つなごう。驚きを。幸せを。
docomo Business

AIエージェントが“企業の最大のセキュリティリスク”になる日

リスクを“見える化”し、制御するための最新アプローチ

NTTドコモビジネス株式会社
城 征司氏

CyberArk Software株式会社
蓑輪 尚毅氏

パロアルトネットワークス株式会社
上野 智弘氏

AI技術の発展により、企業におけるAIエージェントの活用が進んでいるが、その一方でAIの利用を巡るサイバーセキュリティリスクが問題視されている。

2026年3月17日、NTTドコモビジネス主催のオフラインセミナー「AIエージェントが企業最大のセキュリティリスクになる日」を開催。サイバーセキュリティ分野における3名の専門家が集結し、それぞれの視点からAIエージェントのセキュリティリスクを防ぐ対策について講演を行った。その内容を紹介する。

NTTドコモビジネス株式会社　城 征司氏

AIが社員になる日、セキュリティはどう変わるか

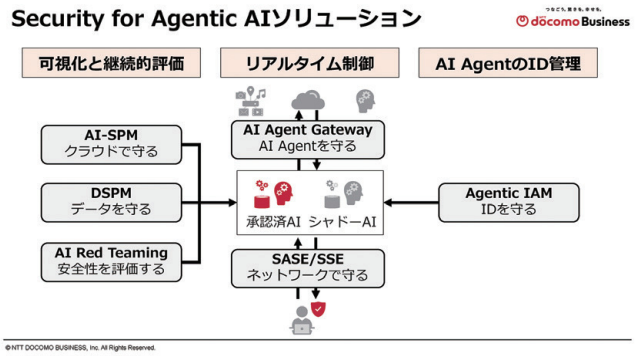
まず最初に講演を行ったのはNTTドコモビジネスのエバンジェリストを務める城氏、AI社員の拡大によるセキュリティの今後の在り方について解説した。

AIは受動的な存在から、人間と協働する存在へと進化、今後は人間のように、アプリ操作やメール送信、会議参加などを自律的に行い数千、数万のAI社員が、産業のフロントラインへと存在意義を高めていくという。

また、AI社員の普及により、社会経済には3つの変化が起こると説明。
1つ目は、専門知識や高度なスキルがAIを通じて誰でも利用可能になるスキルの民主化。2つ目は、24時間365日稼働するAIにより、リアルタイム経営の実現。3つ目は、定型業務をAIに任せ、人間は構想と責任に集中するという役割の変化が起こる。

一方でこうした進化にはリスクもあると指摘する。
生成AIは確率論にもとづいて動作し回答が毎回揺らぐため、従来のITのように完全な制御が難しく、ハルシネーションやプロンプトインジェクション、学習データの偏りなどによる誤判断といった問題が発生する可能性がある。これらが積み重なることで、AIの自律性が暴走し、社会経済を崩壊に導く懸念がある。だからこそ「このAIは信頼できる」と人々が納得できる状態であるAIのトラストワージネスが重要だと語り、それを実現するための3つのテクノロジーを紹介。

1つ目は社員のシャドーAIの利用状況やAIの脆弱性をリアルタイムに可視化・継続評価を行う仕組み。2つ目はAIエージェントに対する適切なID管理、必要最小限の権限付与・はく奪する仕組み。3つ目は文脈にもとづくリスク可視化とリアルタイム制御によりAIエージェント環境のセキュリティレベルを継続的に向上する。
NTTドコモビジネスではこれらの3つのテクノロジーを包含して一体提供する「Security for AI Agentic AIソリューション」でお客さまのトラストワージネスの実現をサポートすると説明した。



CyberArk Software株式会社　蓑輪 尚毅氏

アイデンティティを起点に守る、AIセキュリティ最前線

続いてCyberArk Software社の蓑輪氏が講演し、アイデンティティセキュリティの観点からAI時代のセキュリティのあり方について解説がなされた。
従来アイデンティティは“ヒト”とヒト以外の“マシン”とで分類されていたが、AIの登場により両者の特性を併せ持つ新たな主体として捉える必要があると指摘する。

アイデンティティセキュリティを実現するためには、社員を信用し過剰な権限を付与するのではなく、システムで制御し正しいことのみができる環境にする必要があるという。

また、何を守るのか、守らなければいけないものがどこにあるか、全体像を正しく把握して初めて効果的なアプローチが行えるようになると説明。

続いてAIエージェントを安全に活用するためのセキュリティの考え方として、4つの観点が提示された。
1つ目は、組織内で利用されているAIエージェントの数や役割、利用者を把握し、「何が、何をしているのか」を明確にすること。2つ目は、AIが社内のデータベースや機密情報へアクセスする際の「セキュアなアクセス」を制御すること。3つ目は、AIの不審な挙動を検知し、アラートやキルスイッチによってリスクを抑止するといった「脅威検知と対応」の仕組みづくり。4つ目は、AIエージェントの権限が過剰でないかを定期的に見直し、何をしているのかを適切に管理する「統制とライフサイクル管理」。

CyberArkではこの4つの観点でセキュアなAIを活用するためのサービス展開を進めていることを紹介。
最後に、セキュリティ対策は、守る対象やリスクに応じて必要な対策は全く異なるため優先順位を付け、段階的に導入していく必要がある。
人・マシン・AIすべてに共通して、最小の特権、適切な認証、認証情報管理、アクセス制御、リスク検知と対応、ガバナンスと統制これらを統合的に考える必要があると強調した。

パロアルトネットワークス株式会社　上野 智弘氏

AIエージェントのリスクを未然に防ぐ包括的なAIセキュリティプラットフォームによる対策

最後にパロアルトネットワークス社の上野氏が講演し、パロアルトネットワークス社の製品「Prisma AIRS」を軸にAIエージェントのリスクとその対策について解説した。

初めにOWASP Top 10 for LLM Appsをもとに、AIアプリケーションにおける具体的なリスク事例を解説し、特に注意が必要なリスクとされるプロンプトインジェクションをサンプルに攻撃の指示や流れをもとに対策のポイントを紹介し、プロンプトインジェクションで守るべきはAI単体ではなく、その背後にあるコンポーネントやシステム全体であることをきちんと認識する必要性があると強調。

また、オープンソースのAIモデルでは攻撃者によってマルウェアや、悪意のあるスクリプトを仕込まれる可能性があり、汚染されたAIモデルを使うことによる、情報漏えいや、他のAIアプリケーションのトレーニングデータ汚染などにも注意も促した。

続いて、そのリスクへの対応について、自社製品「Prisma AIRS」を例に3つの対策を紹介。

まず最初に行うことは、AIモデルやデータ、ツールなどAIアプリケーションに関するすべてのコンポーネントの情報を把握し可視化すること。「Prisma AIRS」では具体的にどのような宛先と通信しているか、どういったユーザーが利用しているかなど、全体像を可視化し「AIエコシステム全体を発見」。次に、AIアプリケーションに対するリスクを評価すること。「AIモデルスキャン機能」では、悪意のあるスクリプトが仕込まれていないかをチェックし、問題がある場合はそのAIモデルのデプロイをブロックすることが可能。これに加え、AIエージェントに対して自動で擬似攻撃を行う「AIレッドチーミング」により、脆弱性を事前に評価する。次に、実際のプロンプトインジェクションからAIを守ること。

「AIランタイムセキュリティ機能」では、ネットワーク型のアプライアンスによる検査と、SDKを組み込むAPIインターセプト方式の2つのアプローチにより、保護の精度を高め、期待通りの制御を行う。

AIアプリケーション特有のリスクから守るために必要な機能を、包括的に搭載した製品が「Prisma AIRS」であると説明した。

城氏×蓑輪氏×上野氏

クロストークセッション

クロストークセッションでは、講演を行った3氏が「AI利活用とセキュリティ対策の乖離」「シャドーAIの脅威」「AIエージェントに対するIDセキュリティの重要性」の3テーマについてトークセッションを展開した。

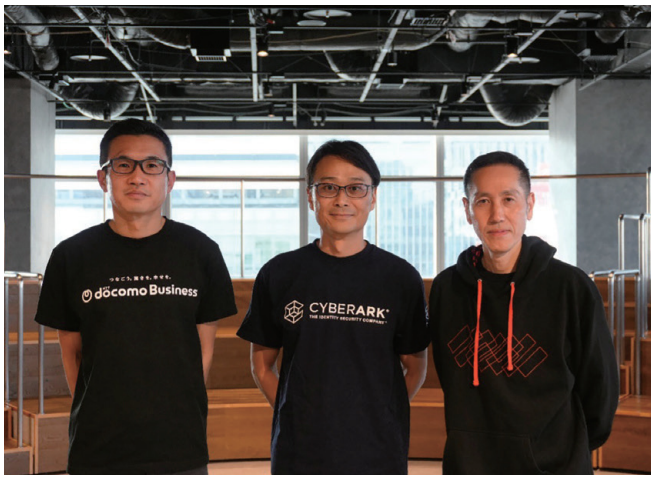
AIのセキュリティ対策の現状について、考えてはいるが、具体策が分からず対策までは進んでいない企業が多いと話す。またAIの活用について、利用を推進する側と、セキュリティを重視する側では考え方が異なると説明。

しかし、セキュリティを重視しすぎるとDXが進まないため、意識せずにセキュリティを担保できる設計と、不自由さを感じない程度の制限をかけるなど、バランスをとることで利活用の促進につながると語る。

シャドーAIの実態については、どれだけ利用を制限してもルールだけでは抑えきれないと全員が指摘。ポリシーに沿った運用は前提ではあるものの、誰が何を使っているかなど、ガバナンスを効かせ、適切に利用をコントロールすることで、社員がセキュアに使える環境を提供することが理想だと話す。

AIエージェントに対するIDセキュリティのテーマでは、CyberArkの特権アクセス管理の意義について触れた。蓑輪氏の講演でも紹介したように、入口で正しい人はすべて正しいことをするとは限らないため、常に行動を監視する必要がある。認証後の行動を監視し、制御することができるのがCyberArkの特権アクセス管理の強みであると強調した。

本講演には、さまざまな所属組織の参加者が集まり、AIセキュリティへの関心の高さが伺えた。NTTドコモビジネスでは、リーディングパートナーであるCyberArk Software社、パロアルトネットワークス社と共に「Security for Agentic AI」構想を実現し、AIによるDXとガバナンスの両立を目指す企業のサポートしていく考えだ。



お問い合わせ
NTTドコモビジネス株式会社
サイト https://www.ntt.com/index.html