

## 導入事例

インテル® Xeon® プロセッサ

intel  
xeon

# インテル® Xeon® プロセッサが支える AI 基盤でオンプレミス環境に 簡単かつセキュアに生成 AI を導入

エフサステクノロジーズが提供する「Private AI Platform on PRIMERGY」

業種や規模を問わず、多様な企業と組織が積極的に導入を進めている「生成 AI」。業務効率化をはじめ生産性向上、働き手不足の解消など、さまざまな導入効果が期待されており、近年もっとも注目されている技術トレンドといっても過言ではないでしょう。

### エフサステクノロジーズ株式会社

所在地：神奈川県川崎市中原区中丸子 13-2  
フロンティア武蔵小杉

設立：2024年4月

資本金：5億円

事業内容：サーバやストレージの開発・製造・販売・  
保守、ネットワーク製品の販売・保守、法人向け PC  
の販売

<https://www.fsastech.com/>

これまで国内企業の IT インフラを支え続けてきた富士通エフサスと、富士通のハードウェア事業を統合し、2024年4月に設立されたエフサステクノロジーズ。インフラ基盤を構成するハードウェアの開発 / 製造 / 販売 / 保守まで一貫した体制を構築し、従来から6万5,000社に及ぶ企業のビジネスに価値を提供してきた会社では、AIに関する企業のニーズをキャッチアップし、今後を見据えた先進的な生成 AI ソリューション「Private AI Platform on PRIMERGY」の提供を開始しました。

本稿では、エフサステクノロジーズ株式会社プロダクトソリューション本部 AI システム統括部統括部長の中嶋 一雄 氏に、本ソリューションの開発コンセプトと特長、インテル® Xeon® プロセッサを採用した経緯、生成 AI 活用における同プロセッサへの期待など幅広い話を伺いました。同社とインテルとの協業により進化を続ける、プライベート環境向け生成 AI プラットフォームの実力を探求します。

### 社内の機密データを含めた生成 AI 活用は、 オンプレミス上に AI 基盤を構築するのが有効な一手

エフサステクノロジーズは、国内に製造工場を持ち、Made in Japan の体制でサーバ・ストレージ・ネットワーク製品など、IT インフラを構成するハードウェア・ソリューションを提供し続けてきました。企業のビジネスをインフラ面から支えてきた会社は、AI がビジネスに与える価値について以前より注視しており、数多くの企業と接しているなかで、進化を続ける生成 AI への期待の高まりを実感していたといいます。同社のプロダクトソリューション本部 AI システム統括部を率い、ストレージ製品の開発や、スーパーコンピュータ「京」の安定稼働などに携わってきた中嶋氏は、AI の活用はより深い領域へと向かっていると市場の動向を分析しています。

「富士通が約 800 社の経営層に AI に関するアンケートを実施したところ、現状はチャットボットなどを用いた顧客対応の自動化や、対話型生成 AI による業務支援といった活用方法が主流でしたが、今後 3 年以内の活用予定を聞くと、商品やサービスの機能強化や経営意思決定の支援、業務プロセスの自動化など、ビジネスのより深い領域での活用を想定している企業が多いという傾向が見えてきました。特に生成 AI への期待が高いことを実感しています。」(中嶋氏)

複雑な意思決定を伴う領域で生成 AI を活用するには、企業内にある機密データを大規模言語モデル (LLM) に読み込ませ、より高精度な回答を得る必要があります。ところが現在の生成 AI 活用は、パブリック・クラウド上で提供されているサービスの利用がほとんどで、社内の機密データを外部 (パブリック・クラウド) に持ち出すことに抵抗を感じる企業も少なくありません。



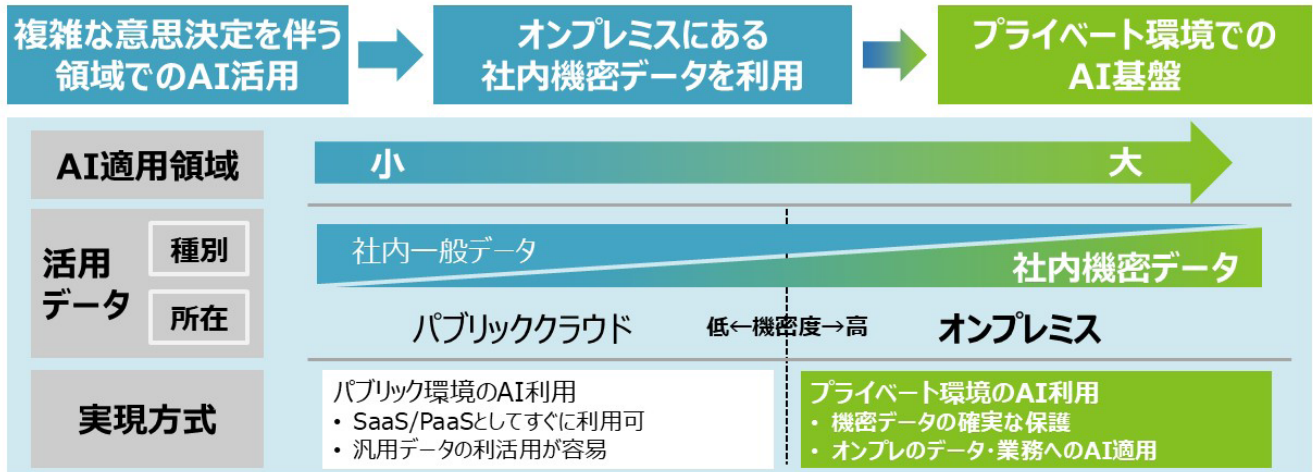
エフサステクノロジーズ株式会社  
プロダクトソリューション本部  
AI システム統括部  
統括部長  
中嶋 一雄 氏



## AI活用における課題



AIを複雑な意思決定が必要な領域で活用するためには、機密データの利用が重要  
機密データを安全に扱えるプライベート環境向けのAI基盤の活用が、新たな課題となります



© 2025 Fsas Technologies Inc.

中嶋氏は「経営の意思決定を支援するには、決算情報やNDAを締結している開発中の製品情報など機微なデータを扱わなければならない、SaaS/PaaSとして利用できるパブリック環境のAIソリューションは使えないと考えるお客様は、やはり一定数いらっしゃいます」と語り、ビジネスのより深い領域で生成AIを活用するには、機密データを安全に活用できる環境を構築する必要があると話を展開しました。

こうした生成AIの活用促進を阻む課題を解決すべく同社が開発したソリューションが、プライベート環境、すなわちオンプレミス上に構築/

運用できる対話型生成AIプラットフォーム「Private AI Platform on PRIMERGY」です。

### AI基盤に必要な要素を網羅した オンプレミス向け生成AIソリューション

前述のとおり、パブリック・クラウド上のサービスとして利用できるAIソリューションは、導入が容易でスピーディーに生成AIの活用を始められる反面、データ・セキュリティの観点から自社特有の業務資料や

## Private AI Platform on PRIMERGY とは



対話型生成AIの業務利用における課題

- 🔍 一般常識だけでなく、**自社特有の業務資料・データ**(※)を使って利用したい
- 🔒 **機密情報**(※)を取り扱うため、セキュリティ上**クラウド**を利用できない

**業務データや機密情報を扱えるようローカル環境に構築したいが  
手間もコストも抑えたい**

※：お客様との取引履歴、お客様向けシステムのソースコード、金融機関の融資関連情報、出願前の技術情報、個人情報など

### Private AI Platform on PRIMERGY

**機密データを社外に出さずに対話型生成AIを活用したい！  
を短期間で実現するソリューション**

© 2025 Fsas Technologies Inc.



機密情報を扱う領域には使いづらいというデメリットも内包しています。このため、事業の根幹を成す領域で生成AIを活用したい場合には、オンプレミス上にAI基盤を構築し、社内データを外部に出すことなくAIアプリケーションを運用するというアプローチが有効となります。

とはいえ、実用に耐えうるAI基盤をオンプレミスに構築するのは難易度の高いミッションであり、ハードウェアの調達から、OSや大規模言語モデル(LLM)の選定/導入、業務に適したAIアプリケーションの開発/運用まで、やるべきことは多岐にわたります。その結果コストの増大はもちろん、導入期間の長期化も避けられず、昨今のビジネスで重要な“スピード”と“柔軟性”を損なう恐れも出てきてしまうでしょう。

エフサステクノロジーズが提供する Private AI Platform on PRIMERGY は、こうした懸念を払拭するソリューションとして開発されました。中嶋氏は、本サービスの特徴として次の4つを挙げています。

- ① 導入後、すぐに使える Ready モデル
- ② AIの専門家が厳選した生成AI
- ③ 堅牢かつ信頼性の高いインフラ基盤
- ④ 業務に合わせた独自のAIアプリケーションを容易に作成可能

「開発コンセプトとしては、オンプレミス上で簡単に使えることを重視しました。生成AI基盤向けのハードウェアとしては、インテル® Xeon® プロセッサを搭載し、高性能GPUも選択できるサーバー製品「PRIMERGY」を採用。その上に検証済みのOSやコンテナ基盤、言語モデル、検索拡張生成(RAG)、OpenAI互換API/WebGUIなど対話型生成AI基盤に必要な環境をすべてセットアップしたうえでお届けする“Readyモデル”として提供しています(①)。電源をオンにするだけで、生成AIのチャットボットがすぐに使えるようになっています」(中嶋氏)

また生成AIの環境を構築するには、さまざまな特徴を備えた言語モデル(LLM)のなかから、日本語処理能力が高く業務活用に適したモデルを選定する必要があり、AIに関する専門知識を備えたエンジニアを確保できていない企業にとってその選定は大きな障壁となります。そこで本サービスでは、経験豊富なエフサステクノロジーズのAIエンジニアによる目利きで、業務に最適なLLMを選定(②)。膨大な選択肢から自社に合ったモデルを探し出す必要はありません。中嶋氏は「LLMはもちろん、チャット用のウェブGUIやAIアプリケーションの開発環境、RAGと呼ばれる検索拡張生成機能も含めて選定していますので、AIに関する知見を持たない企業でも生成AIを効果的に活用できます」と説明します。

インテル® Xeon® プロセッサ搭載のPRIMERGYなど堅牢で信頼性の高いAI基盤は、ハードウェアとソフトウェアを含めて継続的なサポー

## Private AI Platform on PRIMERGY とは



# Private AI Platform on PRIMERGY

### Private AI Platform on PRIMERGY コストパフォーマンスに優れた検証済プラットフォーム

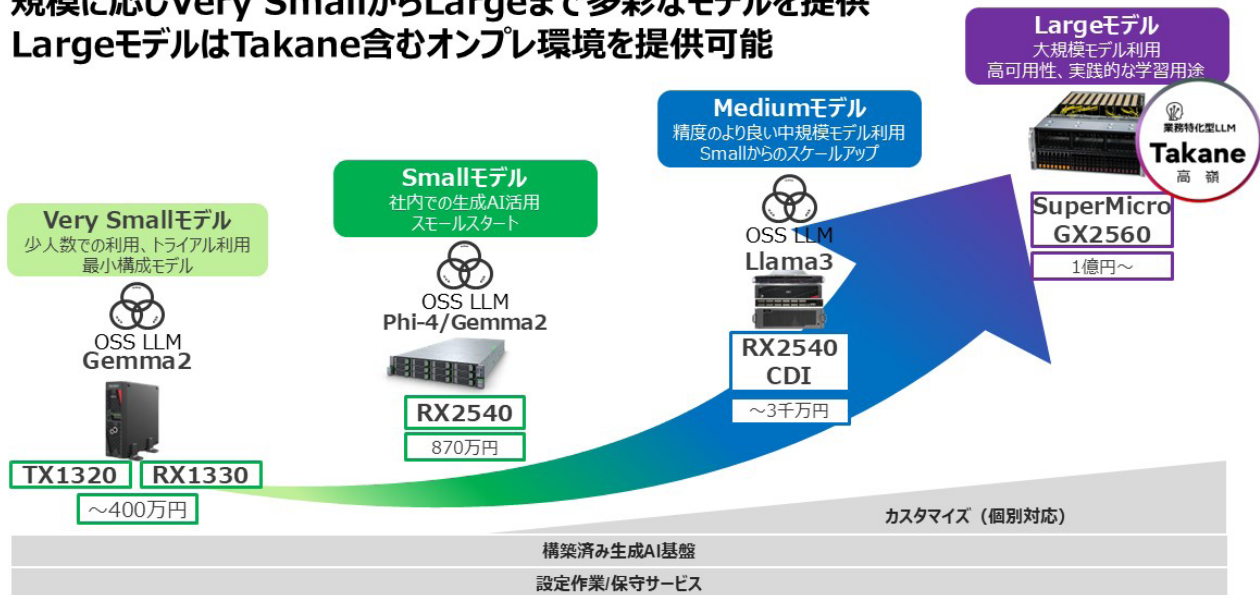


- 導入後、すぐに使えるReadyモデル
- AIの専門家が厳選した生成AI
- 堅牢かつ信頼性の高いインフラ基盤
- 業務に合わせた独自のAIアプリケーションを容易に作成可能

# Private AI Platform on PRIMERGYラインナップ



- 規模に応じVery SmallからLargeまで多彩なモデルを提供
- LargeモデルはTakane含むオンプレ環境を提供可能



© 2025 Fsas Technologies Inc.

トが提供されます(③)。ここには開発から保守までワンストップでインフラ基盤を提供し続けてきた同社の知見が惜しみなく注ぎ込まれており、安定稼働を実現する保守サポートは、ほかのAIソリューションにはない強みとなっています。

さらに本サービスでは規模に応じた4つのラインナップを用意しており、新たに追加された「Very Smallモデル」は、「少人数で利用したい」「トライアルで生成AIを試したい」といった企業のニーズに応えるモデ

ルだと中嶋氏は話します。最上位モデルとなる「Largeモデル」では、富士通が提供している業務特化型 LLM「Takane(高嶺)」が採用されるなど、規模や用途に合わせた選択肢が用意されていることも見逃せないポイントです。

そしてラインナップの豊富さだけではなく、最新の技術と機能を積極的に取り入れ、継続的に改善を図るスタンスも本サービスの大きな魅力となっています。2025年に入っても、先進的なAIアプリ開発プラット

## RAG 参照前の回答

### AI ChatBot

日本のランサムウェア被害状況を箇条書きで300文字以内で教えて

日本のランサムウェア被害状況は以下の通りです。

- **増加傾向:** 企業や個人を対象とした攻撃が増加しており、特に2020年以降顕著です。
- **影響範囲:** 過去には大企業だけでなく、中小企業や公共機関にも影響を及ぼしています。
- **被害額:** 支払われた身代金の総額は数億円に上るケースもあり、経済的な打撃が大きいです。
- **種類の多様化:** REvilやRyukなど、さまざまなランサムウェアが日本でも活動しています。
- **対策の必要性:** バックアップの徹底やセキュリティ対策の強化が急務とされています。
- **政府の対応:** 政府はセキュリティ対策のガイドラインを提供し、情報共有を推進しています。

質問してください

## RAG 参照後のより高精度な回答

日本のランサムウェア被害状況を箇条書きで300文字以内で教えて

RAG検索結果確認

日本のランサムウェア被害状況は以下の通りです：

- **増加傾向:** 国内外での被害が続き、企業や医療機関に影響を与えています。
- **影響の深刻化:** Emotetの活動再開やDDoS攻撃により、国民生活や社会経済に影響が出ています。
- **経済的損失:** 警察庁の調査によると、ランサムウェア被害に関連する調査・復旧費用は、約1億2,528万円の平均被害額が報告されています。
- **対策の強化:** 大型連休中のリスクを考慮し、政府機関が注意喚起を行っています。
- **攻撃手法の多様化:** IoT機器やHTTP/HTTPSポートを狙った攻撃が増加しています。
- **法的対応:** 不正アクセス禁止法違反事件の検挙件数は前年比1件減少しましたが、依然として注意が必要です。

質問してください

フォーム「Dify」を実装。直感的なAIアプリケーション作成や、RAGを利用した自社独自データの取り込みを実現しています(④)。

「例えば、ウェブGUIで“日本におけるランサムウェアの被害状況を教えて”と質問すると、LLMが学習済みの汎用情報を利用し、インターネット接続不要で回答を返してくれます。さらにRAGのデータも簡単な操作で取り込めるため、例えば日本の公的機関が公開しているサイバーセキュリティー被害情報を読み込ませて、より高精度な回答を得ることも可能です。参照したRAGデータも確認できます」(中嶋氏)

Difyを導入したことで、業務用途に合わせた専用のAIアプリケーションを作成することも容易になったと語った中嶋氏は、「例えば総務部門向けのチャットボット作成であれば、就業規則など社内規程の情報を読み込ませて、回答精度の向上が図れます。作成したチャットボットは既存のWebサイトへの組込みも簡単に実施できるため、既存の業務に生成AIを取り込むことも容易です」と、進化を続ける本サービスの魅力に言及しました。一部の部署だけが生成AIの恩恵を受けている状況を打破し、あらゆる業務で生成AIを当たり前で活用できる“AIの民主化”を推進できるソリューションと説明します。



## “GPUありき”という生成AI活用の固定観念を覆す — インテルとの協業でCPU推論という新たな道を模索

エフサステクノロジーズでは、Private AI Platform on PRIMERGYを自社導入し、実践的なアプローチで導入効果の検証を続けています。「弊社のサポート業務に適用し、お客様からの問い合わせ対応業務における工数削減、回答品質の向上、属人化の解消など多くの効果を得られています」と中嶋氏。過去の対応履歴やマニュアル/設計書などの情報をLLMに読み込ませることで、担当者の経験やスキルを問わず、均一的で高精度の問い合わせ対応が可能になったと手応えを口にしています。

実際、本サービスの導入を検討する企業の多くは、必要な情報を引き出すための工数を削減したいというニーズを持っているといいます。例えばパブリック・クラウドの利用が困難な製造業では、過去の設計情報を素早く検索し、流用設計で新製品開発を効率化するという目的で本サービスの採用を検討しています。そのほか、定期的に改訂されるドキュメントの新版と旧版の差分を分析し、どのような改訂があったのかを要約するといった使い方も検討されていると中嶋氏は語り、産業界や公共系、ヘルスケアまで幅広い分野で引き合いをいただいていると力を込めました。

## [参考] 活用例 Very Small, Smallの場合



### 固有の業務向け

種別	活用シーン
製造	過去の設計書や技術文書といった必要な情報を迅速に抽出し、開発を効率化、など
金融	複雑な金融商品の約款等に対する質疑応答、経済見通しのレポートの要約、など
小売	顧客の購買履歴や嗜好を分析し、パーソナライズされたメールの配信、など
政府・自治体・公共・文教	分野外の審査員に対しても伝わる補助金申請書の生成、など
ヘルスケア・製薬	診療室の対話から電子カルテを作成、カルテから紹介状を作成、など
通信・メディア	ニュース記事、ブログ統合の自動生成、など

### 共通業務向け

種別	活用シーン
サポート	チャットボットによる自己解決率向上、24h・365d 対応の実現、など
リサーチ・レポート	学術論文、技術文書の検索、業務報告等の要約、など
マーケティング	顧客フィードバックやレビューの要約、など
開発	仕様書や設計書、Codeの自動生成、など
法務・財務	申請書の作成、稟議ルールの確認、など

また本サービスは、信頼性の高いGPUを搭載したハードウェア構成を基本としている一方で、今後も先進的な技術や最新のハードウェアをキャッチアップし、市場のニーズに応じたブラッシュアップを続けていく予定です。中嶋氏は、CPUによる処理を重視した生成AI活用も今後増えていくと予測しており、AI/HPCのワークロード向けに機能が強化された最新世代のインテル® Xeon® プロセッサのポテンシャルに期待していると話を続けました。

「エフサステクノロジーズでは、インテルと長い協業の歴史があり、単にCPUの供給を受けるのではなく、さまざまな技術協業を行ってきました。AI/DX時代においても、ハードウェアだけにとどまらず協業の幅を広げています。現時点では、大規模な言語モデルを動かすにはどうしてもGPUが必要となりますが、例えばAIアプリの安定稼働など、その他の処理はCPUが行っている部分も非常に多い。LLMが頭脳でGPUが処理を担うとすると、実際に動かすための手足となる部分をCPUが担っているわけです。さらに第4世代以降のインテル® Xeon® スケーラブル・プロセッサに実装されている拡張命令セット「インテル® AMX」を活用すれば、CPUでの推論処理も十分なパフォーマンスが期待できます。将来的には、用途に合わせてCPUで推論を行うPrivate AI Platform on PRIMERGYのモデルも検討しており、製品化を見据えた技術検証も進めています」(中嶋氏)

## インテル® AMXを利用したCPU推論モデルも検討し、ビジネスにおけるAIの民主化を目指す

前述したとおり、Private AI Platform on PRIMERGYは技術の進化や市場のニーズに合わせてラインナップの拡充や最新AI技術の実装

などの改善を続けています。今後もプライベート環境で簡単かつ安全に使える生成AIソリューションというコンセプトを継続しながら、企業のビジネスに価値を届けていきたいと中嶋氏。インテル® AMXの活用に向けた技術検証も含め、インテルとの協業をさらに深めていきたいと今後の展望を語りました。

「1980年代から数十年にわたり続けてきたインテルとの技術協業を継続し、生成AIをはじめとした企業のDX推進、IT活用を支援していきたいと考えています。AIソリューションのラインナップ強化に向けて、インテル® AMXを利用したCPU推論の活用も検討しており、インテルとともに市場を盛り上げ、ビジネスにおけるAIの民主化を牽引していきたいと思っています」(中嶋氏)



Private AI Platform on PRIMERGY

<https://www.fujitsu.com/jp/products/computing/servers/primergy/solution/private-ai-platform/>

インテル® Xeon® プロセッサ

<https://www.intel.co.jp/content/www/jp/ja/data-center/what-is-xeon-processor.html>

# intel xeon

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、<https://www.intel.com/PerformanceIndex/> (英語) を参照してください。

性能の測定結果は、構成に示されている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アップデートが適用されているとは限りません。構成の詳細については、補足資料を参照してください。絶対的なセキュリティを提供できる製品やコンポーネントはありません。

実際のコストや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データの正確さを評価してください。

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。

テックプラス (2025年4月11日) に掲載されたコンテンツから抜粋し、再構成したものです。

インテル株式会社

〒100-0005 東京都千代田区丸の内1-4-1 丸の内永楽ビル25階

<http://www.intel.co.jp/>

©2025 Intel Corporation. 無断での引用、転載を禁じます。

2025年4月

365231-001JA  
JPN/2504/PDF/SE/MKTG/TK