



Technology White Paper

現場適用を加速する アラヤのエッジ AI ソリューション



Contents

はじめに	2
産業別のアプリケーション・ユースケース例	22
産業別ユースケース	3
エッジ AI のメリット：クラウドからエッジ AI へ	44
エッジ AI のメリット	55
クラウドとエッジコンピューティングの連携	66
エッジ AI プラットフォーム	66
ケーススタディ	88
事例 1：エッジ AI 搭載ドローンによるリアルタイムな鉄塔さび検知	8
事例 2：自動車メーカー向けエッジ AI 開発効率化への取り組み	8
エッジ AI 導入における課題	11
レイテンシ	11
消費電力	11
結論	12
アラヤについて	13
お問合せ	13

はじめに

現代において、人工知能（AI）の台頭は、産業界や私たちの日常生活に革命をもたらし、前例のない機会と能力を提供しています。クラウドベースの処理とクラウド AI の普及に伴い、遅延、帯域幅の制限、セキュリティおよびプライバシーの問題が顕在化する中、これに対応するために、現場でローカルに AI 処理を行うエッジ AI の重要性が高まっています。応答性の向上、プライバシーとセキュリティの強化、コスト効率、スケーラビリティ、分散型 AI 処理など、エッジ AI は私たちの世界にさらなる革命をもたらす要素を備え、多様な技術進歩が急速に引き起こされています。一方で、このようなエッジ AI の急速な技術進歩を、日々の現場業務を抱えるエンジニアや責任者がキャッチアップするには高い障壁があります。

本レポートは、エッジ AI の現状を包括的に理解するための実用的な技術的知識を提供するとともに、アラヤが提供するエッジ AI ソリューションを活用することで、エッジ AI の現場導入を短期間で実現できるイメージを提供します。

産業別のアプリケーション・ユースケース例

集中型クラウドデータセンターから、小規模なローカルデータセンター、データソースの近くに存在するエッジデバイスへのデータ処理の移行が進行中です。この流れが、エッジ AI の出現と台頭に寄与しています。現在、多くの企業が AI アプリケーションにエッジ処理機能を組み込み、活用しています。これらの機能により、エネルギー効率に優れ、低レイテンシでリアルタイムなパフォーマンスを発揮するアプリケーションを開発することができます。さらに、エッジ AI の機能により、企業はセキュリティとデータ保護に大きなメリットを得ることができ、プライバシー管理の向上や適用される規制への効果的な準拠につながります。このようなメリットにより、エッジ AI はさまざまな分野の企業にとって非常に魅力的な存在となっており、幅広いユースケースでエッジ AI コンピューティング機能の導入・活用が進んでいます。

Fortune Business Insights によると、エッジ AI 市場規模は 2021 年に 119 億 8,000 万米ドルと評価され、2029 年までに 1,074 億 7,000 万米ドルに達し、予測期間中に 31.7% の CAGR を示す、と予想されています。 <https://www.fortunebusinessinsights.com/jp/エッジ ai 市場-107023>

産業別ユースケース

製造業

製造業におけるエッジ AI の応用は、その革新的な機能によって特に顕著です。工場の自動化、製品の品質管理、供給チェーンの最適化など、様々なプロセスがエッジ AI により変革を遂げています。例えば、製品の品質検査においては、ディープニューラルネットワークを用いた画像解析により、製品の欠陥をリアルタイムで検出し、品質の一貫性を保つことが可能です。また、産業用ロボットの動作制御にエッジ AI を適用することで、低遅延かつ高精度な操作が可能となり、生産効率を大幅に向上させています。3D プリンティングなどの新しい製造技術では、エッジ AI を活用することで機密データを保護しつつ、プロセス効率を向上させています。

ヘルスケア

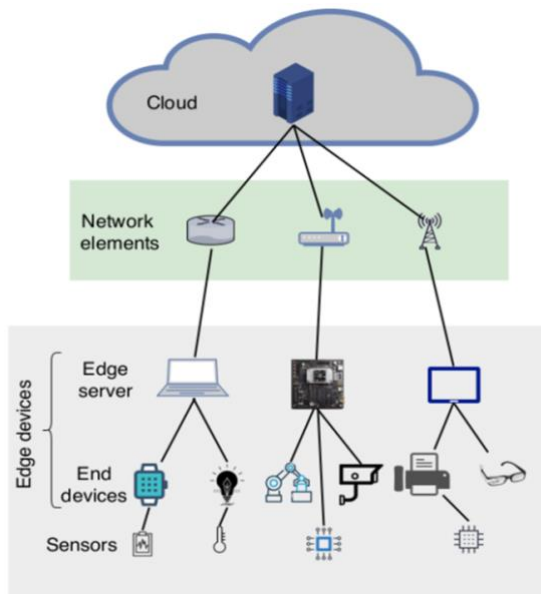
ヘルスケア分野では、エッジ AI は患者のケア品質の向上と医療サービスの効率化に貢献しています。リアルタイムの患者監視システムでは、ウェアラブルデバイスや医療機器が生成する大量のデータを即座に解析し、異常の兆候を早期に検出することが可能です。これにより、緊急時の迅速な対応が可能になり、患者の命を救うことができます。また、遠隔医療においても、エッジ AI は医師が患者の状態を効率的に監視し、迅速な治療を行うための重要な役割を担っています。手術室では、エッジ AI を組み込んだ医療機器が外科医にリアルタイムで正確な情報を提供し、より安全で効果的な手術を支援しています。

スマートシティ

スマートシティでは、エッジ AI の活用が都市運営の効率化と市民生活の質の向上に大きく寄与しています。交通管理システムでは、リアルタイムのデータ解析により交通の流れを最適化し、渋滞の軽減や事故の予防に貢献しています。また、公共の安全を確保するために、監視カメラに組み込まれたエッジ AI が不審行動や緊急事態を迅速に検出し、適切な対応を促します。さらに、エッジ AI を活用したスマートメーターシステムは、エネルギー使用量の効率的な管理や水漏れの早期発見を可能にし、持続可能な都市運営に貢献しています。これらのシステムは、都市全体の運営をよりスマートで効率的なものにし、市民の生活をより快適かつ安全にしています。

エッジ AI のメリット : クラウドからエッジ AI へ

クラウドコンピューティングの導入後、クラウド上でデータを管理・分析することへの関心が急速に高まりました。当時、多くの企業は、クラウドをスケーラビリティ、容量、サービス品質から、データ分析や機械学習機能を実行するための非常に魅力的な選択肢だと考えていました。しかし、時間が経つにつれて、多くの企業がクラウドへ大量のデータを転送・処理することに対して懸念を抱き始めました。特に、機密性の高い企業データを自社で管理していない遠隔地のデータセンターに転送することに抵抗を感じる企業が多くありました。また、クラウドのデータ管理や分析には広域ネットワークを介したデータ転送が必要なため、これがいくつかのアプリケーションにおける遅延の原因となっているという懸念もありました。上記のようなクラウドデータ処理の限界から、多くの企業はデータソースの近くでデータの一部を処理するエッジコンピューティングに注目しています。エッジコンピューティングは、データをクラウドに転送する前に処理することで、処理と分析のためにクラウドに転送するデータ量を削減します。そのため、クラウドでのデータ分析における待ち時間、セキュリティ、プライバシーの課題が軽減されます。



出展: “Machine Learning at the Network Edge: A Survey”, M.G. Sarwar Murshed , 2019

https://www.researchgate.net/profile/M-G-Sarwar-Murshed/publication/334866807_Machine_Learning_at_the_Network_Edge_A_Survey/links/5fe9319392851c13fec9a29a/Machine-Learning-at-the-Network-Edge-A-Survey.pdf

エッジ AI のメリット

クラウドコンピューティングの普及に伴い、データのクラウド上での管理と分析への関心が高まりましたが、データの安全性、プライバシー、遅延などの課題も明らかになりました。これらの課題に対応するため、エッジ AI コンピューティングが注目、クラウドではなく、データソースに近い場所での AI 処理がこれらの課題を解決しつつあります。

エッジ AI の具体的なメリットを以下に示します：

低遅延とリアルタイム性能

エッジ AI はデータソースの直近でデータを処理するため、通信遅延が極めて低くなります。これにより、リアルタイムでの応答が可能になり、応答性や処理時間を重視したアプリケーションの価値を高めます。製造ラインの不良品検出、セキュリティシステム、緊急対応システムなどが、この低遅延のメリットを享受しています。

セキュリティとプライバシー管理の強化

データをエッジで処理することにより、エッジ AI はセキュリティとプライバシーのリスクを減らします。特に、機密性が高いデータをクラウドではなくローカルで処理することにより、外部へのデータ露出が最小限に抑えられ、セキュリティ侵害の可能性が大幅に低減されます。

省エネルギー性とコスト効率の向上

エッジ AI の導入により、クラウドへのデータ転送と処理が削減され、エネルギー消費と CO2 フットプリントの低減に寄与します。これは、電力資源が限られている場所やリモート地域で特に重要なメリットです。

デバイス上の学習と新機能の実現

エッジ AI はクラウドに依存しないため、デバイス上で直接学習を可能にし、新たなアプリケーションの開発を促進します。これにより、従来のクラウド AI アプリケーションを超える、新しいタイプのアプリケーションが生まれます。

エッジ AI のメリットは、高速な処理能力、セキュリティの強化、エネルギー効率の向上、そしてデバイス上での直接学習により、多くの業界に革新をもたらしています。特に、エッジ AI 処理の低遅延性は、リアルタイムのデータ分析や迅速な意思決定が可能にしています。

クラウドとエッジコンピューティングの連携

エッジ AI は多くの点でクラウド AI を上回りますが、データ集約型のアプリケーションではクラウドのスケラビリティが重要です。たとえば、大量のデータポイントを扱うディープラーニングアプリケーションでは、エッジとクラウドの組み合わせが効果的です。エッジで初期データ処理を行い、クラウドで必要なデータセットのみを処理することで、エッジ AI とクラウド AI のメリットを最大化します。このハイブリッドアプローチでは、ローカルデータ処理後に選択したデータをクラウドに転送し、異なるエッジソースからのデータを統合してグローバルな分析を行うことができます。

更に、認証機能を用いた、フェデレーション（認証連携）型の機械学習システムを導入することで、異なるエッジノードで機械学習モデルを訓練し、個々のモデルから得られる知見をクラウドで集約する際に、データのプライバシーを保護しつつ、効率的なグローバルモデルを構築できます。このように、クラウド/エッジ AI の統合によるメリットは、エッジ AI システムのさらなる拡張性と効率性をもたらします。

エッジ AI プラットフォーム

エッジ AI プラットフォームを選択し、利用する際には、プロジェクトの要件に合わせて、ソフトウェア・プラットフォーム、ハードウェア・プラットフォーム両方の視点から、レイテンシ、データ処理能力、消費電力、サイズ、重量、放熱などを評価することが重要です。

例えば、プラットフォームの評価を以下のような視点で行っています。

- プラットフォームが要求に対して十分な処理能力と低レイテンシを備えているか
- 実行環境で消費電力の要求に対して十分なエネルギー効率の高さを備えているか
- プラットフォームの操作に必要な専門知識のレベル

- プラットフォームのコストとして、特定機能に追加費用などがかかるか
- プラットフォームが提供するサポートのレベルやエコシステムの互換性はあるか

特に、強力なサポートシステムと豊かなエコシステムを持つプラットフォームは、開発者に貴重なリソースやドキュメント、コミュニティサポートを提供し、より効果的な利用を可能にします。

多様なニーズに対応し、エッジデバイス上で機械学習モデルを展開できる代表的なプラットフォームとして、PyTorch Mobile、OpenVINO、NVIDIA Jetson、等があります。以下では、これらのソフトウェアおよびハードウェアのプラットフォームについて概要を説明します。

ソフトウェア・プラットフォーム

TensorFlow Lite

TensorFlow Lite は、エッジデバイス上で機械学習モデルをデプロイするために設計されています。低レイテンシ、クロスプラットフォーム互換性、ユーザーフレンドリーな機能が特徴です。TensorFlow Lite は、モバイルプラットフォームやマイクロコントローラプラットフォームなど、さまざまなデバイスで優れたパフォーマンスを発揮します。

PyTorch Mobile

AI モデルの研究開発に広く使用されている PyTorch フレームワークの拡張である PyTorch Mobile は、その使いやすさと柔軟性をモバイルデバイスにもたらしめます。スピードと効率のために最適化され、さまざまなデータ処理機能をサポートしているため、リソースに制約のあるデバイスに適しています。

OpenVINO

インテルが開発した OpenVINO は、ディープラーニングモデルのデプロイに俊敏性と優れたパフォーマンスを提供します。CPU、GPU、FPGA など様々なハードウェア・プラットフォームをサポートしており、リアルタイムアプリケーションに適しています。

ハードウェア・プラットフォーム

NVIDIA Jetson

高い処理能力と低レイテンシで知られる NVIDIA Jetson は、エッジ向けの GPU を活用した幅広いニューラルネットワークと開発ツールをサポートしています。ビデオ解析や自律ロボット工学など、リアルタイムのデータ処理を必要とするアプリケーションに特に適しています。

Hailo

Hailo シリーズのプロセッサは、ディープラーニングのタスクにおいて高効率なパフォーマンスを発揮するように設計されており、低電力で迅速かつ信頼性の高い処理を必要とするアプリケーションに最適です。

Ambarella

画像処理に特化したアンバレラの CV シリーズプロセッサは、集中的なディープラーニング・タスクにおいて効率的です。特に大量のデータを迅速に処理することに長けており、高度な監視システムや IoT 機器に適しています。

NXP

NXP i.MX シリーズはエッジ AI 処理性能に優れており、一部のプロセッサは NPU を搭載しています。AI 主導のタスクで効率的なパフォーマンスを発揮するように設計されているため、データ処理の速度と精度が重要な幅広いアプリケーションに適しています。

ケーススタディ

この章では、エッジ AI の効果的な活用例を具体的に紹介します。

事例 1 : エッジ AI 搭載ドローンによるリアルタイムな鉄塔さび検知

https://www.nedo.go.jp/news/press/AA5_101616.html

アラヤでは、NEDO（国立研究開発法人新エネルギー・産業技術総合開発機構）の「高効率・高速処理を可能とする AI チップ・次世代コンピューティングの技術開発」の一環として、KDDI 株式会社（現 KDDI スマートドローン株式会社）様、と共同で、端末（エッジ）で動作する軽量

な人工知能（AI）モジュール（以下、AI エッジモジュール）を開発し、これを実装したドローンで通信鉄塔を点検する実証を行いました。



本実証では、エッジ AI モジュールには、NVIDIA社のJetson、Hailo社のAIチップ、Intel社のFPGAを採用し、カメラ撮影動画をリアルタイムにAI処理可能な推論パイプラインを構築し、今回開発したエッジ向け錆検知のAIモデルをエッジAIモジュールに組み込みました。

本実証では、ドローンが自律飛行して通信鉄塔を撮影し、映像データからエッジAIモジュールがリアルタイムにさびの検知を行うと共に、4Gや5G通信を用いて、外部記憶装置やPCを経由することなくワンストップでエッジからクラウドへ検知データを自動アップロードできることを確認しました。

事例2：自動車メーカー向けエッジAI開発効率化への取り組み

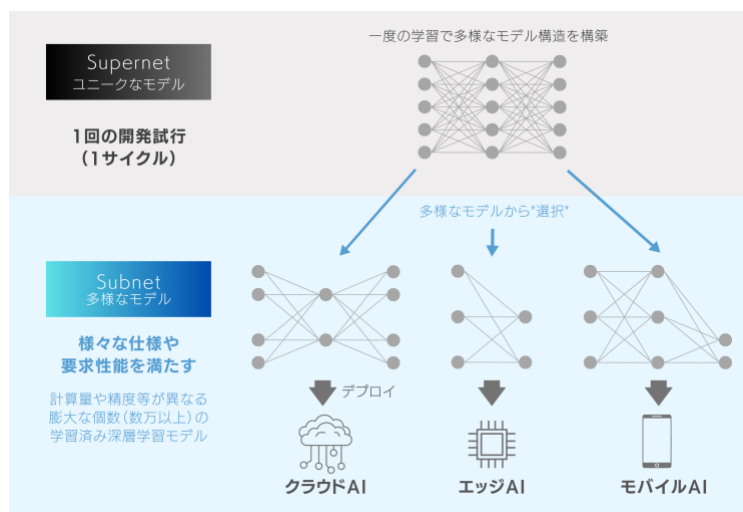
<https://www.araya.org/projects/subnetx/>

ある自動車メーカーでは、AIモデルを運用システムにデプロイする際の最適化が大きな課題でした。AIモデルは、機能要求をもとに先行開発し、運用システムのデバイス仕様に応じた機能要件に合わせて、最適化を行う必要があるためです。しかし、このようなAIモデル最適化の開

発において、スキルを備えた要員を確保して必要な開発時期にプロジェクトへアサインすることは容易ではなく、開発における大きな課題でした。

このような課題に対して、AIモデル最適化手法として知られるOFAベースの技術を、自動車メーカーで想定する開発プロセスに適用することで、エッジデバイスの開発経験がないAI開発者でも、AIモデル最適化のプロセスを自動化し、様々なデバイス向けに最適化されたAIモデルを効率的に開発できるフロー確立に取り組みました。

各モデルに対するOFAベースの手法適用は、大きく2つのプロセスで行いました。



まず、学習フェーズでは、対象となるモデルのパラメータ（カーネルサイズ、深さ、チャンネル幅）について、最も冗長なパラメータを冗長表現したモデル（Supernet）を準備してモデル全体を学習します。

その後、モデル全体のカーネルサイズ、深さ、チャンネル幅に加え解像度を微調整したパラメータを備えるモデルを段階的に学習します。これにより、さまざまなパラメータや解像度に柔軟に対応した多数（例えば $\sim 10^{19}$ 個）のモデル（Subnet）を効率的に学習できます。

次に、デプロイフェーズでは、学習フェーズで得られた多数のモデルの性能評価を行うための精度予測器と速度予測器とを作成します。精度予測器は、学習フェーズで用いたデータセットを使用して非常に低コストで準備できます。また、速度予測器は、単純にはモデルのパラメータ数から算出したMAC数（積和演算数）を利用します。また、デプロイ対象となるデバイス上での実行速度の見積値や実測値を用いることもできます。

このようにして作成した、精度予測器と遅延予測器から得られる精度と速度との情報を用いることで、デプロイ対象のデバイスに最適となる性能（精度・速度）を備えたモデルの取得が可能となりました。

本事例で説明した開発手法は、現在、AI モデル開発効率化サービス「SubnetX」として提供しています（<https://www.araya.org/service/subnetx/>）

エッジ AI 導入における課題

エッジ AI の導入は現代のコンピューティングに多くの利点をもたらす一方で、エッジ AI の実装や利用にはいくつかの課題が存在します。本章ではこれらの代表的な課題を示し、エッジ AI の導入成功への影響を探ります。

レイテンシ

レイテンシはエッジ AI システムのパフォーマンスに大きく影響する重要な要素です。これには、データを取得して AI モデル処理を開始するまでの時間（入力レイテンシ）、AI モデルによるデータ処理時間（処理レイテンシ）、そして AI 予測の生成から結果の伝達までの遅延（出力レイテンシ）が含まれます。入力レイテンシはセンサー反応やデータ伝送の遅れ、前処理のオーバーヘッドが原因であり、処理レイテンシはモデルの複雑さやデバイスの処理能力に依存します。出力レイテンシはネットワーク輻輳や通信プロトコルのオーバーヘッドが原因です。適切なエッジコンピューティングの導入することで、AI 処理をクラウドからエッジに移し、処理の待ち時間を短縮します。

消費電力

エッジ AI システムは高いエネルギー要件に直面することがあり、特に遠隔地や屋外環境では、その採用や影響が制限されることがあります。エネルギー消費の高さは、データをリアルタイムで処理・分析するために必要な高性能コンピューティングリソースに起因します。エッジデバイスは、多くの場合、高性能プロセッサやメモリ、ストレージデバイスを必要とし、エネルギー制約のある環境での電力供給が困難です。これに対処するため、低消費電力のプロセッサやメモリなど、エネルギー効率の高いハードウェアを使用することが重要です。エッジ AI 専用

に設計されたエッジコンピューティングハードウェアの使用は、パフォーマンスを損なわずにエネルギー消費を削減する一つの方法です。

結論 ～エッジ AI 導入実現に向けて～

人工知能の進化の中で、エッジ AI 技術はインテリジェントなデバイスとシステムの将来を形成する前例のない機会を提供しています。ローカルでのデータ処理、待ち時間の短縮、プライバシーの強化、リアルタイムの意思決定を可能にするエッジ AI は、自動運転車、ヘルスケア、産業オートメーション、IoT の展開など、さまざまな業界に革新をもたらし、より繋がりのある、効率的でスマートな世界へと導きます。エッジコンピューティングの実行インフラの拡大と AI アクセラレータの開発により、エッジ AI アプリケーションの潜在能力はさらに広がります。自然言語処理、コンピュータビジョン、ディープラーニングなどの複雑なタスクをエッジで直接実行できるようになると、より洗練された AI モデルがエッジデバイスに導入されることが期待されます。

エッジ AI 導入の成功には、プロジェクトの要件を、上述したレイテンシや消費電力、その他の課題と照らし合わせて慎重に評価することで最適なプラットフォームを選択して、パフォーマンス、専門知識のレベル、サポート、エコシステム、コスト等において達成すべき効果を最大化する必要があります。

アラヤのエッジ AI ソリューションを活用することで、エッジ AI の現場導入を短期間で実現し、エッジ AI の社会実装に貢献することで、革新的なアプリケーションやソリューションを創出し、新たな可能性を切り開いていくものと確信しています。

アラヤのエッジ AI ソリューション： <https://www.araya.org/service/edgeaiconsulting/>

アラヤについて

株式会社アラヤは、認知神経科学の研究者、金井良太、が率いる、人工知能 AI とニューロテックをコア技術とし、「人類の未来を圧倒的に面白く」をミッションに掲げたディープテックベンチャー企業です。製造業、ヘルスケア、建設、アカデミック・リサーチ等の領域において、AI アルゴリズム開発、エッジ AI 実装、生成 AI を活用した先進的なソリューションや DX 支援、また、ニューロテック・ブレインテック領域等における高度な研究開発支援をご提供することで、皆様から信頼されるソリューションパートナーであり続ける事を目指します。

お問い合わせ

株式会社アラヤ

本社所在地

〒101-0025 東京都千代田区神田佐久間町 1-11 産報佐久間ビル 6F

お問い合わせフォーム： <https://www.araya.org/contact/>