AI同士が連携する「AIファクトリー」実現に向けて最適なインフラとは？

セキュアかつ高パフォーマンスな環境でAIエージェントの活用を促進

生成 AI や AI エージェントの登場によって、企業の業務効率化や意思決定の高度化の取り組みは新たな段階に入りつつある。だが実際に AI を業務へ組み込み、持続的に価値を生み出す仕組みである「AI ファクトリー」を構築するには、まだ多くの壁が存在する。本稿では、こうした課題を乗り越え、安全かつ迅速に AI ファクトリーを実現するためのアプローチを紹介する。

AI の役割を変える AI エージェント いかに早く取り入れるかがビジネス革新のカギ

AI のビジネス活用が急速に進む現在、その役割は変化してきている。当初は画像解析や文書の要約といった単一の目的を果たすためのツールであった AI は、いまや業務を独自に遂行し、リードする存在になりつつある。このように「自律的・協調的に活動し、判断を伴う複雑な業務を完遂できる AI」は AI エージェントと呼ばれる。

今後 AI エージェントの業務適用はさらに拡大していくだろう。例えば、営業活動を支援する AI エージェントでは、メールや CRM のデータなどを参照して顧客満足度を向上させるための施策を営業担当者に提案するといった活用方法が考えられる。

AI が持続的に価値を創出し続ける仕組みである「AI ファクトリー」を実現する試みも盛んに行われている。すなわち、AI エージェントが必要なデータをリアルタイムに取得して、より精度の高い推論を行い、その結果を AI 自身にフィードバックして、さらに推論の質を高めていくための土台だ。AI ファクトリーのためのプラットフォームを迅速に展開することで、業務革新をさらに加速することができるだろう。

AI 活用は進むも、セキュリティ対策は後回しに 年々広がる「セキュリティギャップ」

現在の AI プラットフォームはパブリッククラウド上が主流であり、AI エージェントもクラウドサービスを中心に普及する可能性が高い。しかし、その場合に大きく 4 つの課題が残る。

1 つ目がコスト増大のリスクだ。クラウドは従量料金モデルを採用しているケースが多く、AI を適用する業務の拡大や利用対象となる従業員の増加に伴って、コストの増大が問題となるという側面がある。2 つ目が、サービス依存のリスクだ。AI モデルが世代交代して互換性がなくなったり、もしサービス提供が停止されたりした場合、ビジネスを継続できなくなる可能性がある。

3 つ目は、ガバナンス低下のリスクだ。部門や個人で好きな AI サービスを利用していると、契約しているサービスを把握・管理できなくなるおそれがある。4 つ目は、生成物に関するリスクだ。AI の生成物を利用した場合、意図せず著作権を侵害してしまう懸念がある。

重要なのは、AI 活用が進むほど、より強固なセキュリティ対策が求められることである。しかし、その対策には遅れが生じているという実情もある。今後 AI に関するセキュリティ対策の方針としては、「フレームワークと運用モデルの構築」「アクセス管理」「データ保護」、「AI 駆動型の脅威予測や対策自動化」などが求められるという考えも提唱されている。いずれにしても、これらを実践するには、AI の実行環境をプライベートクラウドに作る「AI 主権（ソブリン AI）」のアプローチが有効だ。それを実現可能なソリューションとして提供しているのが Nunitax である。

ソブリン AI を実現して 安全な AI 活用を促進する機能を提供

ソブリン AI の基盤となる「Nutanix Cloud Platform（NCP）」は、セキュリティ・バイ・デザインの理念に基づいて構築されている。データ基盤となるソフトウェア（Nutanix AOS Storage）やハイパーバイザー（Nutanix AHV Virtualization）は、万が一改ざんや変更されてしまった場合でも自動的に復旧することが可能だ。

NCP 上で AI エージェントやアプリを実行する際は、ゼロトラスト型のネットワーク制御を行うことができる。具体的には、必要なネットワーク通信をあらかじめ設定・許可し、

それ以外はすべてブロックする。万が一、この環境下でランサムウェア攻撃を受けた場合は、仮想マシンやコンテナのネットワークを自動的に遮断することが可能だ。

ソブリン AI の実行環境に関しては、「Nutanix Kubernetes Platform（NKP）」が役立つ。AI のフレームワークの多くは Kubernetes に準拠しているため、セキュアなコンテナ基盤導入の検討が求められる。そうした中で NKP では Kubernetes 環境の可視化、ロギング、権限管理、セキュリティポリシー制御などの機能を提供している。

ソブリン AI の管理・統制を担うのが「Nutanix Enterprise AI（NAI）」だ。これは、AI モデルを一元管理するためのモデルカタログを提供すると同時に、どのモデルを誰が使用できるのかを統制できるものだ。組織内の AI 利用者は認可された AI モデルのみ、統一された API を介して利用を行う。もし AI の挙動に問題が生じた場合は AI 管理者が API 公開を停止したり、特定のユーザアクセスを制御することも可能であり、AI 利用の透明性を確保する。

ソブリン AI においては、機密データの連携方法も検討する必要がある。例えば、AI エージェントがデータベースにアクセスする場合、動的にクエリを生成してリアルタイムにアクセスすることがパフォーマンス問題の引き金ともなり得る。

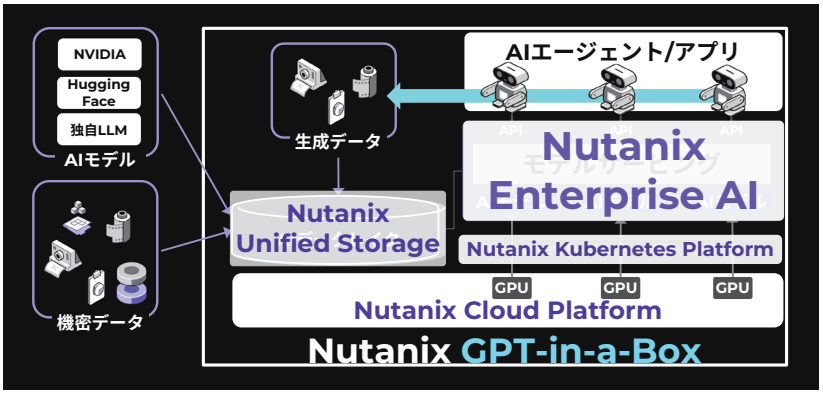
そこで Nutanix Database Service（NDB）を利用すると、仮想マシンにデータベース環境のクローンを作成し、AI がそのクローンにアクセスすることで、元データへの負荷分散が可能になる。加えて、NDB のデータリフレッシュ機能を用いれば、AI エージェントの推論の質を左右するデータの鮮度を維持しながら、AI によるデータアクセスを最適化できる。

非構造化データについては、Nutanix Unified Storage（Nutanix Files Storage ／ Nutanix Objects Storage）が役立つ。マルチモーダルと呼ばれる多様な形式のデータを保存できる同製品には WORM（Write Once Read Many）機能があり、データの改ざんを防止する。AI には、このイミュータブルのストレージ領域にアクセスさせる。なお同製品は NVIDIA の「AI Data Platform」認定を受けており、NVIDIA の各種 AI ソリューションと連携でき、高水準のパフォーマンスを発揮している。

AI ファクトリーの構築を省力化する Nutanix GPT-In-a-Box

AI ファクトリーを実現する上では、AI エージェントが稼働するためのコンピューティング環境や、AI と連携するためのデータ領域の導入・連携など、多くの時間と労力が必要となる。そこで迅速な構築を実現する上で役立つのが、「Nutanix GPT-In-a-Box」だ。

これはいわば Nutanix が提供するさまざまな製品群を統合した「AI 基盤セットアップパッケージ」であり、ここまで解説してきたようにセキュリティ・バイ・デザインを取り入れたセキュアな AI 基盤を構築できる。加えて、NKP 環境があれば、NAI のセットアップは NKP マーケットプレイスでボタンを 1- クリックするだけですぐに利用可能だ。Nutanix は、AI ファクトリーの構築を支援するパートナーエコシステム（NVIDIA、Dell、Cisco、HPE など）とも連携して、企業の迅速な AI 導入を後押ししていく。



Nutanix GPT-In-a-Box はプライベート AI ファクトリーをフルスタックで提供

Nutanix Japan 合同会社

E-mail:contact-jp@nutanix.com
https://www.nutanix.com/ja/contact-us