# Edge AI Server

ARX EDGE OS

ハードウェアとソフトウェアを統合した
エッジAIサーバー＆オペレーティングシステム




ARDGE

## GPUサーバーを、すぐ使える社内AIサービスへ

ARX OSは、エッジAI GPUサーバーのためのオンプレミスAI管理プラットフォーム。
オフィスや支社、工場、分散拠点へのAIハードウェア導入が進む今、企業が求めているのは演算能力だけではありません。
オンプレミスモデルの迅速な導入、一元管理、サービスとしての提供、さらに最新のAIエコシステムとの継続的な連携を可能にするOSレイヤーが求められています。
ARX OSは、単体のオンプレミスAIハードウェアを、すぐに利用でき、管理・共有可能な企業向けAIサービスへと変えます。AIへの投資を、より迅速に実際の業務成果へとつなげます。

## モデルとアプリケーションの導入を 数時間から最短15分へ

ARX OSを利用することで、企業はGPU環境、推論エンジン、モデルサービスを一から構築する必要がなくなります。また、限られた専門エンジニアへの依存を軽減できます。
企業や開発者は、複雑なインフラの導入や保守ではなく、アプリケーションとビジネスプロセスの革新に集中できます。
ノーコードのグラフィカルインターフェースから、プライベートモデルの導入、APIキーの作成、LLMエンドポイントの公開、接続テストを迅速に実行できます。
さらに、検索拡張生成（RAG）ナレッジベース、NASデータソース、リモート運用機能を統合することで、PoC、システム統合、検証にかかる期間を大幅に短縮します。
これにより、導入・管理・保守が容易なオンプレミス推論環境を構築できます。

## 主なメリット



**小さく始めて、段階的に拡張**
PoCからスタートし、利用状況に応じて部門・拠点単位へ段階的に拡張できます。



**迅速な検証とイノベーション**
複数のオープンソースモデルを自由に切り替えて検証し、用途に最適なモデル構成をより早く見つけられます。




**ノーコードによる導入と公開**
統合されたAPIエンドポイントとAPIキー管理により、アプリケーションとの接続を迅速に確立できます。



**IT部門の運用負荷を軽減**
GPUインフラ、推論ランタイム、モデルサービス、アクセス制御環境の統合に必要な時間と人的リソースを削減します。

ARDGE: www.ardge.com 営業・事業提携に関するお問い合わせ: usasales@ardge.com

# ARX OSのコア機能

複雑なオンプレミスAIの導入・管理・運用を、1つのプラットフォームでシンプルにします。



ARX EDGE OS

### 01 最短15分で迅速に導入

ノーコードのグラフィカルインターフェースから、モデルと推論ランタイムをワンクリックで導入し、RAGナレッジベースやNASデータソースに接続できます。

### 02 IT管理を簡素化

GPU、モデル、APIキー、モデルへのアクセス権、有効期限、1分当たりのリクエスト数（RPM）、利用記録、リモート運用を一元管理します。

### 03 AIモデルをサービス化

異なるオンプレミスモデルを標準化されたLLMエンドポイントとして公開し、アプリケーション、チーム、複数拠点で安全に共有できます。

### グラフィカル管理コンソール

ブラウザからリモートでアクセスし、グラフィカルインターフェースを通じて、モデル、エンドポイント、APIキー、アプリケーション、GPUリソース、システム状態を管理できます。

### ワンクリックモデル導入

用途特化型モデル、企業独自モデル、または承認済みのHugging Faceオープンソースモデルをワンクリックで導入できます。
ARX OSが推論エンジンとGPUリソースを自動的に設定するため、導入期間を大幅に短縮し、専門的なMLOps人材への依存を軽減します。

### モデルライフサイクル管理

モデルの登録、バージョン管理、更新、複数バージョンの比較テストを含む、モデルのライフサイクル全体を管理します。
各モデルのバージョンを体系的に管理し、必要に応じて以前の安定バージョンへ手動でロールバックできます。

### 安全なアクセス制御

許可リスト（Allowlist）、拒否リスト（Blocklist）、不正・異常アクセスの遮断、APIアクセス制御に対応し、より安全なプライベートAI利用環境を構築します。

### バックアップ＆リストア

アプリケーション、モデル、ベクトルデータベースをバックアップし、別のマシンへ復元できます。
ハードウェア障害発生時にも、サービスを迅速に再構築できます。

### プロジェクト単位のモデル分離

部門またはプロジェクトごとに独立したモデルインスタンスを起動できます。1対1または1対多など、柔軟な導入方式にも対応します。
各チームに専用のモデルサービスとナレッジスペースを割り当てることで、データ分離を実現し、部門間でのデータ混在や情報漏えいのリスクを低減します。

### 統合モデルプラットフォーム＆マルチランタイム対応

複数の推論エンジン、オンプレミス導入、GPUリソースの割り当てに対応した統合モデルプラットフォームです。
GPUリソース管理、APIキーの作成・失効・有効期限設定、モデル単位のアクセス制御、RPM管理、利用状況ダッシュボードを提供します。
1つのプラットフォーム上で異なるタイプのモデルを柔軟に導入・運用できるため、複数の異種推論環境を個別に構築・統合する負担を解消します。

### AIエージェント・MCP（Model Context Protocol）・CLI連携

既存の社内システム、ナレッジベース、データベース、各種業務ワークフローと連携できます。
企業やサードパーティーの開発者は、自動化アプリケーションを迅速に構築し、プラットフォームの価値を拡張できます。
また、将来の機能拡張やAIエコシステムとの連携にも柔軟に対応できます。

※導入時間は、検証済みモデル、標準構成および事前設定済み環境での参考値です。モデルサイズ、ネットワーク環境、ハードウェア構成により異なります。

# エッジAIプラットフォームの包括的ソリューション

**迅速なサービス展開、オープンソースエコシステムとの連携、PoCから本番環境への移行を想定して設計されています。**

## 1つのARX OSで、迅速なPoCから全社・多拠点展開まで対応

各段階のニーズに応じて、適切なコンピューティング性能と提供方式を選択できます。
モデルの導入、管理、サービス化においては、一貫した操作体験を維持できます。

01 POC 検証 → 02 Deploy 導入 → 03 Scale 多拠点展開 → 04 Operate 継続運用








### ARX-100

**小型エッジAIアプライアンス**

NVIDIA GB10

GPUハードウェアとARX OSを統合した、導入後すぐに利用できるオールインワンのエッジAIシステムです。

**最適な用途**
支社、小規模部門、PoCチーム、分散拠点。

**導入メリット**
GPU環境を一から構築することなく、プライベートモデルと生成AIアプリケーションを迅速に稼働できます。

参考価格：180万円～（税別）

### ARX-200 / 300

**企業向けプライベートAIシステム マルチGPU・ノーコードモデル運用対応**

NVIDIA RTX Pro 6000/ GB300 DGX AI Station ; BTO

安定性、管理性、ガバナンスを備えた主権型AIプライベートクラウド（Sovereign AI Private Cloud）です。

**最適な用途**
企業のIT部門、AIチーム、プライベートモデルのガバナンスを必要とする組織。金融、官公庁、医療、製造などの業界に適しています。

**導入メリット**
最短15分でモデルサービスを構築し、GPUインフラ構築の負担や、MLOps・推論分野の専門家への依存を軽減します。

OSソフトウェアライセンス：デュアルGPUノード単位で課金*

### vARX

**プライベートAI向けGPUコンピューティングリソースのレンタル＆リモートPaaS開発環境**

PaaS, by Subscription

月額ですぐに利用でき、専用の物理ハードウェアに近い操作体験を提供するプライベートクラウド型Platform as a Service（PaaS）です。

**最適な用途**
開発者、ISV、AI SaaSチーム、企業のAI開発チーム。

**導入メリット**
高額なハードウェアへの初期投資なしで、リモート開発、PoC、モデル導入、性能検証を実施できます。
料金は、GPUの機種・演算性能、利用期間、プロジェクト要件に応じてお見積もりします。詳細はWebサイトをご覧ください。

次世代GPUサーバーを、すぐに使える企業向けAIサービスへ進化させます

ARDGE: www.ardge.com 営業・事業提携に関するお問い合わせ: usasales@ardge.com

# データを、行動につながるインテリジェンスへ

### AIツールをワンクリックで導入｜App Market

モバイル端末のApp Storeのような操作体験を提供します。検証済みのAIツールをワンクリックでインストール・管理でき、基本的な導入・管理操作にコマンドラインは不要です。

### AIモデルセンター｜Model Hub

Hugging Faceエコシステムと深く統合されています。GUIによるモデルのダウンロード、推論管理、コンテナ化された導入に対応し、異なる環境間での一貫性と導入の安定性を確保します。




### ARGENT｜インテリジェント・ナレッジエージェント構築プラットフォーム

セルフサービス型ナレッジベース＆AIエージェントプラットフォーム
中小企業や技術的な専門知識を持たないユーザー向けに設計されています。
文書をアップロードすると、ベクトル化と構造化処理が自動的に実行され、最小限の設定で自然言語による質問応答システムを構築できます。

### PreciseMD｜高精度文書認識エンジン

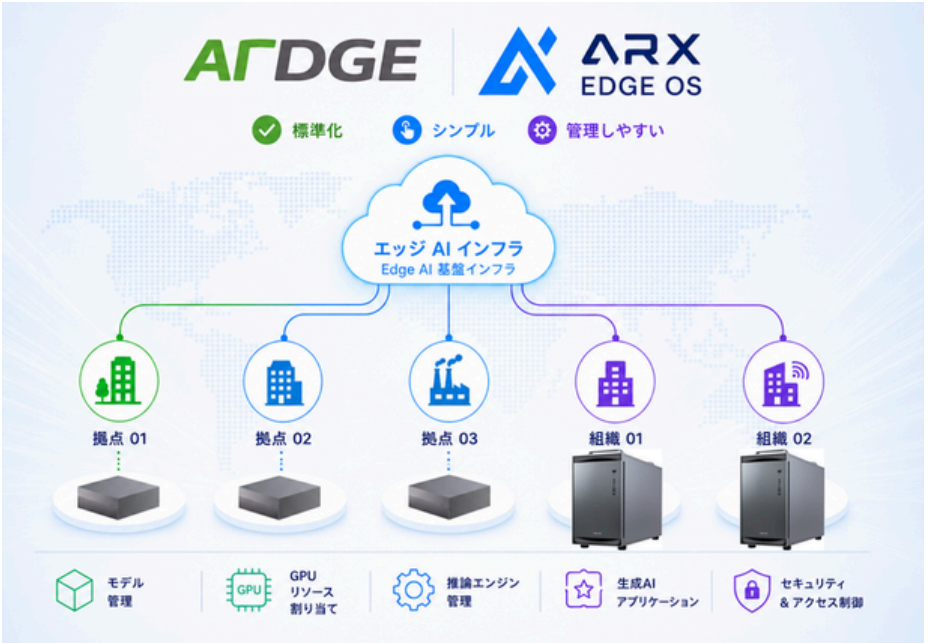
非構造化データをオンプレミスで一括処理する、高精度OCRデータ前処理ツールです。
CLIやMCPなどの連携インターフェースを提供し、ISVパートナーが同様の機能を一から開発する負担を軽減します。
複雑なレイアウトの認識と構造化データの抽出に対応し、AIによる回答の精度と品質を向上させます。

### ScribeSTT｜セマンティック動画検索

音声・映像コンテンツを構造化ナレッジへ
音声を高精度な文字起こしデータへ変換すると同時に、ベクトル化処理を実行します。
検索可能な動画ナレッジベースを構築し、自然言語を使ってコンテンツを迅速に検索・参照できます。

## オンプレミスAIを スケーラブルな製品・サービスへ

コンプライアンス、データガバナンス、知的財産の保護、主権型AIへの対応を背景に、企業がモデルのオンプレミス導入を求めるケースが増えています。
これは、AI SaaSやISVチームにとって新たな市場機会となります。
しかし、オンプレミス導入を一度限りの個別開発プロジェクトで終わらせることはできません。
異なるハードウェア構成、GPUプラットフォーム、顧客環境に対応するには、標準化され、管理可能で、継続的にサポートできる導入モデルが必要です。
簡素化された一貫性のあるオンプレミス・エッジ基盤により、ソフトウェアチームは複雑なインフラ統合に多くのリソースを投入する必要がなくなります。
各顧客環境を、標準化された制御可能なサービスノードへと変え、モデル、AIエージェント、ユーザー体験の革新に集中できます。




**日本での販売・事業提携について相談する**

次世代GPUサーバーを、すぐに使える企業向けAIサービスへ進化させます

ARDGE: www.ardge.com 営業・事業提携に関するお問い合わせ: usasales@ardge.com