

生成 AI のための基盤モデル・プラットフォーム

主なメリット

- ▶ ユーザーは InstructLab を使用して大規模言語モデル (LLM) を更新および強化できる
- ▶ LLM をプロプライエタリーなデータと安全かつ確実に整合させ、ビジネス要件に合わせてモデルをカスタマイズできる
- ▶ 生成人工知能 (生成 AI) をすぐに使い始め、信頼できるセキュリティ重視の Red Hat® Enterprise Linux® プラットフォームで成果を得る
- ▶ インストールとアップデートのためのブート可能な Red Hat Enterprise Linux コンテナイメージとしてパッケージ化されている

製品概要

Red Hat Enterprise Linux AI は、生成 AI ビジネスユースケース向けの LLM を開発、テスト、デプロイするための、エンタープライズグレードの生成 AI 基盤モデル・プラットフォームです。

Red Hat Enterprise Linux AI には次が含まれます。

- ▶ オープンソース LLM の Granite ファミリー。
- ▶ InstructLab モデル・アラインメント・ツール。LLM ファインチューニングに対するコミュニティ主導型アプローチを提供します。
- ▶ Red Hat Enterprise Linux のブート可能イメージと、生成 AI ライブラリおよび依存関係 (PyTorch や、NVIDIA、Intel、AMD 用の AI アクセラレーター・ドライバー・ソフトウェアなど)。
- ▶ Red Hat が提供するエンタープライズレベルのテクニカルサポートとモデルの知的財産の補償。
- ▶ Red Hat Enterprise Linux AI は、信頼できる Red Hat Enterprise Linux プラットフォームを提供し、生成 AI 採用プロセスを開始して結果を得るために必要なコンポーネントを追加。

AI の未来はオープンで透明

Red Hat Enterprise Linux AI には、Red Hat が完全に補償するオープンソースの Granite 言語モデルおよびコードモデルが含まれます。オープンソースの Granite モデルは、さまざまな生成 AI のユースケースに合わせて、コストとパフォーマンスが最適化されたモデルを組織に提供します。Granite モデルは、Apache 2.0 ライセンスの下でリリースされました。モデルがオープンソースであることに加えて、モデルのトレーニングに使用されるデータセットも透明でオープンです。

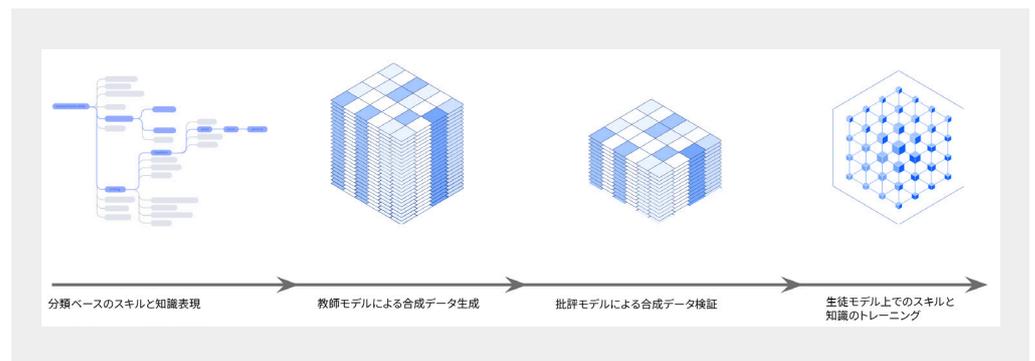
表 1. Red Hat Enterprise Linux AI の Granite モデル

IBM Granite 言語モデル	Granite-7B-Starter Granite-7B-RedHat-Lab
IBM コードモデル	Granite-8B-Code-Instruct Granite-8B-Code-Base

アクセスしやすい生成 AI モデルトレーニングで価値実現までの時間を短縮

Red Hat Enterprise Linux AI には、オープンソースの Granite モデルに加えて、Large scale Alignment for chatBots (LAB) 技術に基づく InstructLab モデル・アラインメント・ツールも含まれています。InstructLab により、組織内のチームはスキルや知識を効率的に LLM に提供し、ビジネスの特定のニーズに合わせてそれらのモデルをカスタマイズできるようになります。

- ▶ スキル：何かを行う方法をモデルに教えることを目的とした機能ドメイン。スキルは2つのカテゴリに分類されます。
 - ▶ 構成的スキル
 - ▶ AI モデルに特定のタスクや機能を実行させます。
 - ▶ グラウンディング (コンテキストを含む) または非グラウンディング (コンテキストを含まない) のいずれかです。
 - ▶ グラウンディングの例としては、マークダウン形式のテーブルを読み取る機能をモデルに提供するスキルを追加することが挙げられます。
 - ▶ 非グラウンディングの例としては、モデルに韻を踏む方法を教えるスキルを追加することが挙げられます。
 - ▶ 基礎的スキル
 - ▶ 数学、推論、コーディングなどのスキルです。
- ▶ 知識：より高い正確性で質問に答えるための追加のデータと情報をモデルに提供する、データおよび事実。



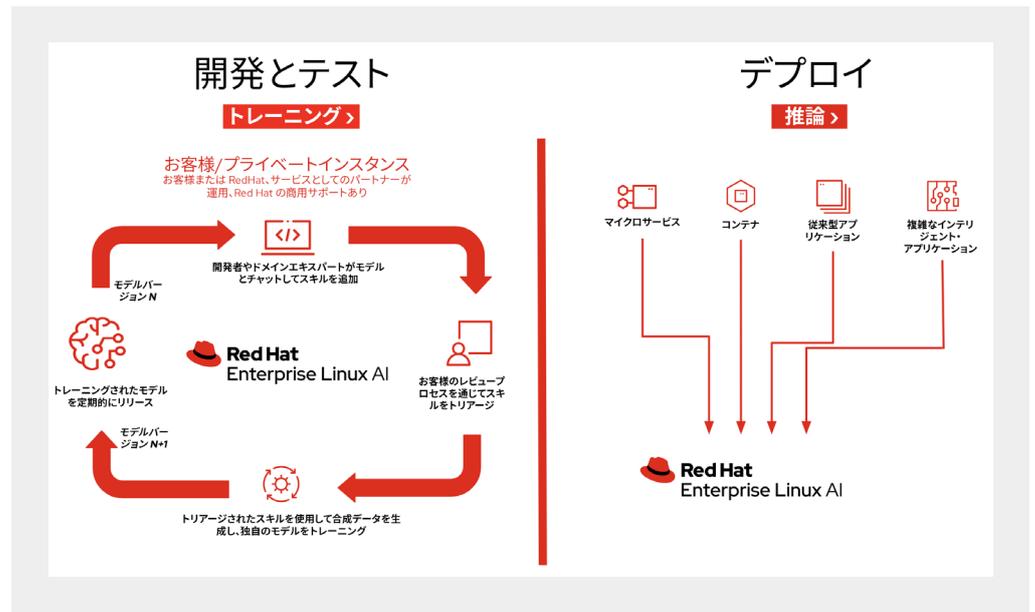
上の図は、InstructLab でモデルをファインチューニングするワークフローの概要を示しています。

1. スキルと知識の貢献は、分類ベースのデータリポジトリに配置されます。
2. LLM を正常に更新および変更するのに十分な量のデータセットを作成するために、分類データを使用して大量の合成データが生成されます。
3. この合成データ出力は、批評モデルによってレビュー、検証、ブルーニングされます。
4. モデルは、人間が生成した手動入力に基づく合成データを使用してトレーニングされます。

InstructLab は、LLM のファインチューニングに通常必要とされるデータサイエンスの専門知識が不足している開発者やドメインエキスパートでも利用できます。InstructLab 手法を使用すると、チームは選択したモデルに対して、中でもビジネスユースケースの要件に適したデータまたはスキルを追加し、共同でトレーニングできるため、価値実現までの時間を短縮できます。

場所を問わないトレーニングとデプロイ

Red Hat Enterprise Linux AI は、必要なすべてのツールと、ハイブリッドクラウド上のどこでもデータが存在する場所でモデルをトレーニング、チューニング、デプロイできる機能を提供することで、組織が概念実証からプロダクション・サーバーベースのデプロイメントへと移行するプロセスを加速できるよう支援します。デプロイされたモデルは、社内のさまざまなサービスやアプリケーションで使用できるようになります。



組織の準備が整えば、Red Hat Enterprise Linux AI は Red Hat OpenShift® AI への足掛かりともなります。このプラットフォームにより、Red Hat Enterprise Linux AI デプロイメントで使用されているものと同じ Granite モデルと InstructLab アプローチを使用して、分散クラスタ環境全体でモデルを大規模にトレーニング、チューニング、提供できます。

特長とメリット

特長	メリット
Apache 2.0 ライセンスに基づいてオープンソース化されている、フルサポートされた Granite 言語モデルおよびコードモデル	オープンソースで透過的な LLM、およびオープンにアクセスできるトレーニングデータにより、データの透明性が高まり、データのコンテンツとソースに関する倫理的懸念に対処し、最終的にビジネスリスク全体が軽減されます。

特長

メリット

Granite モデルの知的財産補償

Red Hat Enterprise Linux AI 内の Granite モデルに対する補償は、これらのモデルの厳密な開発とテストに対する Red Hat と IBM の強い自信を反映しています。この補償はお客様にとって強力な保証となり、成功に向けた Red Hat の取り組みに対する大きな信頼と自信を持って Granite モデルを使用できるようになります。

InstructLab LLM アライメントツールによるスケーラブルでアクセスしやすいモデル・ファインチューニング

InstructLab は、LLM をファインチューニングするためのアクセスしやすい方法を提供します。これにより、データサイエンスに関する深い専門知識の必要性が軽減され、組織内のさまざまな役割の人が貢献できるようになります。その結果、企業は生成 AI を導入し、価値実現までの時間を短縮し、投資対効果を最大化できます。

最適化されたブート可能なモデル・ランタイム・インスタンス

Red Hat Enterprise Linux AI は、ブート可能なコンテナイメージとして提供されます。これは、Red Hat Enterprise Linux のイメージモードと呼ばれるデプロイメント手法です。このテクノロジーにより、インストール、構成、更新の複雑さが軽減され、シンプルなセットアップと変更管理プロセスが可能になります。

生成 AI パッケージの依存関係と AI ハードウェアのソフトウェアドライバ

PyTorch、vLLM、NVIDIA ドライバーなどの必須パッケージやドライバを含む包括的なツールセットを使用して生成 AI の使用をすぐに開始し、初日から生成 AI のビジネスユースケースに取り組む準備を整えることができます。



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

オーストラリア

1800 733 428

インド

+91 22 3987 8888

インドネシア

001 803 440 224

日本

03 4590 7472

韓国

080 708 0880

マレーシア

1800 812 678

ニュージーランド

0800 450 503

シンガポール

800 448 1430

中国

800 810 2100

香港

800 901 222

台湾

0800 666 052

[f fb.com/RedHatJapan](https://fb.com/RedHatJapan)
[X twitter.com/RedHatJapan](https://twitter.com/RedHatJapan)
[in linkedin.com/company/red-hat](https://in.linkedin.com/company/red-hat)