TechTarget Japan SPECIAL

“AIが業務の主体になる時代”に備える

AIエージェント導入を阻む「インフラの壁」解決の糸口を専門家に聞く

生成 AI の進化が次の段階に入り、企業の関心は目的を理解して自律的に行動する「AI エージェント」や「エージェンティック AI」の実戦投入に移りつつある。こうした AI 時代の新たなインフラ要件とその解決策について、専門家に話を聞いた。

生成 AI の利用が広がり、AI エージェントやエージェンティック AI への関心が高まっている。アイティメディアの「キーマンズネット」が実施した読者調査では、「生成 AI に関連するトレンドについて興味があるものを選んでください」という質問に対して「AI エージェント（エージェンティック AI）」と回答した人が「RAG」（検索拡張生成）などを抑えて最も多かった。

NTT ドコモビジネス 末松慎司氏（ジェネレーティブ AI タスクフォース 担当課長）

AI エージェントは特定のタスクを自律的に処理するものだ。さらに、複数の AI エージェントが連携してより高度なタスクを処理するのがエージェンティック AI であり、これが普及すると AI が単なる業務支援ツールではなく AI が意思決定や実行の主体になる可能性がある。

エージェンティック AI を含む AI サービスを利用するには、機密データの取り扱いやコストなどの課題があるが、オンプレミスで AI を利用する際にもさまざまな課題が付きまとう。企業が AI を本格的に活用する際に直面するインフラ課題と対策は何か。専門家に話を聞いた。

生成 AI の活用が進み、企業の関心は自律的にタスクを処理する「AI エージェント」や「エージェンティック AI」の実戦投入に移りつつある。そのような高度な AI システムを利用するためのインフラ要件について、NTT ドコモビジネスの加賀山英之氏（クラウド&ネットワークサービス部 担当課長）、末松慎司氏（ジェネレーティブ AI タスクフォース 担当課長）、管原康弘氏（ジェネレーティブ AI タスクフォース 主査）の 3 人に話を聞いた。

エージェンティックAIを自社専用に構築する意味

生成 AI ソリューションの開発を担当している末松氏は、エージェンティック AI が描く未来について次のように説明する。

「エージェンティック AI が普及すれば、既存の業務システム内のデータを自在に活用し、個々の業務を担う複数のエージェントが連携、判断しながら、一連のビジネスプロセスを自律的に完遂するようになります。この段階になると、AI は単なる “ 業務支援ツール ” ではなく、業務を回す “ 主体 ” になるでしょう」

現在、生成 AI の業務適用では、人間が主体となって進める作業を高速化することで、業務時間を短縮する使い方が多い。しかし、これでは一般的な業務の効率化にとどまる。近年では、ただ作業を速くするのではなく、より付加価値のある活用に向けて生成 AI をコア業務に適用し、社内のデータを連携させようとしている企業が増えている。これを実現しようと考えるとき、特に金融や医療、防衛分野のように機微な情報を取り扱う業界では、

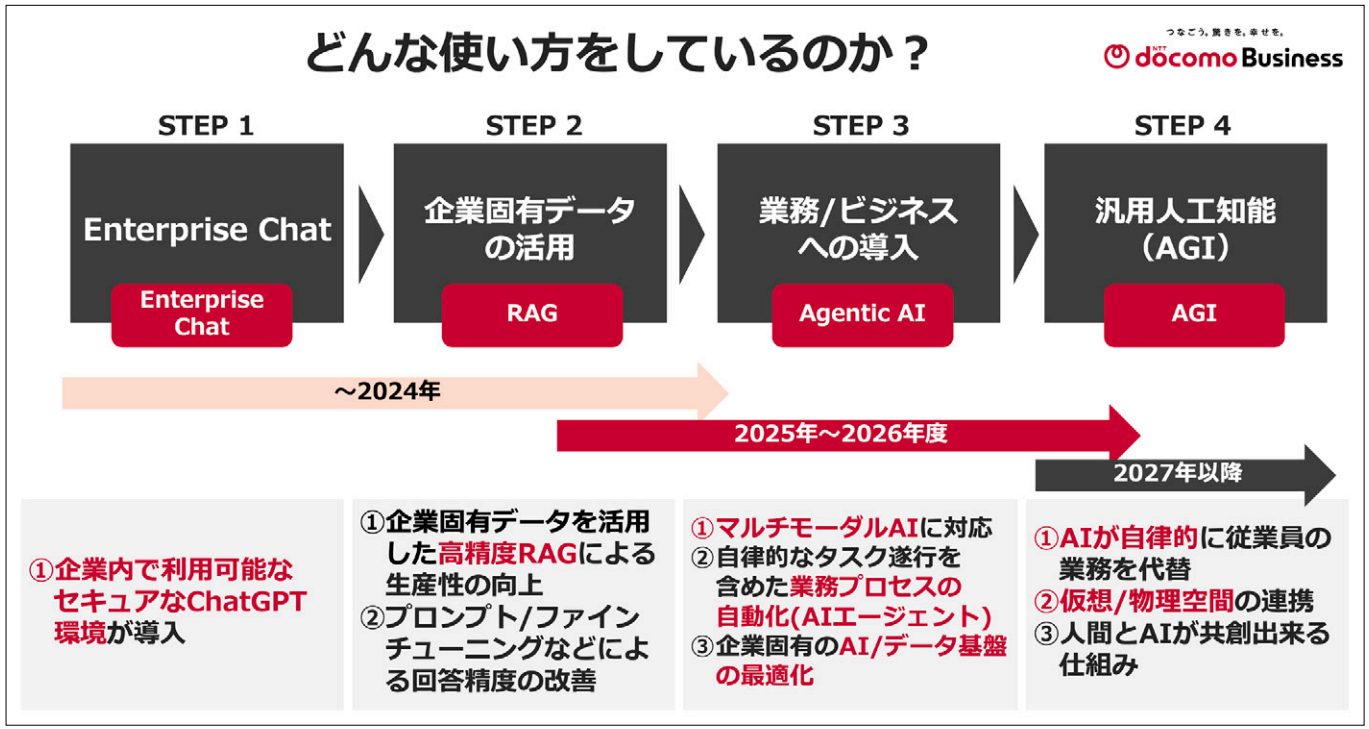

図 1　企業における AI 活用の変遷（提供：NTT ドコモビジネス）

それらの情報を AI に学習させたり参照させたりするために自社専用のインフラを構築する必要がある。

AIの本格導入を妨げるセキュリティとコストの課題

既に AI サービスの PoC（概念実証）を実施して一定の成果を得ている企業もある。しかし、管原氏によれば本番導入フェーズに移行する際にコストとセキュリティの課題に直面する企業が多いという。

「PoC では生成 AI や AI エージェントが自分たちの業務にどの程度適合するかというアプリケーションレイヤーの評価に重点が置かれるので、クラウド型の AI サービスで素早く実現性を確認するケースが多いです。しかし、本格的な商用利用や全社展開に移行するフェーズになるとコストやセキュリティの問題が顕在化します」（管原氏）

コスト面を見ると、AI モデルを特定の用途に特化させるために、自社のデータを使ったモデルの調整（ファインチューニング）を繰り返し実施するケースで、ハイパワーな GPU を長期間使用することになるため GPU を従量課金で使うサービスは不向きだ。

セキュリティ面の課題としては、クラウド型の AI サービスではデータ漏えいリスクが高まることを考慮しなければならない。

「基幹システムのデータなど、ビジネスの根幹に関わる重要なデータを社外のサービスに送信することは多くの場合、企業ポリシーやコンプライアンスで禁止されています。データの種類によっては、法令や規制によって保管場所を日本国内にとどめなければならないケースもあります。このため、オンプレミスやプライベートクラウドでの AI インフラ構築に注目する企業が増えています」

NTT ドコモビジネス 管原康弘氏（ジェネレーティブ AI タスクフォース 主査）

また、コストとセキュリティの両面に絡むのがエージェント同士の通信だ。クラウド型の AI サービスを利用する場合は外部ネットワークとの通信が発生するため、適切な管理をしなければセキュリティホールになりかねない。通信トラフィックが予測しにくく、時と場合によって増減するため、スケーリングを考慮せず帯域に余裕を持たせるとコスト増の原因にもなり得る。

AI時代のインフラ課題に挑むNTTドコモビジネス「AIインフラソリューション」

企業が本格的な AI 活用に踏み出す際に直面するこれらの課題に対して、NTT ドコモビジネスはデータセンターやネットワーク、GPU サーバなどをワンストップで提供する「AI インフラソリューション」で支援する。

同ソリューションには、顧客専有型 IaaS で GPU 基盤を提供する「GPU プラットフォーム」、液冷方式のサーバ機器に対応する省エネ型データセンターサービスの「Green Nexcenter」、IOWN 構想に基づく APN（All-Photonics Network）技術を利用した「docomo business APN Plus powered by IOWN」（以下、docomo business APN Plus）、セキュリティー体型統合ネットワークサービス「docomo business RINK」などが含まれる。

docomo business APN Plus と docomo business RINK はネットワークを仮想化してソフトウェアで制御する（Software-Defined）NaaS（Network as a Service）だ。加賀山氏はこれらのサービスでコスト面の課題を解決できると説明する。

「NaaS を利用することでトラフィックの増減に応じて帯域を柔軟に変更でき、コストを最適化できます。利用する場所や用途によって最適なサービスをお客さまにご提案します」

顧客の要件に応じて柔軟な GPU 構成をマネージドサービスとして提供する「GPU プラットフォーム」もある。これは国内データセンターで運用する専有型 IaaS であり、定額で利用できることから AI モデルの学習といった GPU 稼働率が高い用途で特に高いコストパフォーマンスを期待できる。従量課金で GPU を使いたい顧客向けに、GPU ベアメタルサーバをオンデマンドで利用できるメニューもある。

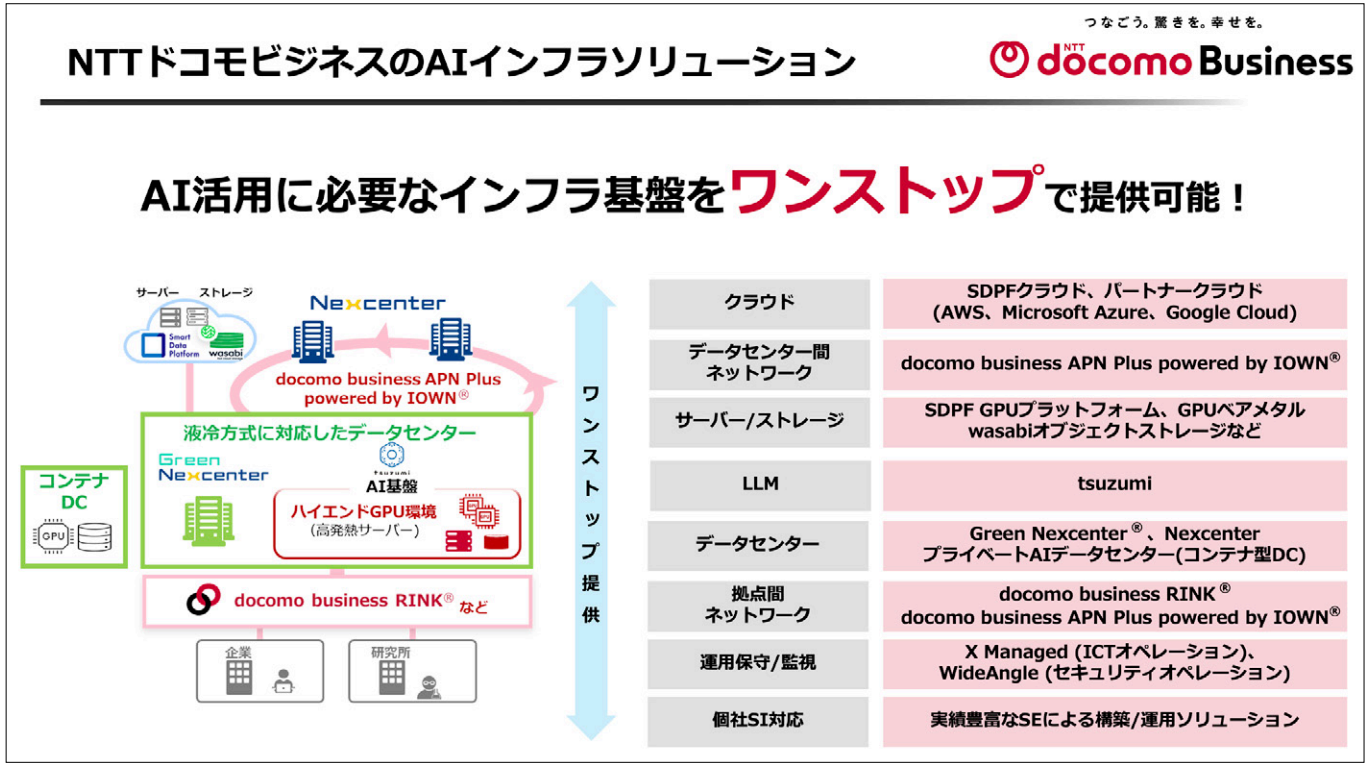

図 2　AI インフラソリューションの全体像（提供：NTT ドコモビジネス）

AI ワークロードの増加によって GPU サーバの発熱量が増加しており、従来のデータセンターでは対応が難しいことから必要な GPU リソースを調達できていない企業もある。また、工場などの近くでデータを処理したいというニーズもある。それらに対し、高発熱サーバの冷却に対応した液冷方式のデータセンターサービス「Green Nexcenter」や、コンテナ型データセンター「プライベート AI データセンター」などで、全国の AI 需要とさまざまな要望に対応する。

プライベート AI データセンターはクライアントの敷地内に設置することで外部のネットワークへの接続を減らせるため、セキュリティ面でも効力を発する。分散したインフラ全体を防御するなら、docomo business RINK の「WAN セキュリティ」機能が役に立つ。

「工場の産業機械やロボットのデータ、IoT 機器で集めたセンサーデータなども AI が処理する重要な情報源です。しかし、それらの機器は EDR を入れられないことも多く、サイバー攻撃の標的になるリスクが高まります。WAN セキュリティはネットワーク組み込み型のセキュリティ機能で、WAN 内の不正な通信や不審なパケットを検知して顧客に通知します。速やかな通信の遮断

NTT ドコモビジネス 加賀山英之氏（クラウド&ネットワークサービス部 担当課長）

まで一貫してサポートし、サイバー攻撃による被害の拡大を未然に防ぎます」

AIの自律的な通信が増えることで、インシデント発生時に原因や影響範囲の調査が困難になる可能性もある。docomo business RINKの「フローコレクター機能」は、通信キャリアのネットワーク内で通信ログを取得・保存することによって、インシデント発生時の早期対応を支援する。

“AIが業務を回す時代”に備える基盤

エージェンティックAIを含むAIの活用は、ビジネスやわれわれの働き方に抜本的な変革をもたらす。末松氏は、この変革の潮流について次のように語る。

「人がAIを活用する段階から、AIが業務を回す形に少しずつ変わってきています。既に一部の業務をAIに任せる企業も出てきており、今後はさらに適用範囲が拡大するでしょう。AIは人間に寄り添って共に仕事をする仲間のような存在に進化していきます。マルチモーダルAIやフィジカルAIの活用もさらに進み、各企業、各産業の在り方も大きく変わっていきます」

NTTドコモビジネスはそのような時代にビジネスを展開する企業を支援するため「AI-Centric ICTプラットフォーム構想」を推進している。企業がAIを活用して競争力の強化やビジネスモデルの変革を実現するための、セキュアで柔軟な環境の実現を目指す。加賀山氏によると、この構想は企業がAI時代に求める分散、柔軟、安全、リーズナブルという4つのニーズに応えるものだという。分散は、データや計算リソースがデータセンターやクラウドなどのさまざまな場所に分散していること。柔軟は、ハードウェアリソースやネットワークトラフィックの増減に柔軟に対応できること。安全は、企業が持つさまざまな機器をカバーする緻密なセキュリティ。リーズナブルは、さまざまなサービスやリソースの組み合わせによってコスト最適化が可能なことだ。

「当社は、AI-Centric ICTプラットフォーム構想とそれを実現するAIインフラソリューションによって、お客さまのニーズに合わせたAI活用を支援し、企業競争力の強化に貢献します」（加賀山氏）

生成AIを活用して自社の競争力を増す流れは今後一層強化されていくだろう。企業は今、生成AIを活用することで発生するリスクと“活用しないリスク”を天秤にかけている。AIエージェントやエージェンティックAIは活用するかしないかで生成AI以上の差が出る可能性を秘めている。本格的にAI戦略に取り組むなら、大切なのは安定した基盤の上で戦うことだ。

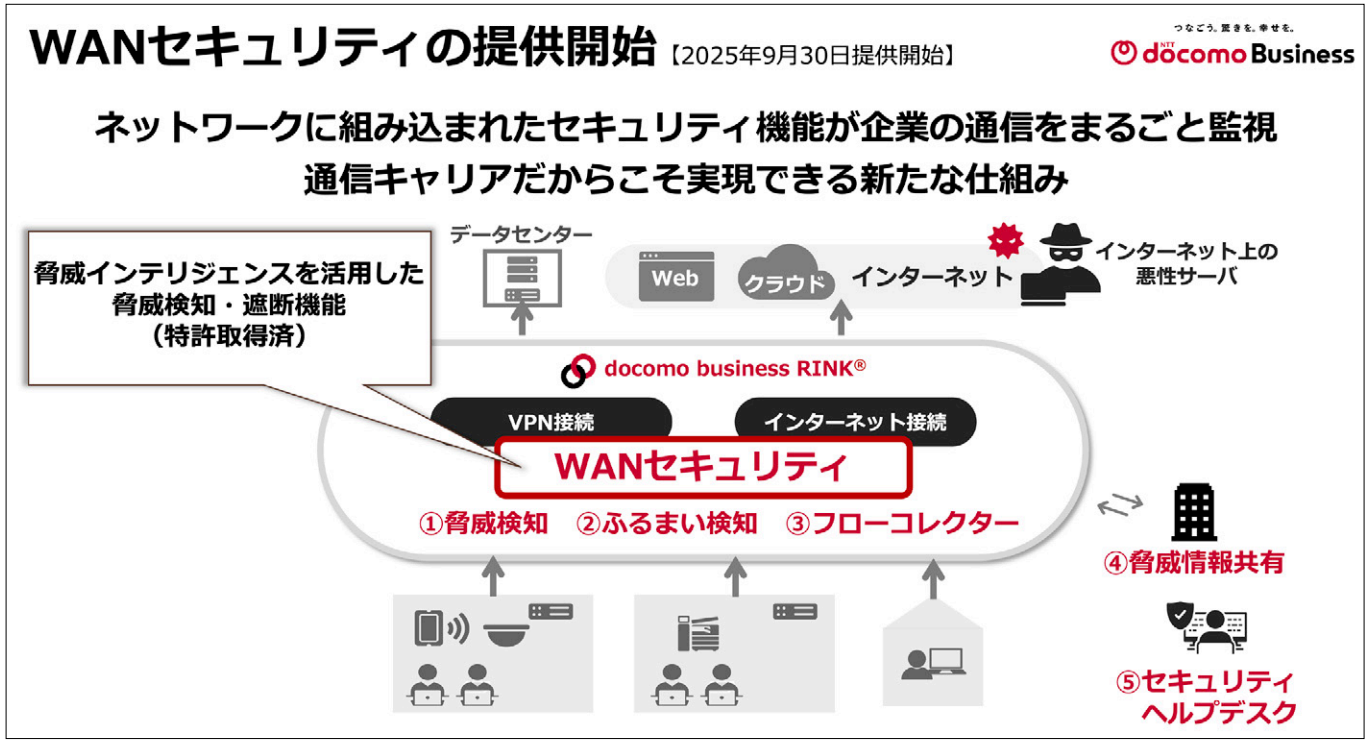

図3　WANセキュリティによる脅威検知と遮断（提供：NTTドコモビジネス）

NTTドコモビジネス株式会社　プラットフォームサービス本部　クラウド & ネットワークサービス部

AIインフラソリューションに関する詳細はこちら　https://www.ntt.com/business/lp/aiinfra.html
お問い合わせ　https://www.mkt.ntt.com/jp_PS_inq_AI-Infra_25_223_wtc_reg.html

※この冊子は、2025年12月に掲載されたアイティメディア編集局制作コンテンツを再構成したものです。
https://techtarget.itmedia.co.jp/tt/news/2512/12/news02.html