVP から始まり、完全な生成 AI モダナイゼーションに向けたステップを概説する包括的なロードマップを作成します。
- 実装の各フェーズの明確なマイルストーンと成果物を定義します。

実践的なステップ：

- 優先順位付けマトリックス: アンケートの回答を [6 つの結果](#) にマッピングして、特徴と労力の優先順位付けに役立つマトリックスを作成します。
- 反復アプローチ: 一連の計画されたリリースの最初の反復となるように MVP を設計します。各リリースは、完全なターゲットアーキテクチャに向けて構築されます。
- ステークホルダーの調整: アンケートの結果を使用して、ステークホルダーを MVP の範囲とすべての成果を達成するための段階的なアプローチに合わせます。
- 継続的なフィードバックループ: MVP デプロイ後にフィードバックを収集するメカニズムを実装し、インサイトを使用して後続のフェーズの計画を絞り込みます。
- アジャイル実装: MVP で最も重要な結果から始めて、時間の経過とともにすべての結果に柔軟に対応できるアジャイル手法を採用します。

次のステップ

生成 AI ワークロード評価を完了したら、次のステップに従います。

1. 詳細なターゲットアーキテクチャを提供する

- 目的: ソリューションアーキテクトは、組織の目標と評価の結果に沿った包括的なターゲットアーキテクチャを作成します。
- コンポーネント: このアーキテクチャには、スケーラビリティ、信頼性、パフォーマンスの最適化を確保するためのデータインジェスト、統合ポイント、システムの相互運用性の設計が含まれます。

2. ユースケースにどの程度 AWS のサービス 適合するかを説明する

- サービスマッピング: 特定されたユースケースに最も適 AWS のサービス した特定の を特定してマッピングします。
- メリット: これらのサービスが特定のビジネスニーズにどのように対応し、効率を高め、スケーラビリティを提供するかを強調します。

3. 長所と短所を備えたオプションの代替ソリューションを提供する

- 代替方法: 組織の要件を満たす代替ソリューションを提示します。
- 分析: コスト、複雑さ、ビジネス目標との整合性などの要因を考慮して、各代替案の利点と欠点の詳細な分析を提供します。

4. の詳細な価格見積もりを提供する AWS のサービス

- コスト分析: 潜在的な使用シナリオや料金モデルなど AWS のサービス、提案された の詳細なコスト見積もりを提供します。
- 予算の調整: コストが組織の予算上の制約と一致し、財務上の影響を明確に理解していることを確認します。

5. 提案されたアーキテクチャに関するフィードバックを取得する

- ステークホルダーのエンゲージメント: ステークホルダーと協力して、提案されたアーキテクチャを提示し、フィードバックを収集します。
- 反復的な改善: フィードバックを使用してソリューションを改良および改善し、すべての利害関係者のニーズと期待を満たしていることを確認します。

よくある質問

生成 AI ワークロード評価の主な目的は何ですか？

評価の主な目的は、生成 AI ワークロードをモダナイズする組織の準備状況を評価し、ユースケースを特定し、ターゲットソリューションアーキテクチャを開発することです。モダナイゼーションの要件を定義し、実装範囲を決定し、生成 AI モダナイゼーションを成功させるための準備をすることを目指しています。

この評価は誰が使用すべきですか？

この評価は、生成 AI モダナイゼーションの技術的側面を評価したいソリューションアーキテクト、エンタープライズアーキテクト、アプリケーションアーキテクトを対象としています。また、プログラママネージャーと人事マネージャーは、全体的な準備状況、リソースの割り当て、有効化のニーズを測定するのに役立ちます。

評価で評価される主要なコンポーネントは何ですか？

この評価は、全体的な準備状況、ユースケース、アーキテクチャ、ストレージ、規制とコンプライアンス、統合、テスト、デプロイ自動化、データ戦略を対象としています。これらのコンポーネントは、生成 AI モダナイゼーションの導入に向けた技術的および組織的な準備状況を決定するために不可欠です。

評価はターゲットアーキテクチャの定義にどのように役立ちますか？

この評価は、現在のシステムを評価し、改善点を特定するための構造化されたアプローチを提供します。これは、適切なテクノロジーを選択し、ビジネス目標とユースケースの要件に合ったスケーラブルなアーキテクチャを設計するのに役立ちます。

生成 AI ワークロード評価を実施する利点は何ですか？

メリットには、効率の向上、意思決定の改善、コンプライアンス保証、イノベーションの促進、スケーラビリティの準備などがあります。この評価では、生成 AI モダナイゼーションへの戦略的アプローチを確立し、リスクを軽減しながら潜在的な利点を最大化します。

組織は評価後に実装を確実に成功させるにはどうすればよいですか？

組織は、定義されたマイルストーンを含む明確な実装計画を策定し、ステークホルダーを早期に関与させ、反復的なアプローチを採用する必要があります。センターオブエクセレンス (CoE) を確立し、人材開発に焦点を当てることも、推奨されるベストプラクティスです。

評価中に組織が直面する可能性のある課題

課題には、変更への抵抗、データ品質の問題、コンプライアンスの複雑さなどがあります。これらの課題に対処するには、イノベーションの文化を育み、データの準備を整え、堅牢なセキュリティ対策を実装する必要があります。

評価では、規制とコンプライアンスの要件にどのように対処していますか？

評価では、現在のコンプライアンス対策を評価し、ギャップを特定します。これにより、ターゲットソリューションが関連する規制やデータプライバシー法に準拠し、機密情報を保護するためのセキュリティのベストプラクティスが組み込まれます。

ステークホルダーエンゲージメントは評価プロセスでどのような役割を果たしますか？

ステークホルダーの関与は、賛同を得て、モダナイゼーションイニシアチブをビジネス目標と整合させ、実装を成功させるために不可欠です。早期の関与とメリットの明確なコミュニケーションは、抵抗を克服し、サポートを促進する上で重要です。

組織は評価後に生成 AI モダナイゼーションイニシアチブの成功をどのように測定できますか？

成功は、ビジネス目標に沿った主要業績評価指標 (KPIs) を使用して測定できます。これらのメトリクスの定期的なモニタリングと評価は、意思決定をガイドし、生成 AI モダナイゼーションの価値を利害関係者に示すのに役立ちます。

さまざまな規模 (小規模、中規模、エンタープライズ) の組織や業界では、評価アプローチはどのように異なりますか？

小規模な組織:

- 包括的な評価のためのリソースと専門知識が限られている可能性がある
- 企業全体の導入ではなく、特定の影響の大きいユースケースに焦点を当てる可能性が高い
- 評価のためにサードパーティーのツールやサービスに大きく依存する可能性がある
- 評価プロセスは形式が低く、俊敏性が高い場合があります

中規模の組織:

- 多くの場合、専任の IT チームまたはデータチームがありますが、AI に関する専門知識が不足している可能性があります。
- 主要部門のパイロットプロジェクトから段階的なアプローチを取る可能性がある
- イノベーションと既存のシステムやプロセスのバランスを取る必要がある
- 評価には部門横断的なチームが含まれる可能性が高い

エンタープライズ組織:

- 通常、包括的な評価のために、専用の AI/ML チームやその他のリソースがあります。
- 既存のエンタープライズシステムとの複雑な統合を検討する必要がある
- 考慮すべき業界固有の規制要件がある可能性がある
- 多くの場合、評価には正式なガバナンスプロセスが含まれます。

リソース

- [の生成 AI AWS](#)
- [AWS は、AI 戦略を計画するための新しい人工知能、機械学習、生成 AI ガイド](#)を提供しています (AWS ブログ記事)
- [で生成 AI アプリケーションを構築するためのベストプラクティス AWS](#) (AWS ブログ記事)
- [での Generative AI Application Builder AWS](#) (AWS ソリューションライブラリ)
- [生成 AI 機能](#) (AWS セキュリティリファレンスアーキテクチャ)
- [AWS 生成 AI ベストプラクティスフレームワーク](#) (AWS Audit Manager ユーザーガイド)
- [生成 AI サービスの選択](#) (AWS 決定ガイド)
- [Amazon Bedrock とは](#) (Amazon Bedrock ユーザーガイド)
- [Amazon SageMaker AI とは](#) (Amazon SageMaker AI デベロッパーガイド)

ドキュメント履歴

以下の表は、本ガイドの重要な変更点について説明したものです。今後の更新に関する通知を受け取る場合は、[RSS フィード](#) をサブスクライブできます。

変更	説明	日付
初版発行	—	2024 年 11 月 6 日

翻訳は機械翻訳により提供されています。提供された翻訳内容と英語版の間で齟齬、不一致または矛盾がある場合、英語版が優先します。