knowbe4

日本版

# 調査レポート：エージェンティックAIのリスクを「人間の強み」へ

## ハイライト

- **ガバナンスなきAIの実態を知る：**シャドーAIの広がりと、正式なガバナンスなしにAI活用が進む日本企業の実態を明らかにする
- **AIエージェントの自律稼働がもたらすリスクを読む：**自律的に行動するAIエージェントを導入している組織の割合と、ガバナンスが追いついていない実態を明らかにする
- **ディープフェイクと心理的脅威の最前線を見る：**脅威の手口がAIを活用した精巧な心理操作へと進化する中、日本の従業員がどれほどその脅威にさらされているかを示す
- **「報告しにくい文化」が生む、見えないリスクを特定する：**ミスを安心して報告できない組織文化が、日本のサイバーセキュリティにどれほど深刻な影響を与えているかを示す
- **自社のセキュリティ成熟度を評価する：**セキュリティ成熟度の「ゴールデンスタンダード（最高基準）」とは何かを理解し、日本の組織が新たなAI主導の脅威に対応できているのかを確認する

## はじめに

日本のビジネス界において、理論的なAIの実証実験（実用性の検証）の段階はすでに過去のものである。現在は、自律型エージェントや機械的なスピードで進む業務プロセス（ワークフロー）が、実際のビジネス業務を自律的に実行している。

しかし、多くの組織が効率性の向上を急ぐあまり、セキュリティ体制にほころびが生じつつある。KnowBe4の調査レポート『エージェンティックAIのリスクを「人間の強み」へ』（From Agentic Risk to Human Wins）は、日本の組織が今まさに苦悩している「重大な盲点」「心理的な脆弱性」そして「受け入れ態勢の不備（準備不足）」を浮き彫りにしている。

本レポートは、13の国・地域のIT・セキュリティ部門に携わる役員・管理職（セキュリティリーダー）800名と、セキュリティ業務を担当しない一般従業員3,200名の計4,000名を対象に実施したグローバル調査をもとに、日本在勤のセキュリティリーダー75名、一般従業員300名のデータを抽出、分析したもの。

## Section 1: シャドーAIとガバナンスの欠如

日本の組織が公式なIT環境へのAI導入を進める一方で、その裏では、管理者へ報告のない「シャドーAI」が、従業員の間で急速に広がっており、管理の目が届かない大きなレイヤーとなっている。脅威アクター（サイバー犯罪者）は、「見えないものを守ることはできない」というセキュリティチームの弱みを理解している。

- 日本のセキュリティリーダーの79%が、自律的に行動するAIエージェントをすでに業務プロセスに組み入れていると回答しており、 これは世界平均（58%）を大きく上回り世界最多の水準だ。そのうち34%は人間の監視が限定的なままAIが自律稼働しており、この割合も世界平均（17%）の2倍で世界最多。セキュリティリーダーの40%（世界平均：37%）は、「ガバナンスが限定的または不明確」と認めており、さらに11%（世界平均：14%）は、「正式な承認なしでAIが使われている」と回答している。
- 日本のセキュリティリーダーの44%（世界平均：51%）が、過去12か月間に自社のサイバーセキュリティに最も大きな影響を与えた人的な問題として、「未承認の外部ソフトウェアやシャドーAIツールの使用」を挙げている。
- 日本の従業員の28%（世界平均：34%）が、組織が公式に提供・承認したAIツール以外のツールを業務で使用していると回答している。

## Section 2: ディープフェイクと忍び寄る心理的脅威

脅威の手口は、従来のフィッシング詐欺から、AIによって精巧に仕組まれた「きわめてリアルな誘導・心理操作」へと根本的に変化している。もはや、AIが生成したディープフェイクなどのメディア（音声や映像、テキスト）に対して、人間の認知的な防御力は通用しなくなっている。

- 日本の従業員の89%（世界平均：86%）が「ディープフェイクの音声・動画コンテンツは、本物と区別がつかないレベルに達しており、何を信用すべきか判断が難しくなっている」と回答している。
- 日本の従業員の71%が「社内の関係者や経営幹部になりすましたディープフェイク詐欺に実際に騙されてしまう可能性がある」と認めている。この割合は世界平均（64%）を上回るほか、米州（南北アメリカ）の60%、EMEA（欧州・中東・アフリカ）の61%をも超えており、この指標において日本はより脆弱だと言える。

- フィッシングメールは依然として最大のリスク経路だが、脅威はメールにとどまらない。日本の従業員の63%（世界平均：59%）がフィッシングやなりすましメールを人的サイバーリスクの主要な原因として挙げる一方、TeamsやSlackなどのコラボレーションツールを経由した攻撃を主要リスクとして挙げたリーダーも27%（世界平均：30%）にのぼる。さらに、コラボレーションツール経由のなりすましを見抜く自信が「ない」と答えた従業員は40%（世界平均：19%）に達しており、攻撃経路の多様化に対して従業員の対応力が追いついていない実態が浮かぶ。

## Section 3:「ミスを安心して報告できない」――日本に潜むセキュリティの危機

日本における本質的な脅威は、攻撃の巧妙さやAI活用のスピードだけではない。それは、従業員がミスを安心して報告できない組織文化が深く根付いているという現実であり、処罰への恐れがリスクを組織の内側に封じ込めてしまうという構造的な問題だ。

- 「ミスを犯した際、責められる恐れなく安心して報告できる」と強く感じている日本の従業員はわずか21%。これは13の国・地域の中で最も低い割合であり、世界平均（43%）の半分以下である。米州（南北アメリカ）では54%、EMEA（欧州・中東・アフリカ）では42%の従業員が安心して報告できると感じており、同じAPJ地域内のオーストラリア&ニュージーランドも43%に達している。ミスが適切に報告されないと、原因の特定も再発防止も困難になる。
- この「ミスを報告しにくい環境」がもたらすリスクは、セキュリティリーダーと従業員との間でインシデント後の対応に対する認識が大きく異なるという事実によって、さらに増幅される。日本のセキュリティリーダーの60%が「偶発的なセキュリティミスに対し、学習・改善支援型の対応をしている」と回答しており、この割合は世界平均（49%）を上回り13の国・地域の中で最多。しかし従業員への調査では、同様の対応を「実感している」と答えたのはわずか 32%（世界平均：39%）にとどまる。この28ポイントの認識ギャップは世界で最大であり、米州（南北アメリカ）の 6 ポイント、EMEA（欧州・中東・アフリカ）の9ポイントと比べても際立っている。また従業員への調査では、偶発的なインシデントの後に「懲戒処分を受けた」と答えた割合が13%（世界平均：11%）にのぼるのに対し、セキュリティリーダーへの調査での実施率はわずか3%（世界平均8%）。セキュリティリーダーが認識・意図していない形で、従業員が処罰的な対応を経験している可能性が高く、それがさらなる報告抑制につながっていると考えられる。
- 人間とAIエージェント双方のリスクを管理できる組織になるには、強いリーダーシップと明確な責任体制のもとで、セキュリティへの「意識」「行動」「文化」が組織に深く根付いた状態――「ゴールデンスタンダード（最高基準）」を実現する必要がある。しかし、日本でそのレベルに達している組織はわずか 8%にすぎず、世界平均（19%）の半分以下で13の国・地域の中で最低水準となっている。人間の監視が限定的なままAI エージェントが自律稼働する割合は世界最多でありながら、いざ問題が起きたときに対処できる組織の成熟度が最も低い――これが日本の現実だ。

## 結論

日本の労働環境は変化した。今や従業員とAIエージェントは、単なる「人とツール」という関係を超えて現場の新たな労働力として深く結びつき、ひとつの強固な防御層として機能することになる。これにより、「優れたセキュリティ」の定義そのものが変わる。自社の新たな労働環境のリスクを明確に捉え、それをもとに「セキュリティ意識の向上から行動の変容、そしてそれらが定着するセキュリティ文化の醸成まで」を一連の流れとして捉えて取り組む組織は、単にリスクを管理する以上の成果を上げている。彼らは、自らを極めて「強固な標的（攻めにくい組織）」へと変貌させている。脅威が消え去ることはない。しかし、高度な備えを固めた組織の壁を打ち破ることは、攻撃者にとって極めて困難なものとなるのである。

労働環境が劇的に変化し、自律的に動く「エージェンティックAI」が人の能力の拡張、あるいは「新たな同僚（デジタルな労働力）」となりつつある今、シャドーAIの使用は、もはや「従業員が会社の許可なく、勝手に外部の人間を職場に招き入れて業務を任せている」のと同義である。現場にとってAIは、時に不条理な上司や他部門のスタッフ、あるいは指示待ちの部下よりも、遥かに有能で頼りになる存在かもしれない。だからこそ、現場は無断であってもAIを頼ってしまうのである。

これまでのサイバーセキュリティは、「やってはいけないこと」を明確にし、違反に対する抑制と失敗への「処罰」を行う、いわば性悪説的なアプローチが当たり前とされてきた。しかし、心理的安全性やミスの報告しやすさが世界最低水準である日本において、この罰則型アプローチを続けていれば、現場はAIの不具合やセキュリティリスクを徹底的に隠蔽するようになるだろう。

今後は、単なる管理や禁止ではなく、現場と上司の密な対話、現場のリアルなニーズに対するIT部門からの積極的なヒアリング、そしてダイバーシティを踏まえた自由な議論の促進など、「物事を本質的に良くするための前向きな取り組み」へと舵を切るべきである。私たちは今、AIという「デジタルなワークフォース（労働現場）」を組織としていかにガバナンスし、そしてそれらを扱う現場の人間をいかに信頼し対話していくかという、新たなステージの議論に向き合うべき局面に来ている。

## 調査概要

本調査は、米州（南北アメリカ地域）、EMEA（欧州・中東・アフリカ）、およびAPJ（アジア太平洋・日本）地域の13の国・地域における、800名のIT・セキュリティ部門に携わる役員・管理職（セキュリティリーダー）と3,200名のセキュリティ業務を担当しない一般従業員の計4,000名を対象に実施されたグローバル調査に基づいています。本レポートで引用しているデータは、そのうち日本在勤のセキュリティリーダー75名と一般従業員300名の回答を抽出・集計したものです。

回答者は従業員数250名以上の組織に所属しており、情報技術（IT）、医療・ヘルスケア、消費者サービスなど、幅広い業界の民間企業および公共部門（公的機関）を網羅しています。

KnowBe4, Inc. | Midtown Tower 18F, 9-7-1 Akasaka, Minato-ku, Tokyo, Japan
www.KnowBe4.com | Sales@KnowBe4.com