

# 製品概要

最適化された AI クラウドサービス

## インテル® デベロッパー・クラウドで AI を高速化およびスケーリング



### 最新のインテルのハードウェアとソフトウェアにアクセスして AI を高速化

AI を活用して、最高のパフォーマンス、スケーリング、投資収益率 (ROI) を達成するための適切なハードウェアとソフトウェアを見つけることは大きな課題です。これらすべてに加えて、柔軟性、選択肢、エンドツーエンドのワークフローの最適化も提供するのが最高のソリューションでしょう。また、将来新しいアーキテクチャーに移行する際に、投資の価値が引き続き提供されるという保証も必要でしょう。

**インテル® デベロッパー・クラウド**は、開発者や企業が AI を高速化できる、最新のインテルのハードウェアとソフトウェアを備えた開発環境を提供します。クラウド、エンタープライズ、クライアント、エッジにデプロイする、AI、マシニング、HPC、セキュリティー・アプリケーションを開発してテストすることができます。無料の Standard サービス、有償の Premium サービスおよび Enterprise サービスがあります。

インテル® デベロッパー・クラウドは、オープン・ソフトウェアと最先端のソフトウェア・ツールとともに、AI、GPU、その他のアクセラレーター向けに構築された最先端のハードウェア上に構築されています。開発者や共同研究者は、製品が発売される数カ月から 1 年前に、現在、新規、将来のハードウェア・プラットフォームに早い段階で効率良くアクセスすることができます。企業は、インテルのさまざまなテクノロジーに基づいて構築された新機能やパフォーマンスに合わせて、自社の製品やソリューションをいち早く最適化し、迅速に市場に投入することができます。

### AI の価値を高める

現在および次世代のハードウェア・システムから AI 計算の価値を最大限に引き出します。ソフトウェアの最適化によりパフォーマンスと生産性を向上させます。

- **アクセラレーターでマルチプラットフォームの利点を得る**—AI に最適な第 4 世代インテル® Xeon® スケーラブル・プロセッサ、メモリー帯域幅の影響を受けやすいワークロードに最適なインテル® Xeon® CPU マックス・シリーズ、ハイパフォーマンス、高効率なトレーニングと推論が可能な、大規模な LLM および生成 AI に最適なインテル® データセンター GPU マックス・シリーズおよびフレックス・シリーズとインテル® Gaudi®2 AI アクセラレーター上でアプリケーションとソリューションを構築してテストできます。
- **AI アクセラレーション**—小規模および大規模な AI トレーニング、モデルの最適化、推論ワークロードを実行して、最適化されたコストとパフォーマンスでデプロイします。インテル® oneAPI ベース・ツールキット、OpenVINO™ ツールキット、最適化されたディープラーニング・フレームワーク、その他の AI、HPC、レンダリング・ツールを使用して、インテル® アーキテクチャー独自の高度な AI 機能を有効にします。
- **スケーラブルで柔軟**—強力なマルチプラットフォーム・システムとクラスターを活用します。ユーザーのさまざまな柔軟性と計算ニーズを満たすため、3 つの異なるサービスとサポートを提供しています。開発者向けの無料サービス (Standard)、チーム向けの有償サービス (Premium または Enterprise) から選択できます。インテル® デベロッパー・クラウドは、オープン・ソフトウェアの oneAPI マルチアーキテクチャー、マルチベンダー・プログラミングに基づいており、特定のハードウェアや独自のプログラミング・モデルに依存することなく、アクセラレーテッド・コンピューティング、コードの再利用と移植性を提供します。
- **市場投入までの時間の短縮**—量産開始前のハードウェアにアクセスして、新機能とパフォーマンスを備えた AI 製品を最適化し、迅速に市場に投入できます。使いやすい UI、最新のインターフェイス、合理化されたワークフローにより、エンドツーエンドの AI パイプラインを最適化できます。迅速かつ簡単に開始できます。ハードウェアのインストールや取得、ソフトウェアのダウンロード、構成のセットアップは必要ありません。クラウドクレジットを利用して開始することもできます。

#### エンタープライズ AI をより高い ROI で

インテル® デベロッパー・クラウドと [cnvrg.io](https://cnvrg.io) を使用して、管理された、費用対効果の高いインフラ上に、あらゆる規模の AI を構築してデプロイ。

**intel**  
Developer Cloud

インテル® デベロッパー・クラウドは [cnvrg.io](https://cnvrg.io) メタクラウド・プラットフォーム経由でも利用できます。

**cnvrg.io**

## AI のテスト、開発、デプロイをより簡単に —主な使用方法

ユーザーや開発者は、インテルの最新のプラットフォームでテストして、特定のニーズに最適なアクセラレーターを決定できます。開発者、ISV、学術研究者、大企業など、10,000 人以上のユーザーが、すでにインテル® デベロッパー・クラウドを使用しています。

**開発者:** インテルのハードウェア上で高速化された、インテルにより最適化された AI ソフトウェアにアクセスして使用するための簡単なパスを提供します。一般的な用途には、アーキテクチャーの評価、アプリケーションの開発と最適化、モデルとワークロードの最適化、研究と学術学習、oneAPI と LLM/MLOps 向けの教育/トレーニングに加えて、AI 開発と設計認証が含まれます。LLM ワークロードがインテル® アーキテクチャー上でどのように動作するか、サンプルコードで確認できます。

クラウドには、生成 AI のモデル、トレーニングのエッセンシャル・モジュール、開発を支援するソフトウェア・ツールが事前にロードされた Jupyter\*、Notebook が含まれています。Visual Studio\* Code もサポートします。

**企業およびエンタープライズ:** システム・インテグレーター (SI)、独立系ソフトウェア・ベンダー (ISV)、サードパーティー SaaS 組織の用途には、大規模な AI トレーニングと推論プロダクション・ワークロードの実行とテスト、認証とベンチマーク、サードパーティー AI SaaS 向けのコンピューティング・サービスが含まれます。

**AI 計算:** 小規模および大規模な AI トレーニング (LLM または生成 AI)、モデルの最適化、推論ワークロードを実行します。インテルの GPU、CPU、インテル® Gaudi®2 AI アクセラレーターを搭載したシステム上で、小規模から大規模な仮想マシン (VM)、フルシステムまたはクラスターを利用します。パラメーター数が 70 億から数千億までのモデルを利用できます。

次の表は、インテル® デベロッパー・クラウドで提供されるさまざまなサービスとサポートオプションの概要です。

インテル® デベロッパー・クラウド サービスとサポートオプション			
	Standard	Premium	Enterprise
対象ユーザー	AI および HPC 開発者、研究者、データサイエンティスト、アカデミア 1 人のユーザーがアクセス	エンタープライズ・ユーザー/開発者 1 人のユーザーがアクセス	エンタープライズ・ユーザー 複数のユーザーがアクセス
使用方法	<ul style="list-style-type: none"> <li>評価—さまざまなアーキテクチャーで、アプリケーション、ワークロード、LLM ワークロードのサンプルコードをテスト</li> <li>AI および HPC アプリケーションを開発して、新機能と最高のパフォーマンスのために最適化</li> <li>GPU アクセスのスケジュール</li> <li>トレーニングと開発、AI 開発と設計向けのインテルにより認定された開発者認証評価の取得</li> </ul>	<ul style="list-style-type: none"> <li>Standard サービスの使用方法 +</li> <li>AI トレーニングと推論のプロダクション・ワークロード</li> <li>モデルの最適化とデプロイ</li> <li>認証、ソフトウェア検証、ベンチマーク・テスト</li> </ul>	<ul style="list-style-type: none"> <li>Premium サービス +</li> <li>サードパーティー AI SaaS プロバイダー向けのハイパフォーマンスでコストが最適化されたインテルのコンピューティング・サービス</li> </ul>
ハードウェアアクセス	<ul style="list-style-type: none"> <li>第 4 世代インテル® Xeon® スケーラブル・プロセッサ - ベアメタルおよび仮想マシン (VM)</li> <li>インテル® Xeon® CPU マックス・シリーズ - ベアメタル</li> <li>インテル® データセンター GPU マックス・シリーズおよびインテル® データセンター GPU フレックス・シリーズ - ベアメタル</li> <li>インテル® Gaudi®2 AI アクセラレーター (ディープラーニング・タスク向け) - ベアメタルアクセス (事前に承諾された一部のユーザー)</li> </ul>	Standard サービスのアクセス + <ul style="list-style-type: none"> <li>第 5 世代インテル® Xeon® スケーラブル・プロセッサ、その他のインテル® Xeon® スケーラブル・プロセッサ、インテル® データセンター GPU マックス・シリーズ、インテル® データセンター GPU フレックス・シリーズ - シングルノードのシステムとクラスターへのベアメタルアクセス</li> <li>k8s クラスターへのアクセス</li> </ul>	Premium サービスと同じ
テクニカルサポート	コミュニティ・フォーラムによるサポート (SLA なし)	インテルのテクニカルエンジニアによるサポート 平日午前 8 時から午後 5 時 (ユーザーの地域ごと、制限あり)、1 営業日 SLA	インテルのテクニカルエンジニアによる優先サポート (電話、チャット、ヘルプ・リクエスト・チケット) 1 時間から 1 営業日 SLA、24 時間
費用	無料 + クラウドクレジットで利用可能 (ベアメタルサービスなどの特定のインスタンス・タイプ) 従量課金へのアップグレード・オプション	クラウドクレジットで利用可能 + インテル® デベロッパー・クラウド・ポータルに記載の 1 時間当たりの料金に基づく  長期契約/予約価格割引があります。詳細は、インテルの担当者までお問い合わせください。	クラウドクレジットで利用可能 + インテル® デベロッパー・クラウド・ポータルに記載の月額サブスクリプション料金に基づく  特別価格および長期契約/予約価格割引があります。詳細は、インテルの担当者までお問い合わせください。

## ユーザーにもたらされる価値とメリット

インテル® データセンター GPU マックス・シリーズ上で LoRA を使用して予測、電子商取引、パーソナル・アシスタント向けに LLM をファインチューニング—Moonshot AI は、LLM を活用して予測を行うという新しい分野を開拓しています。SiteMana は、LLM を利用して電子商取引のマーケティングを自動化しています。Selecton Technologies は、LLM を利用してゲーマー向けの AI パーソナル・アシスタントを開発することによりニッチな市場を開拓しています。[詳細](#) (英語)

エンタープライズ向け LLM のリスク回避—Prediction Guard のプラットフォームを使用すると、企業はプライバシーやセキュリティを犠牲にすることなく、ChatGPT で使用されているような AI モデルの最新のトレンドを導入できます。[詳細](#) (英語)

電子メール生成の革命—SiteMana は、高度なマシンラーニングを用いて購入意欲が高い訪問者を特定し、パーソナライズされたメッセージでリターゲティングすることにより、匿名トラフィックに関与する方法を企業に提供します。このアプローチは安全かつ合法であり、個人のプライバシーを尊重し、個人の身元を明らかにすることはありません。[詳細](#) (英語)

粒子ベースのシミュレーション・ソフトウェアを使用して、エンジニアが優れた製品を開発できるように支援—「インテル® HPC ツールキットのコンポーネントを使用して、インテル® Xeon® プロセッサ上でコスト効率に優れた方法でソルバーのパフォーマンスを最大限に引き出すことができました。現在、CPU および GPU アーキテクチャーでワークロードを実行できるようにする準備を進めています。」

— Dive Solutions GmbH 最高技術責任者 (CTO)  
Johannes Gutekunst 氏

## 実際に使ってみる

インテル® デベロッパー・クラウド・アカウントのセットアップは簡単です。AI LLM/MLOps および oneAPI のトレーニングが迅速に提供されます。[今すぐ始めましょう。](#)

## インテル® デベロッパー・クラウドで利用可能なテクノロジー

<p><b>ハードウェア・カタログ:</b> CPU</p>	<ul style="list-style-type: none"> <li>● <b>第 4 世代インテル® Xeon® スケーラブル・プロセッサ</b>。高度な AI アクセラレーターと次の機能を含みます。                     <ul style="list-style-type: none"> <li>– インテル® アドバンスド・マトリクス・エクステンション (インテル® AMX)</li> <li>– インテル® アドバンスド・ベクトル・エクステンション 512 (インテル® AVX-512)</li> <li>– インテル® クイックアシスト・テクノロジー (インテル® QAT)</li> <li>– インテル® データ・ストリーミング・アクセラレーター (インテル® DSA)</li> <li>– インテル® In-Memory Analytics Accelerator (インテル® IAA)</li> <li>– VNNI/bfloat16</li> </ul> </li> <li>● <b>第 5 世代インテル® Xeon® スケーラブル・プロセッサ</b> (一部の Premium および Enterprise ユーザー)</li> <li>● <b>インテル® Xeon® CPU マックス・シリーズ</b>。高帯域幅メモリー (HBM) を搭載した唯一の x86 ベースのプロセッサで、インテル® AMX、インテル® AVX-512、インテル® ディープラーニング・ブースト (VNNI, bfloat16)、インテル® DSA を含みます。</li> </ul>	<ul style="list-style-type: none"> <li>● 仮想マシン (VM) およびベアメタルアクセス (2 ソケット、256GB メモリー、2TB ディスク)。</li> <li>● ベアメタルアクセス - DDR5 (8 チャンネル)、4,800MT/S (1DPC)、PCIe* 5 (80 レーン)、64GB HBM2e、最大 56 コア、1TB/S メモリー帯域幅。HBM 専用モード、フラットモードまたはキャッシュモード。</li> </ul>
<p><b>GPU</b></p>	<ul style="list-style-type: none"> <li>● <b>インテル® データセンター GPU マックス・シリーズ 1100 および 1550</b>。ベクトルエンジンとマトリクスエンジン (インテル® X<sup>e</sup> マトリクス・エクステンション (インテル® XMXX)) を搭載し、AI および HPC 向けの重要なデータ形式に対応する強力な Op/CLK を処理可能な X<sup>e</sup> コアによる革新的な機能が含まれています。SIMT モデルと SIMD モデル、インテル® X<sup>e</sup> リンク、データ型の柔軟性、レイトレース・ハードウェア・アクセラレーションなどをサポートします。</li> <li>● <b>インテル® データセンター GPU フレックス・シリーズ</b>。メディア・ストリーミング、AI 視覚推論、クラウドゲーム、仮想デスクトップ・インフラストラクチャー (VDI)、仮想化とデジタルコンテンツ作成をサポートします。組み込みのインテル® XMXX AI アクセラレーション、AV1 ハードウェア・エンコードとデコード、レイトレース・ハードウェア・アクセラレーションにより、さまざまなレイトレーシング、シミュレーション、画像補正ワークロードを高速化します。</li> </ul>	<ul style="list-style-type: none"> <li>● ベアメタルアクセス - AI および ML トレーニング向けのバッチサービスを使用。最大 128 基のレイトレーシング・ユニット、128 個の X<sup>e</sup> コア、最大 64MB の L1 キャッシュ、408MB L2 キャッシュ、最大 128GB の HBM2e。</li> <li>● ベアメタルアクセス - 最大 32 個の X<sup>e</sup> コアとレイトレーシング・ユニット、最大 4 個の X<sup>e</sup> メディアエンジン。</li> </ul>
<p><b>AI アクセラレーター</b></p>	<p><b>インテル® Gaudi®2 AI アクセラレーター</b> (英語)。パフォーマンスとコスト効率に優れた、LLM のディープラーニングのトレーニングと推論、生成 AI に最適な AI アクセラレーターです。Gaudi®2 向けに最適化されたソフトウェアにより、小規模なコンピューター・ビジョンおよび NLP モデルから数十億のパラメーター・モデルの効率的な処理まで、最先端のモデルに簡単にアクセスできます。効率的なスケーラビリティを実現するように設計された Gaudi®2 アクセラレーターは、費用対効果が高く、困難なワークロードで優れたパフォーマンスが必要なユーザーに最適です。</p>	<p>ベアメタルアクセス - 一部の Premium および Enterprise ユーザー。</p>

## インテル® デベロッパー・クラウドで利用可能なテクノロジー

<b>ソフトウェア・カタログ:</b> <b>最適化されたソフトウェアとツール</b>	<ul style="list-style-type: none"> <li>● <b>インテル® oneAPI ベース・ツールキット</b> <ul style="list-style-type: none"> <li>– インテル® oneAPI DPC++/C++ コンパイラー</li> <li>– ディープ・ニューラル・ネットワーク、数学、データ・アナリティクス、DPC++、スレッディング、集合通信、画像、信号処理、データ圧縮、暗号化向けライブラリー。</li> <li>– 解析、デバッグ、コード移行ツール—インテル® VTune™ プロファイラー、インテル® Advisor、インテル® ディストリビューションの GDB、インテル® DPC++ 互換性ツール</li> </ul> </li> <li>● <b>インテルの AI ツールとインテルにより最適化されたフレームワーク</b> (英語) <ul style="list-style-type: none"> <li>– インテルにより最適化された TensorFlow* および PyTorch* 向けディープラーニング・フレームワーク</li> <li>– インテル® ディストリビューションの OpenVINO™ ツールキット</li> <li>– インテル® ニューラル・コンプレッサー</li> <li>– scikit-learn* 向けインテル® エクステンション、XGBoost 向けインテル® オプティマイゼーション、Modin 向けインテル® ディストリビューション、インテル® ディストリビューションの Python*</li> <li>– インテル® アーキテクチャー向け Model Zoo</li> </ul> </li> <li>● <b>インテル® HPC ツールキット</b>: インテル® oneAPI DPC++/C++ コンパイラー、インテル® Fortran コンパイラー、インテル® Inspector、インテル® MPI ライブラリー、インテル® Trace Analyzer &amp; Collector</li> <li>● <b>インテル® レンダリング・ツールキット</b>: インテル® Embree、インテル® オープン・イメージ・デノイズ、インテル® オープン・ポリウム・カーネル・ライブラリー、インテル® オープン・バス・ガイディング・ライブラリー、インテル® インプリシット SPMD プログラム・コンパイラー、インテル® OSPRay</li> <li>● <b>インテル® Quantum SDK</b> (英語)</li> <li>● <b>Jupyter* Notebook</b>: 使い慣れた Jupyter* Notebook で開発を始めます。生成 AI のモデルを実行し、インテルの最新の CPU と GPU で Python* コードをインラインで作成して実行できます。</li> </ul>
<b>AI サービス</b>	<p>オープンソースの基本的な AI モデルを実行します。次のモデルを含みます。</p> <ul style="list-style-type: none"> <li>● Technology Innovation Institute* (TII) Falcon LLM</li> <li>● MosaicML* MPT</li> <li>● Hugging Face* BigScience Large Open-science Open-access Multilingual Language Model (BLOOM)</li> <li>● Stability.AI* Stable Diffusion*</li> <li>● Meta AI* Llama 2</li> <li>● Databricks* Dolly</li> </ul>

### 関連情報

- [インテル® デベロッパー・クラウド](#)
- [AI ソフトウェア・ソリューション](#)
- [インテルの AI ツール、ライブラリーとフレームワークの最適化](#) (英語)
- [インテルが提供するエンタープライズ・ソフトウェア・ソリューション](#)
- [インテル® Liftoff for Startups](#)



#### 法務上の注意書き

性能は、使用状況、構成、その他の要因によって異なります。詳細については、<http://www.intel.com/PerformanceIndex/> (英語) を参照してください。結果は異なることがあります。

性能の測定結果はシステム構成の日付時点のテストに基づいています。また、現在公開中のすべてのセキュリティー・アップデートが適用されているとは限りません。絶対的なセキュリティーを提供できる製品またはコンポーネントはありません。

実際の費用と結果は異なる場合があります。インテルのテクノロジーを使用するには、対応したハードウェア、ソフトウェア、またはサービスの有効化が必要となる場合があります。

インテルは、サードパーティーのデータについて管理や監査を行っていません。ほかの情報も参考にして、正確かどうかを評価してください。

© Intel Corporation. Intel, インテル, Intel ロゴ, その他のインテルの名称やロゴは、Intel Corporation またはその子会社の商標です。

\* その他の社名、製品名などは、一般に各社の表示、商標または登録商標です。