

Red Hat AI Enterprise により統合 AI プラットフォームでの構築体制を実現

AI の急速な台頭

AI はすでに広く導入されており、次のような結果をもたらしています。

- ▶ 55% の組織が生成 AI を使用。¹
- ▶ 50% 以上の企業がエージェント型 AI の PoC または厳選したユースケースを実施。²
- ▶ AI 対応アプリケーションの 30% が 2026 年末までにエージェント型 AI を使用する見込み。²

AI の適用機会の進化によって複雑さが増大

AI により、企業は運営や競争の方法を根底から作り直すことができるようになりました。AI は急速に今日のビジネス環境における要となりつつありますが、AI の導入が進んでいるにもかかわらず、多くの企業はさまざまな課題のために AI の価値を最大限に引き出すことができていません。

そうした課題には、モデルコストの上昇、複雑なカスタマイズのニーズ、デプロイに関する厳格な制約、イノベーションの加速への対応などがあります。

企業が生成 AI のイノベーション、エージェント型 AI ワークフロー、ソブリン AI へと軸足を移し始める中で、そうした課題は複雑さを増し、対処が難しくなっています。これを解決するためには、推論コストの削減、スケーリングと監視の効率化、オンプレミスのソブリン AI の構築、変化への迅速な適応に役立つ AI プラットフォームへと戦略的に移行する必要があります。

エンタープライズ向けのプラットフォームとツールで AI の複雑さに対処

AI の導入を開始したばかりの段階から、既存の AI イニシアチブを組織全体に拡張しようとしている段階まで、Red Hat® AI はあらゆる組織が AI イノベーションを加速させ、ハイブリッド環境およびマルチクラウド環境全体での AI ソリューションの開発と提供にかかる運用コストを削減できるよう支援します。

Red Hat AI は、最適化されたモデルと効率的な推論によるコスト効率の良いソリューション、プライベートデータとの統合の最適化、スケーラブルで柔軟なプラットフォーム上でのエージェント型 AI の迅速な提供を実現します。また、単一サーバー環境から高度にスケールアウトされた分散プラットフォームまで、予測 AI モデルと生成 AI モデルのライフサイクルを大規模に管理・監視することができます。

実績あるオープンソース・テクノロジーをベースとして構築された Red Hat AI は主要なオープンソース AI モデルを広くサポートしており、企業がイノベーションを加速し、重要なツールおよびテクノロジーへのアクセスを民主化できるよう支援します。

さらに、AI パートナーのエコシステムがこれを補完します。このエコシステムは、パフォーマンス、安定性、アクセラレーターのサポートに重点を置いたテスト済みおよび認定済みの製品とサービスをインフラストラクチャ全体で提供しており、企業が AI によってビジネス上および技術上の主な課題を解決できるよう支援します。

1 IDC Tech Buyer Presentation, 『Worldwide Generative AI Industry Use Case Early Adoption Trends, 2025: Executive Summary』、Document #US53280825、2025 年 4 月 (購入が必要)

2 IDC 調査, 『Agentic AI Impact on Digital Infrastructure Strategies』、Document #US53418526、2025 年 10 月 (購入が必要)

Red Hat AI は以下を提供します。

- ▶ AI モデルライフサイクルの完全な監視と管理
- ▶ 使用するハードウェア・アクセラレーターと AI モデルを自由に選択できる柔軟性
- ▶ 最適化された AI モデルの検証済みコレクション
- ▶ ハイブリッド環境またはマルチクラウド環境全体での柔軟でコスト効率の高い推論
- ▶ 企業のプライベートデータ資産との信頼できる統合
- ▶ エージェント型 AI ワークフローの提供の最適化
- ▶ 必要に応じてリソースを効率的に使用してスケーリングする能力

Red Hat AI Enterprise は Red Hat AI ポートフォリオの一部として提供される統合 AI プラットフォームであり、今日の AI 導入における課題を解決できるよう支援します。

AI プラットフォームに Red Hat AI Enterprise を選ぶべき 6 つの理由

Red Hat AI Enterprise は、AI の取り組みのあらゆる段階にある企業をサポートするために必要なすべてのツールと機能を提供します。そのため、統合 AI プラットフォーム上のハイブリッドおよびマルチクラウド環境全体で、効率的でコスト効率の高い AI モデル、エージェント、アプリケーションの開発とデプロイを行うことができます。

このプラットフォームは、すぐに使用できる開発環境と主要な AI 機能を提供します。そうした機能には、モデルのチューニング、高性能な推論、エージェント型 AI ワークフローの管理などがあります。また、優れた柔軟性により、あらゆるモデルをサポートし、あらゆるハードウェアを使用し、どこにでもデプロイし、進化するデータロケーション要件に対応できます。

柔軟かつ効率的な推論で AI のデプロイを効率化

AI のイノベーションが加速し続ける中でデプロイの複雑さも増しており、その結果リソースの非効率的な使用、運用コストの上昇、結果品質の低下が発生する可能性があります。

Red Hat AI Enterprise は、オンプレミス、ハイブリッドおよびマルチクラウド環境、ネットワークエッジなど、さまざまな環境で最適化されたモデル推論を提供します。これにより、使用するリソースを減らし、推論コストを削減し、精度を維持しながらモデルのデプロイを効率化できます。

これを可能にするのが、vLLM で駆動する最適化されたランタイムです。vLLM は、GPU メモリをより有効に利用してスループットとレイテンシーを最適化することで、生成 AI ベースのアプリケーションの出力を高速化する推論サーバーです。また、モデルのサイズと計算要件を軽減してコストを最適化するのに役立つ高度な圧縮技法も使用できます。

このプラットフォームでは、拡張 (モデル数を数個から数十へ、ユーザー数を数十から数百へ) が必要となった場合、llm-d を使用できます。このテクノロジーは、vLLM の中核となるイノベーションを基盤とし、処理量の増加に応じてパフォーマンスを向上させることで、生成 AI モデルでコスト効率が高く予測可能なパフォーマンスを大規模に実現できる分散推論フレームワークを提供します。

プライベートな企業データで AI モデルの精度を改善

一般に公開されている生成 AI モデルは多様な機能を備えていますが、企業が必要としているのは、自社データでトレーニングされた、ドメイン固有のユースケースに対処するための AI モデルです。

Red Hat AI Enterprise は、企業がモデルの効率的なカスタマイズをサポートしたり、AI エンジニアやデータサイエンティストがモデルの応答の精度と関連性を向上させるのを支援したりするための最適化された一貫性のあるエクスペリエンスを提供します。これにより、モデルを組織のデータに接続し、検索拡張生成 (RAG) のユースケースに活用することが可能になります。

このプラットフォームはファインチューニングや継続的学習、強化学習もサポートしています。さらに、ドキュメント処理と解析、合成データ生成、モデル評価とタスクに対するモジュール式の柔軟なアプローチを提供します。

エージェント型 AI ワークフローでイノベーションを加速

AI テクノロジーは、比較的短期間のうちに急速に進化してきました。まず予測型 AI により多くの新しいユースケースが生まれ、続いて生成 AI モデルが普及し、そして現在では、AI 戦略をさらに拡張し、企業により多くのビジネス価値をもたらす次の進化へとつながるエージェント型 AI ワークフローに焦点が当てられています。

Red Hat AI Enterprise は、企業がエージェント型 AI を活用して価値を実現できるよう支援する、アジャイルで安定したプラットフォームを提供します。このプラットフォームを支えているのは、統合されたアプリケーション・プログラミング・インタフェース (API) レイヤー、エンタープライズユースケース向けにエージェント型 AI ワークフローの構築を効率化するすぐに使えるコンポーネント、専用のユーザーエクスペリエンス、エージェント型 AI システムのデプロイメントと管理をサポートする柔軟でスケーラブルな基盤です。

また、このプラットフォームはモデルコンテキストプロトコル (MCP) のサポートも提供します。これはエージェント型デプロイメントに不可欠なコンポーネントであり、幅広いツールや機能と大規模言語モデル (LLM) をつなぐ標準化された変換ツールとして機能します。

柔軟性と一貫性を維持しながら AI をハイブリッドクラウド全体に拡張

多くの業界の組織が、初期のテストユースケースで AI の価値を証明することに成功しています。しかしその多くが、コスト効率や、規制遵守、プライバシー、セキュリティへの準拠を維持しながらその価値を企業全体に拡張することは、いまだできていません。ハイブリッド環境やマルチクラウド環境などの多様な IT 環境や、ネットワークエッジでの AI デプロイメントが含まれる場合には、これは特に困難になります。

Red Hat AI Enterprise は、予測型および生成 AI モデルのトレーニング、チューニング、デプロイ、管理、拡張を、あらゆるハードウェアや IT 環境の全体で組織の AI ワークロード戦略に最も適した場所で行うために必要な制御と一貫性を提供します。

このプラットフォームは、任意のハードウェア、相手先ブランド製造会社 (OEM)、クラウドプロバイダー、データセンターを組み合わせられた環境であらゆる AI モデルをサポートする柔軟性を備えています。これを可観測性および監視機能の包括的なスイートと併用すれば、パフォーマンス最適化、コスト制御、ガバナンス維持、セキュリティ重視を堅持しつつ、ビジネスニーズの進化に応じて拡張していくことができます。

オンプレミスの AI デプロイメントでデータ主権を維持

世界中で、AI の使用やデータの保存および処理方法に関する政府の規制が次々に施行されています。そのため、多くの企業は主権とプライバシーを維持するために、AI をオンプレミスにデプロイする方法を模索しています。

機密性の高いプライベートデータをオンプレミスで保管および処理することにより、進化し続ける多くの規制に準拠できるだけでなく、攻撃対象領域を縮小し、セキュリティを強化してそれらのデータのプライバシーを保護することができます。

Red Hat は AI セキュリティに対する包括的で階層化されたアプローチを提供しており、組織は自信を持って AI イノベーションを導入できます。Red Hat AI Enterprise はこのセキュリティアプローチを信頼できるオープン・ハイブリッドクラウド基盤上で提供し、AI ライフサイクル特有のセキュリティおよび安全性の課題に対処できるよう支援します。

あらゆるモデル、あらゆるハードウェア、そしてすべてのクラウドおよびエッジ環境 (エアギャップ環境を含む) の全体でセキュリティ重視の GPU 共有をサポートしており、真の主権の実現を促進します。組織がセキュリティを重視したプライベートクラウド環境に AI をデプロイできるよう、Red Hat はソブリンクラウド・プロバイダーと緊密に連携しています。

また、Red Hat AI Enterprise には、監査可能な信頼性の実現をモデルの説明可能性、公平性、出力バイアスの適用を通じてサポートするツール、および再現可能なパイプラインと監査への対応によってライフサイクルのトレーサビリティを実現するツールも含まれています。

信頼できる Kubernetes プラットフォーム上に AI を構築

Red Hat AI Enterprise はこの完全な AI 機能セットを、Kubernetes という信頼される基盤上に構築された、すぐ使えるエンドツーエンドの統合 AI スタックの一部として提供します。

そのため、Kubernetes とコンテナ化をすでに導入している企業は、オンプレミスのデータセンター、ハイブリッド環境またはマルチクラウド環境、エッジでの AI デプロイメントのすべてで、馴染みのある一貫したエクスペリエンスを通じてこれらを使用できます。チームは一元化された単一のプラットフォーム上で、すでに知っているツール、フレームワーク、スキルを使用して、クラウドネイティブの先進的な AI を活用したアプリケーションを構築、トレーニング、デプロイ、管理できます。

Red Hat AI Enterprise の価値の詳細

Red Hat AI Enterprise は、AI 導入のあらゆる段階をサポートする統合プラットフォームにより、効率的でコスト効率の高い AI をハイブリッドクラウド全体にデプロイできるように支援します。[詳細についてはこちらをご覧ください。](#)



Red Hat について

エンタープライズ・オープンソースソフトウェア・ソリューションのプロバイダーとして世界をリードする Red Hat は、コミュニティとの協業により高い信頼性と性能を備える Linux、ハイブリッドクラウド、コンテナ、および Kubernetes テクノロジーを提供しています。Red Hat は、クラウドネイティブ・アプリケーションの開発、既存および新規 IT アプリケーションの統合、複雑な環境の自動化および運用管理を支援します。[受賞歴のあるサポート](#)、トレーニング、コンサルティングサービスを提供する Red Hat は、[フォーチュン 500 企業に信頼されるアドバイザー](#)であり、オープンな技術革新によるメリットをあらゆる業界に提供します。Red Hat は企業、パートナー、およびコミュニティのグローバルネットワークの中核として、企業の成長と変革を支え、デジタル化が進む将来に備える支援を提供しています。

アジア太平洋 +65 6490 4200 apac@redhat.com	インドネシア 001 803 440 224	マレーシア 1800 812 678	中国 800 810 2100
オーストラリア 1800 733 428	日本 03 4590 7472	ニュージーランド 0800 450 503	香港 800 901 222
インド +91 22 3987 8888	韓国 080 708 0880	シンガポール 800 448 1430	台湾 0800 666 052

f fb.com/RedHatJapan
X twitter.com/RedHatJapan
in linkedin.com/company/red-hat