

ハイブリッドAIで 経済性とパフォーマンスを向上

著者

Steven Dickens

最高技術顧問 | The Futurum Group

Ron Westfall

研究ディレクター | The Futurum Group

2024年8月

提携企業





エグゼクティブサマリー

今日のIT部門とビジネス意思決定者（ITBDM）は、所属する企業のAI導入を先導するという前例のないタスクを任されています。選択肢は幅広く、革新のスピードはとても激しいものです。AI導入に最も適した方法を探求する中で、総合的な観点から最良の方法として、ハイブリッドAIモデルを選択する意思決定者はますます増えています。なぜハイブリッドAIなのでしょう？

ハイブリッドAIは、機械学習、深層学習（ディープラーニング）、ニューラルネットワークをベースとしたAIモデルを人間の専門知識で最適化または拡張したものであり、精度を向上させたドメイン固有のAIモデルや、精度または予測確率を最大限に高めたAIユースケース固有のモデルを作成するために使用します。ハイブリッドAIは、データの重要性を守り、モバイルやデスクトップデバイス、プライベートデータセンター、エッジコンピューティング、さらにはクラウドなど、必要なあらゆる場所にAIソリューションと機能をもたらすアイデアです。そのため、ユースケースとデータ管理戦略をもとに、セキュリティ、コスト、ネットワークインフラストラクチャ、データの重要性、信頼性などの要素を考慮しながら、AI機能を導入する場所と方法が決まります。

このアプローチは、単一モデルまたは単一インフラストラクチャメソッドの限界を克服するものであり、柔軟なデータ管理とアプリケーションサポートを求める顧客の要求と一致しています。ハイブリッドAIには、オープンソースAIとOpenAIのChatGPT機能などのクローズドソース機能が統合されており、生産性、俊敏性、データ保護、AIの民主化を促進します。

CIO、CTO、最高データ責任者、最高AI/データ責任者、アーキテクトなどの主要な意思決定者は、最初にまずインフラストラクチャと適切なサイズのモデルを適切に組み合わせて、問題を定義し、データソースを特定し、ユースケースとAIアプリケーションを選択する必要があります。通常、このプロセスには戦略開発、AIスタック、AIファクトリー、またはセンターオブエクセレンス（COE）の構築、そして選択されたソリューションの導入などが含まれます。

ハイブリッドAIを実装するには、AI技術、適切なアーキテクチャ、ビジネスに適したモデルを組み合わせたソリューションを設計できるように、幅広いパートナーエコシステムを有するエマージングテクノロジーを評価、パイロット化し、拡張する必要があります。

本誌では、Lenovo Smart AI for AI製品ポートフォリオとLenovoのビジョンが、ハイブリッドAIを組織全体に実装するという独自の要求を満たすことができる理由を検証しています。Lenovo Smart AI for AIは、AIPC、ワークステーション、AI対応ソリューション、専門知識、AI向けに最適化されたストレージ、コンピューティング、インフラストラクチャ、ソフトウェア、

および提携能力を適切な組み合わせと規模で提供し、検証済みのパートナーエコシステムと連携させることで、ビジネス成果を重視したアプローチを優先させます。これには、個人および業界のエッジ（PC、ワークステーション、事業拠点）、データセンター（CoLos、オンプレミス）、および高性能コンピューティング（HPC）環境全体をカバーする、カスタマイズされ、俊敏性、拡張性、エネルギー効率性に優れた業界およびユースケースに基づくAIソリューションを、パブリッククラウドとの統合や相乗効果と共に提供する能力が含まれます。

LenovoとNVIDIAは共に、AI向けに最適化された、エンタープライズAI環境を実現するAIファクトリーを構築しており、このAIファクトリーでは、AIアプリケーションをプライベートかつ大規模に、安全に開発、デプロイおよび管理することが可能です。

急速に進化するハイブリッドAI環境

組織も従業員も生成AI（GenAI）の変革力を活用したいと考えており、初期の実験やAI/MLパイロット段階から自社のデータの活用へと移行しています。しかしながら、法律、コンプライアンス、ガバナンス上の課題、専門知識のギャップ、データのサイロ化、セキュリティ上の懸念、不十分なAIインフラストラクチャ、エコシステムの複雑さによって、その普及が妨げられることが少なくありません。予算上の制約も問題をさらに複雑化しています。CIOの96%が、今後1年間でAIへの投資が増加すると予想していますが、IT予算全体が10%以上増加すると予想しているのは20%にすぎません（CIO Study 2024）。

ハイブリッドAIは、ビジネスAI戦略の重要なコンポーネントとして台頭しており、組織は自社の目標に沿ったAIおよびデータソリューションを開発し、AIをデータソースの近くにもたらし、リアルタイムの相互作用とエクスペリエンス創作ポイントを実現することができます。無料のAIツール、公開基盤モデル、既存のクラウドサブスクリプションなどによって初期導入は容易ではあるものの、多くの組織は永遠に続く概念実証サイクルとコスト上昇に陥っています。これは多くの場合、プライベートクラウドとパブリッククラウド環境全体でAIを統合するための堅牢なハイブリッドAI戦略の欠如、カスタマイズされたAIモデルの欠如、質の高い組織データへの限定的なアクセス、AIコンピューティングのサイジングとプランニングが不十分であることが原因です。

私達は、GenAIとAIをデバイス上を含めたエッジでローカルに実行することは、データやエッジロケーションが分散している組織（小売、製造、病院、ハイブリッド職場など）、リアルタイムの分析とアクセスを必要とする企業、および切断されたコンピューティングニーズを持つ組織（遠隔の石油およびガス事業、高度に保護されたエッジ環境など）にとって特に不可欠であると考えています。

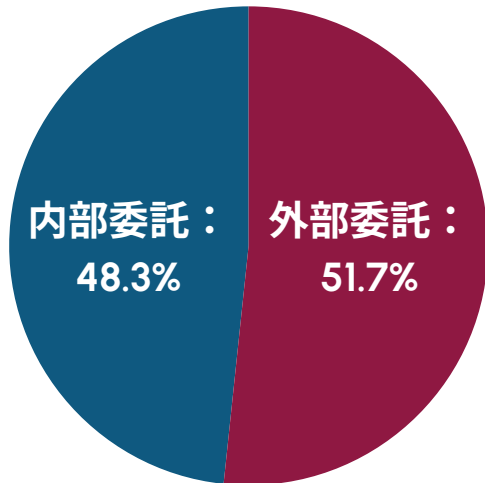
エッジロケーションとデバイスからは膨大な量のデータが生成されますが、そのほとんどは分析されていません。AIおよびGenAIは、これらの未開拓のリソースを活用できる大きな機会となります。エッジユースケース向けに開発されたAIアプリケーション、ますます効率化する小型モデル、AIPCなどのAI対応デバイスなどが登場する中で、エッジコンピューティングは現実化しています。これにより、リアルタイムで分析するためにすべてのデータまたはAIモデルをパブリッククラウドや一元化されたデータセンターに転送する必要がなくなり、セキュリティ、遅延および帯域幅上の制約に対処できます。

ローカルGenAI処理により作成時点でのデータ分析が可能になり、遅延が削減されて効率が向上します。これは、製造プロセスやロボット工学などのリアルタイムアプリケーションにとって不可欠です。さらに、プライバシーに関する懸念により、機密データをデバイス上に残すためにローカル処理が必要になることも少なくありません。

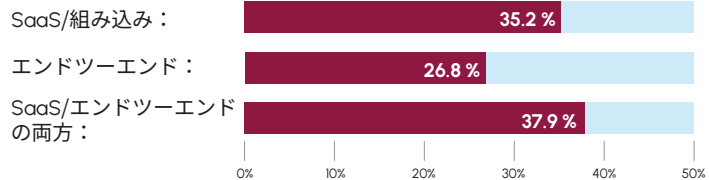
クラウドベースの大規模言語モデル（LLM）には、組織がますます避けようとしている課題があります。これには、ハルシネーションのリスク（論理的には矛盾するが、説得力のある出力）と、学習データからの不正確さの伝播などの問題があります。さらに、クラウド管理型サービスのインフラストラクチャと実装では、組織に可能な制御や可視性が限られているため、セキュリティ上の懸念が高まっています。

よく知られている例としては、2023年の3月に発生したOpenAI漏洩問題があり、OpenAIのChatGPTがオープンソースライブラリの脆弱性により機能停止に陥り、顧客の支払い情報が流出する可能性が生じました。このインシデントは、クラウドベースのAIサービスを利用する際には、厳格なLLM保護テストと堅牢なサイバー脅威防御メカニズムを設けることの重要性を強調しています。

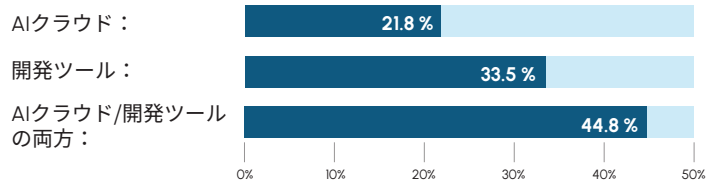
さらに、AIサービスをデプロイする際には、組織はバランスの取れたアプローチを取っていることがわかりました。



外部委託



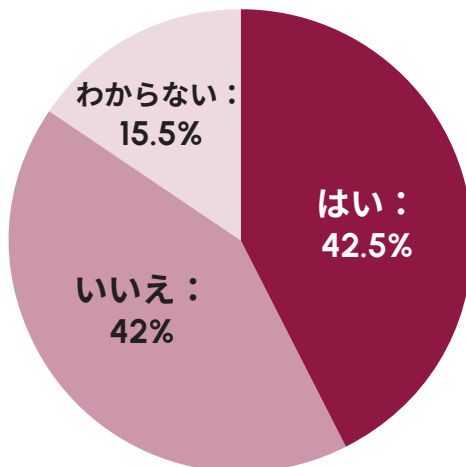
内部委託



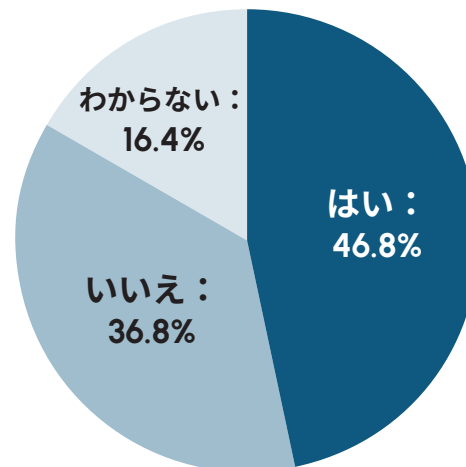
出典：The Futurum Group

ハイブリッドAIへの関心が急速に高まっており、組織はベンダーを変更したり新たに追加することを検討し始めており、新たな競争が生まれています（2023～24年）。

外部委託



内部委託



出典：The Futurum Group

ハイブリッドAIが成功するかどうかは、組織のニーズに合わせてカスタマイズされたAIソリューションをキュレートするエコシステムが果たす重要な役割にかかっています。エコシステムが、多様なバリューチェーンの利害関係者を結びつけ、ハイブリッド環境全体でAIシステムをシームレスに統合することで、オーダーメイドのAIソリューションの作成へとつながります。AIソリューションには、テクノロジーと専門知識の組み合わせが必要になることが多々あり、AIの導入が進むにつれて、ビジネスの課題に直接対処できるソリューションを開発するために、幅広いソフトウェア、アプリケーション、ツール、および機能にアクセスできるエコシステムに依存します。

さらに、ハイブリッドAIに固有の柔軟性は、企業内のAIに関わるさまざまなペルソナの多様なニーズを満たすために不可欠な要素となります。AIアプリケーション開発者やデータサイエンティストは、その配分に関係なく、使い慣れたAIツールやフレームワークを使用することを好みます。エッジでは、小売店や製造などの業界でデータを迅速に処理する能力とともに、AIシステムによるリアルタイムの対応も期待され、求められています。

しかしながら、ハイブリッドAIを効果的に使用するには、テクノロジー、人材、プロセス、責任あるAIアプローチを考慮した綿密な計画と実行が必要です。企業は、サイロ化されたイニシアチブ、一貫性のないAIソリューションの選択と予算に頭を悩ませており、その一方でITチームはインフラストラクチャを近代化し、さまざまなAIコンピューティングの要求、コンプライアンス、セキュリティ基準、およびハイブリッド運用を実行する必要性を満たすのに苦労しています。データ管理とパブリッククラウドの使用が制御されなくなるにつれて、セキュリティ、ガバナンス、およびクラウドにかかるコストの増大が差し迫った課題となっています。これらの課題と、AIから有意義なビジネス成果を達成するために必要なスキルと変更管理との間でバランスを保つ必要があります。

たとえば、データサイエンティストやMLエンジニアはデータの品質や規制、コンプライアンスに多くの時間を費やすべきではありません。企業はAI関連のIT意思決定にIT&Oチームを関与させる必要があります。同様に、AIを企業に導入しようとするITチームからのボトムアップアプローチでは、ビジネスニーズやユースケースが十分に理解されることなく、プロビジョニングの過不足が生じる可能性があります。



LenovoとNVIDIAの共同エンジニアリング製品 ポートフォリオ

AI業界の成長は複数の要因によって牽引されており、中でも自動化と最適化に対する需要の高まりは各分野で最重要課題となっています。本稿で概説したように、企業はコンピューティングパワーとクラウドインフラストラクチャの進歩を利用して、より効率的なAIアプリケーションを実装しています。このシフトにより、プロセスの簡素化に役立つコンサルティング、設計、展開、保守サービスなど、カスタマイズされたAIソリューションに対する需要が急増しています。

Lenovo Smart AI for Allポートフォリオは、AI対応デバイス、スマートインフラストラクチャ、堅牢なクラウドソリューションでこれらの進化するニーズに対応するものであり、さまざまなプラットフォーム間でのシームレスなAI統合を促進します。LenovoとNVIDIAの戦略的パートナーシップにより、AI分野のイノベーションが加速しています。2社によって共同エンジニアリングされたポートフォリオにより、AIの採用が簡素化され、より迅速な成果がもたらされることで、ハイブリッドAI環境におけるパフォーマンスと効率が向上し、業界の成長をさらに促進します。

本レポートに関する重要な情報

寄稿者

Steven Dickens

最高技術顧問 | The Futurum Group

Ron Westfall

研究ディレクター | The Futurum Group

出版者

Daniel Newman

CEO | The Futurum Group

お問い合わせ

本レポートに関してご質問がある場合は当社までご連絡ください。
The Futurum Groupから早急に回答いたします。

引用

本稿は、正式に認可されている報道機関やアナリストが引用できますが、引用する際には著者名、著者の役職、および「The Futurum Group」を明記してください。報道機関およびアナリスト以外の方が引用される場合は、The Futurum Group から書面による事前の許可を得る必要があります。

ライセンス

本文書は、添付資料を含めThe Futurum Groupが所有しています。本出版物は、The Futurum Groupから事前による書面による許可を得ることなく複製、配布、または共有することはできません。

開示

The Futurum Groupは、本稿で言及されている企業を含めた多くのハイテク企業に調査、分析、助言、コンサルティングを提供しています。当社に、本文書で引用されている企業の株式を保有する従業員はおりません。



LENOVOおよびNVIDIAについて

Lenovoは、AIを活用したイノベーションの新時代をすべての人に提供します。当社のフルスタック製品ポートフォリオは、パワフルかつ柔軟性と信頼性のあるAIソリューションを提供し、業界を変革して個人の能力を高めます。すべての人に、よりスマートなAIの未来を創造します。Lenovoでは、AIの未来にはパブリックAIとエンタープライズAIの共存が不可欠であると信じています。Lenovoは、お客様とお客様のデータにAIをもたらします。

NVIDIAとのパートナーシップにより、ハイブリッドAIソリューションはエンジニアリング提携を通じて、ユーザーが最も必要とする場所とタイミングでAIを効率的に顧客データにもたらすことを目的に構築され、すべての人にAIを提供するというLenovoのビジョンを推進し、次世代の生成AIのための画期的な技術とアーキテクチャの市場投入までの時間をサポートします。Lenovoのハイブリッドソリューションは、セキュアでサポートされた安定した本番環境AIを実現するために、NVIDIA AI Enterpriseソフトウェア向けにすでに最適化されており、開発者にNVIDIA NIMsおよびNeMo Retrieverを含めたNVIDIAマイクロサービスへのアクセスを提供します。



THE FUTURUM GROUPについて

[TheFuturum Group](#)は独立した調査、分析およびアドバイザリー企業であり、デジタル革新と市場を揺さぶる存在となるテクノロジーやトレンドに焦点を当てています。当社のアナリスト、調査員およびアドバイザーは日々、世界中のビジネスリーダーが業界の構造的変動を予測し、破壊的イノベーションを活用して市場における競争優位性を獲得または維持できるように支援しています。



お問い合わせ先

The Futurum Group LLC | futurumgroup.com | (833) 722-5337 |