

オープンソースの小規模言語モデルを使用する 4 つの理由

小規模言語モデルが企業の AI 戦略を変革

プロプライエタリーな大規模言語モデル (LLM) は汎用アプリケーション向けには優れていますが、エンタープライズ人工知能 (AI) ソリューションには必ずしも最適ではありません。その多大な計算要件、不透明な意思決定プロセス、高額なライセンスコストにより、柔軟性が制限され、運用が複雑になる可能性があります。小規模言語モデル (SLM)、特にオープンソースの原則に基づくモデルは、カスタマイズされた AI ソリューションの開発、データの制御の維持、およびコストの効果的な管理を求める組織にとって代替手段となります。

オープンソース SLM が今後の AI プロジェクトに最適な選択肢となる 4 つの理由を以下に示します。

1 コミュニティによるイノベーションへのアクセス

オープンソース SLM は柔軟性、コラボレーション、イノベーションを兼ね備えており、適応性の高い特殊な AI アプリケーションを構築するための基盤を提供します。オープンソース AI プロジェクトは、ソフトウェア・コンポーネントと事前トレーニングされたモデルの重みの両方へのアクセスを提供します。これにより、開発者や研究者のグローバルコミュニティと協力して、生成 AI テクノロジーを継続的にリファインおよび改善することができます。この共有されたイノベーションにより、先進的で高度なツールを使用して、エンタープライズ AI ソリューションの技術的ニーズを満たすようにそれらのテクノロジーを調整できるのです。

生成 AI モデルの **IBM Granite ファミリー**などのオープンソース SLM を使用すると、自分の知識やドメインの専門知識を提供して基盤モデルに直接貢献できます。プロプライエタリーな LLM の更新を待つのではなく、オープンソース SLM を積極的にカスタマイズして、AI アプリケーションにおける関連性とパフォーマンスを向上させることができます。このインタラクティブなアプローチにより、イテレーションを高速化し、進化するビジネスニーズに合わせてモデルを最新の状態に保つことができます。

オープンソース SLM は、オンサイト・データセンターとパブリッククラウド・インフラストラクチャにわたる動的な環境でのデプロイメントに不可欠な柔軟性を提供します。モデルを完全に制御できるため、厳しいコンプライアンス環境からリアルタイムでの AI 処理まで、さまざまなデプロイシナリオに合わせてモデルを最適化できます。また、オープンソース SLM では AI テクノロジースタックの制御を維持できるため、技術的ニーズやビジネスニーズの変化に合わせて、革新的な AI ソリューションを常に適応および拡張させることができます。

2 トレーニングデータの制御

オープンソース SLM は、プロプライエタリーな手段と比較して透明性が高くなります。信頼できるプロバイダーはこれらのモデルの事前トレーニングに使用されたデータを開示しています。そのため、モデルの品質を徹底的に評価し、有害な情報やバイアスのかかった情報が含まれていないことを確認できます。この透明性により、モデルの適応とデプロイについて情報に基づいた意思決定を行うことができます。その結果、自社独自の機密データを組み込む前に、AI ソリューションが倫理基準とビジネス目標を満たしていることを確認できます。

さらに、SLM はオンサイトのデータセンターとエンタープライズ IT 環境全体のプライベートクラウド・リソースにデプロイできるため、トレーニングデータの完全な制御を維持できます。この制御性は、機密性の高いデータや規制されたデータを扱う組織にとって非常に重要です。自社独自の情報が外部プロバイダーに決して公開されないようにできるからです。また、自社独自の環境内で生成 AI モデルを管理することで、アクセスを制御し、法令遵守を効率化し、データセキュリティを強化し、AI ソリューション全体の透明性を高めることができます。

最後に、**IBM Granite** ファミリーのモデルを使用する場合、提供しているオープンソースソフトウェアまたは AI モデルが第三者の知的財産権を侵害しているという主張を受けたお客様に対する補償を提供する **保証ポリシー** が適用されます。これらのモデルとベンダーを選択すると、変化する複雑な AI テクノロジー環境における組織の保護をさらに強化できます。

3 AI ソリューションのカスタマイズ

オープンソース SLM を使用すると、特定のビジネス要件に合わせて AI ソリューションを迅速かつ効率的に開発できます。対象を絞ったユースケース向けに設計および構築されたこれらのモデルにより、汎用 LLM の複雑さとリソースの要求を回避しながら、ドメイン固有の課題に正確に対処できます。

企業データを使用して SLM をチューニングすることで、組織の知識とドメインの専門知識をモデルのパラメーターに直接組み込むことができます。このアプローチにより、SLM のレスポンスの関連性が向上し、再トレーニングの頻度とコストが削減され、重要な AI アプリケーションとサービスの開発タイムラインが短縮されます。

SLM は LLM と比較してサイズがコンパクトでデータ要件が少ないため、カスタマイズが容易であり、特定のタスクやドメインに最適化された正確で効率的なモデルを開発できます。また、リソースに制約のある環境やエッジデプロイメントでは、SLM によってリアルタイム・アプリケーションをユーザーデバイス上で直接実行できるため、開発が単純化され、外部のクラウド・インフラストラクチャが不要になります。

IBM Granite モデルなどの SLM により、実験からプロダクションへの移行も効率化されます。SLM はさまざまなハードウェアおよびソフトウェア・インフラストラクチャと容易に統合できるため、生成 AI ソリューションをエンタープライズ IT 環境に合わせて調整することができます。この適応性は、デプロイメントとパフォーマンスの制御を維持しながら、運用の複雑さを軽減するのに役立ちます。

4 AI モデルのコストの削減

多くのエンタープライズ組織にとって、経費を効果的に管理するには、AI の計算需要を軽減することが重要です。オープンソース SLM は、LLM と比較してトレーニングと推論のコストや必要な計算能力を低く抑えつつ、高度な生成 AI ソリューションに必要なパフォーマンスを提供します。

SLM は多くの場合、主要な LLM の数千分の 1 のサイズであり、必要なコンピュートリソース、データ、エネルギーがはるかに少なく済みます。この効率性により、トレーニング時間の短縮、ファインチューニングの容易化、および AI 開発へのより持続可能なアプローチが実現します。

さらに、オープンソース SLM は、コストのかかるハードウェアのアップグレードを必要とせず、複数のプロジェクトや組織にわたって効率的に拡張できます。これらのモデルを既存の IT インフラストラクチャ内にデプロイすることで、パフォーマンスを犠牲にしたり、予算の制約を超えたりすることなく、カスタマイズされた AI ソリューションを作成できます。

コストの削減はインフラストラクチャ以外にも広がります。オープンソース SLM では、プロプライエタリーなモデルに関連するライセンス料が不要になり、ベンダーが課す制限や制約を受けることなく高いコスト効率で高度な生成 AI 機能を活用することもできます。

Red Hat と IBM のオープンソース SLM でイノベーションを実現

IBM によって開発され、Red Hat® Enterprise Linux® AI に組み込まれているオープンソース生成 AI モデルの Granite ファミリーは、エンタープライズ AI アプリケーションのさまざまな要求に対応できます。

オープンソースの生成 AI モデルの詳細について

タスク固有の SLM とオープンソースの生成 AI ソリューションについて詳しくは、e ブック『オープンソースモデルで AI イノベーションを最大化』をお読みください。



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。

アジア太平洋
+65 6490 4200
apac@redhat.com

オーストラリア
1 800 733 428

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
1 800 812 678

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