

適切な AI モデルを選択するための 4 つの考慮事項

人工知能 (AI) と生成人工知能 (生成 AI) の技術進化に伴い、エンタープライズ組織ではそれらを活用した価値実現へのプレッシャーが高まっています。そのプロセスにおいて重要なのが、適切な生成 AI モデルの選択です。これによって AI を組織の戦略的目標の達成にどれだけ効果的に活用できるかが決まります。このチェックリストでは、AI モデルを選択する際に考慮すべき 4 つの重要なポイントを示し、[Red Hat® Enterprise Linux® AI](#) が、お客様独自のユースケースに合わせたソリューションを構築するために必要なモデルをどのように提供できるかを説明します。

1 どのようなサイズのモデルを使用すべきか

組織とその特定のユースケースに適したモデルサイズの選択は、きわめて重要な出発点となります。そのモデルの実行コスト、管理に必要なスキル、結果の精度に直接影響を与えるからです。

より多くのパラメーターを持つ大規模なモデルを使用すると、複雑なシナリオを評価する能力を高めることができます。しかし、大規模なモデル (そしてそのモデルを動かすデータ) には、より多くの計算リソースと複雑なインフラストラクチャが必要となり、運用コストに加えてモデルの管理に必要なスキルが大幅に増加する可能性があります。

お客様の企業、ユースケース、または業界に関連するデータで大規模なモデルをカスタマイズする必要がある場合には、これらのコストや必要とされるスキルはさらに増加する可能性があります。

Red Hat Enterprise Linux AI に付属する [Granite ファミリー大規模言語モデル \(LLM\)](#) のような、より小規模でより特化したモデルを使用することで、コスト効率の高いアプローチを実現できます。これらのモデルはお客様のエンタープライズデータに基づいてトレーニングできるため、多大なコスト、スキル、時間を費やすことなく、特定のニーズに対応できます。

2 モデルの透明性はどの程度か

生成 AI モデルの精度と効率に影響を与えるもう 1 つの重要な要素は、モデルのトレーニングに使用されるデータです。これには、データの出所、所有権、品質、アクセス性、信頼性、関連性、多様性、量などが含まれます。

トレーニングに使用されたデータに関する透明性のないモデルを使用すると、組織は重大なリスクにさらされることになります。たとえば、データが不正確なために結果が不正確になるとか、さらに深刻な問題としては、著作権で保護された素材がトレーニングで使用されているためにモデルの使用が法的に問題となりかねないなどがあります。

Red Hat は Red Hat Enterprise Linux AI に付属する Granite ファミリーモデルのような、透明性が高いデータソースとデータセットを備えたモデルの使用を推奨しています。Red Hat Enterprise Linux AI はトレーニングデータセットとモデルコンテンツの完全な透明性と、Red Hat による完全な補償とサポートを提供します。

これによって組織は、モデルがすべての規制を考慮して倫理的に構築およびトレーニングされたこと、そして機密データの保護が実証されていることを保証する、信頼できるソースによって裏付けられているという安心感を得て、生成 AI の取り組みを自信を持って推進できます。

3 使用するべきモデルはオープンソースか、プロプライエタリーか

LLM には、オープンソースライセンス (Apache 2.0、一般公開ライセンスなど) や、通常はプロプライエタリーまたはサブスクリプションベースの商用ライセンスなど、複数の種類のライセンスがあります。

オープンソースモデルとプロプライエタリーモデルのどちらを選択するかは、以下に示すような重要な結果に大きな影響を与える可能性があります。

- ▶ 運用効率およびコスト
- ▶ 許容されるカスタマイズレベル
- ▶ そのモデルを使用して構築されたアプリケーションとサービス、およびそこに入力されたデータの所有権

プロプライエタリーのモデルは運用コストが高く、ユースケースや業界に合わせたトレーニング/チューニングが難しく、カスタマイズ性が限られており、コミュニティによる採用や貢献の恩恵を受ける小規模なオープンソースモデルが提供できるような効率性に欠けることがあります。さらに、商用ライセンスを使用する場合、そのモデルを使用して構築したサービスやアプリケーションを販売する際にモデルの所有者への料金の支払いが必要になったり、モデルに入力されたデータの所有権が所有者に譲渡されたりすることがあります。

Red Hat Enterprise Linux AI を通じて提供されるオープンソースライセンスの Granite LLM は、運用効率と資金効率に優れています。また、データの完全な所有権を保持しながらモデルを自由に使用、変更、配布できます。

4 モデルをどのようにトレーニングすればよいか

モデルをカスタマイズすることで、効率性、精度、コストの最小化を実現しながら、特殊なユースケースのニーズを満たすことができます。そのモデルをニーズに合わせてどのように調整するかは、モデルチューニングの効率と効果に大きな影響を与える可能性があります。

Red Hat Enterprise Linux AI では、サポート付きでライフサイクルベースの [InstructLab](#) ディストリビューションを通じて、モデルアライメントツールにアクセスできます。

LLM 向けの InstructLab トレーニング手法は、これらのモデルの開発における課題と、従来の LLM トレーニングに見られるようなスケールアップの課題に対処します。これは、分類に基づいた合成データ生成プロセスと、モデルパフォーマンスを向上させる多段階チューニング・フレームワーク、開発者があらゆるモデルに新しいスキルと知識を共同で提供できるオープンソース・コミュニティによって実現されます。

InstructLab モデルアライメントツールを使用することにより以下が実現します。

- ▶ LLM のアライメントを改善するためのコスト効率の高いソリューションが提供される
- ▶ コミュニティアプローチによりオープンソースモデルの改善が促進される
- ▶ アクセス可能なツール、トレーニング、コミュニティの知識を通じて、独自のデータセットでモデルをカスタマイズするプロセスを民主化する
- ▶ トレーニングデータセットとモデル IP の完全な所有権を保持できる

さらに詳しく

[この e ブック](#) では、お客様が AI イノベーションを開始できるように Red Hat が提供する支援についてご紹介しています。

今すぐ始める

[Red Hat の担当者にご相談](#) ください。お客様の組織が AI の取り組みを始めるうえで Red Hat が提供できる支援についてご説明します。



Red Hat について

Red Hat は、[受賞歴のある](#) サポート、トレーニング、コンサルティング・サービスをお客様に提供し、複数の環境にわたる標準化、クラウドネイティブ・アプリケーションの開発、複雑な環境の統合、自動化、セキュリティ保護、運用管理を支援します。

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

[jp.redhat.com](#)

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