つなごう。驚きを。幸せを。
NTT docomo Business

2026 年 5 月 12 日

NTT ドコモビジネス株式会社

AI エージェント経済圏の信頼の基盤となる「AI エージェント属性情報レジストリ(仮称)」のプロトタイプを開発

～検証可能なデジタル証明書(VC)により AI のデジタルトラストを担保～

NTT ドコモビジネス株式会社(旧 NTT コミュニケーションズ株式会社、以下 NTT ドコモビジネス)は、AI エージェント※1 同士が自律的に取引や連携する社会を見据え、AI エージェント自身の信頼性を確認する仕組みの技術検証(以下 本検証)を開始しました。本検証では、AI エージェントの属性情報を一元的に管理・検証する AI エージェント属性情報レジストリ(仮称)のプロトタイプ(以下 本プロトタイプ)を開発し、その有効性を検証します。

1. 背景

生成 AI の進化により、特定の業務を自律的に実行する AI エージェントの活用が広がっています。近年では、複数の AI エージェントが連携し、複雑な業務を完結させるマルチ AI エージェント※2 も台頭しており、AI エージェント間の通信規格 Agent2Agent Protocol(A2A)※3 の登場によって、人間を介さずとも AI 同士が経済活動を自律的に行う世界が現実のものとなりつつあります。

こうした AI エージェント主導の経済圏が拡大する一方で、従来の人間中心のセキュリティ対策では防ぎきれない、AI 特有の新たなリスクへの対処が急務となっています。セキュリティ向上を目的とした国際的なコミュニティーである OWASP(Open Worldwide Application Security Project)※4 などは、生成 AI を利用するアプリケーションの代表的なセキュリティリスクとして、過剰な権限を与えられた AI の不正動作や、機密情報の漏えいを報告しています※5。特に、企業が AI エージェントを用いたサービス連携や自動取引を導入する際、相手となる AI の開発主体や権限を客観的に確認できない現状が、実運用上の課題となっています。

AI が自律的に活動する環境では、その信頼性や指示の正当性を担保するために、AI の属性情報を厳格に管理・提示・検証する必要があります。そのため、なりすましや改ざんを防ぐ AI のデジタルトラストサービスの構築が、安全で信頼できるオンライン経済活動を実現するための重要な基盤となります。

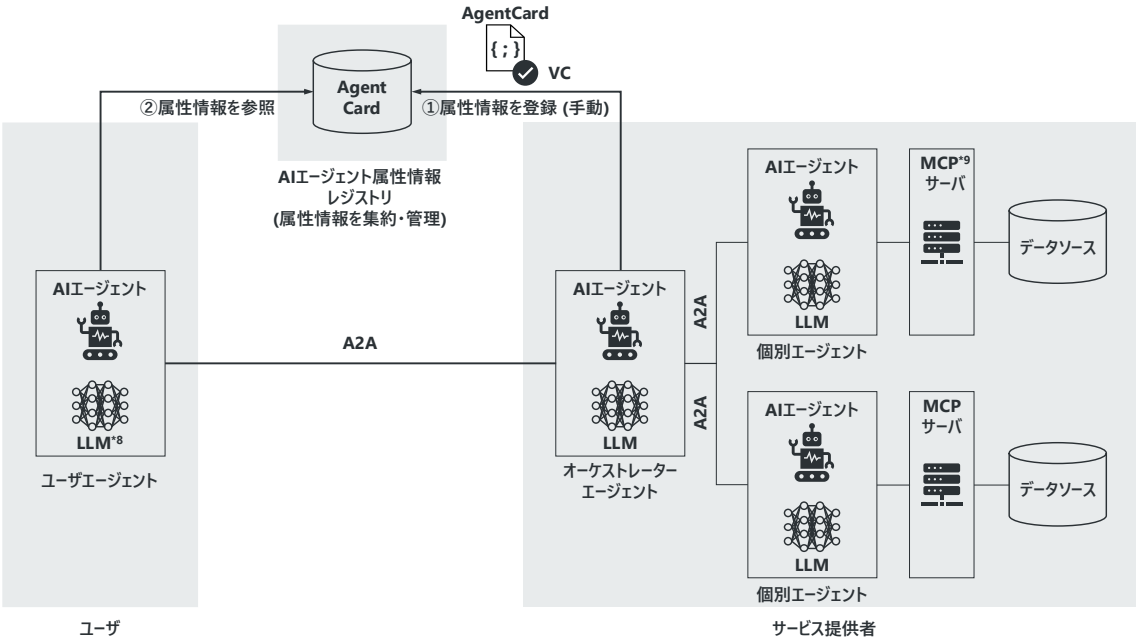
2. 概要

こうした背景を踏まえ、NTT ドコモビジネスは、AI エージェントの信頼性を担保する属性情報を一元的に登録・管理・公開できる基盤として本プロトタイプを開発し、技術検証を開始しました。

本プロトタイプでは、AI エージェントの身分証明書にあたる AgentCard※6 を用いて、属性情報を集約・管

理します。本検証はこれを活用し、AI エージェントそのものの振る舞いを直接保証するのではなく、その発行・運用主体や実行権限に関する属性情報をデジタル証明書の標準規格である Verifiable Credentials(以下 VC) [※7] により検証可能にすることで、AI 主体のデジタル取引における信頼性向上をめざします。これにより、サービス提供者が AgentCard をレジストリに登録する際、発行元の正当性を保証し、なりすましや改ざんの検知を可能にする仕組みを実装します。

管理の対象となる属性情報は、アイデンティティ情報(AI の開発者・運用主体、運用環境の所在国、データの流用ポリシーなど)から、実行権限の正当性である決済やデータアクセスなどまで多岐にわたります。これらを取引時に検証可能にすることで、AI エージェント同士が互いの信頼性を確認し合い、自律的かつ安全に経済活動を完結できるデジタル社会基盤の構築をめざします。



＜AI エージェント属性情報レジストリ(仮称)のプロトタイプ構成＞

3. 今後の展開

今後は、開発したプロトタイプをデジタル ID ウォレット[※10]や、VC の発行・検証機能と連携させ、より実用的なトラスト基盤へと拡張していきます。併せて、お客さまやパートナー企業、国内外の関連団体との共同実験を企画・実施し、実際のビジネスシーンにおける運用性や有効性の検証を加速させます。

そして、本プロトタイプの信頼性をさらに強固なものとするため、NTT 株式会社が培ってきた高度なデジタルトラストおよび AI 関連技術(マルチパーティ選択的開示技術[※11]、ウォレットの分散鍵管理技術[※12]、水平連合学習技術[※13])を順次取り込んでいく予定です。

NTT ドコモビジネスは、本検証を通じて、国際的に相互運用可能なトラスト基盤の整備をめざします。個人、法人、IoT デバイス、AI といった多種多様な主体が、国や組織の垣根を越えて互いの信頼性を担保し

合えるデジタルトラストの仕組みを通信事業者として提供することで、経済安全保障をデジタルの側面から支援するとともに、誰もが安全につながる高度なデジタル社会の実現に貢献していきます。

※1 AI エージェントとは、業務の目標を理解した AI がさまざまなツールを自動で使い分けながら、タスクに取り組む自律型ソフトウェアです。
※2 マルチ AI エージェントとは、異なる専門領域や役割を持つ複数の AI エージェントが、自律的に相互連携・協調することで、単一の AI では対応が困難な複雑な課題や大規模な業務フローを解決する仕組みです。
※3 Agent2Agent Protocol (A2A) とは、自律型 AI エージェント同士が、互いに情報の受け渡しやタスクの連携を安全かつ円滑に行うために策定された通信プロトコルです。
※4 Open Worldwide Application Security Project (OWASP)とは、Web アプリケーションのセキュリティ向上を目的とした国際的な非営利組織です。
※5 https://genai.owasp.org/llm-top-10/ (大規模言語モデル特有のセキュリティ上の脅威ワースト 10 を定義)
※6 AgentCard とは、利用者が AI エージェントの信頼性を客観的に判断できるよう、その機能・制約・安全性に関する情報を標準化された形式で明文化したものです。
※7 Verifiable Credentials (VC) とは、発行元による電子署名が付与され、改ざんが困難であるとともに、オンライン上でその妥当性を即座に検証できるデジタル証明書の標準規格です。
※8 Large Language Model (LLM) とは、大量のテキストデータから学習し、人間のように高度な自然言語の理解や生成を可能にする AI モデル(大規模言語モデル)です。
※9 Model Context Protocol (MCP) とは、AI モデルと外部のツールやデータソースを標準的な方式で安全かつ容易に接続し、AI エージェントの機能を拡張するための通信プロトコルです。
※10 デジタル ID ウォレットとは、個人の ID や資格情報 (VC) を安全に保管し、ユーザー自身が提示する情報の範囲をコントロールしながら管理・利用するためのアプリケーションです。
※11 マルチパーティ選択的開示技術とは、複数の発行者が発行した異なるデジタル証明書 (VC) を組み合わせ、提示先が必要とする特定の項目のみを、他のプライバシー情報を秘匿したまま正当に証明・開示する技術です。
※12 ウォレットの分散鍵管理技術とは、秘密計算や秘密分散技術を用いて、デジタルウォレットの秘密鍵を複数のサーバーなどへ分散して管理することで、鍵の紛失時の復旧と第三者による不正利用の防止を両立させる技術です。
※13 水平連合学習技術とは、データ項目が共通する複数の組織間で、生データを外部に持ち出すことなく、各拠点で学習させたモデルのパラメータのみを統合することで、プライバシーを保護しながら共同で AI モデルを構築する技術です。
