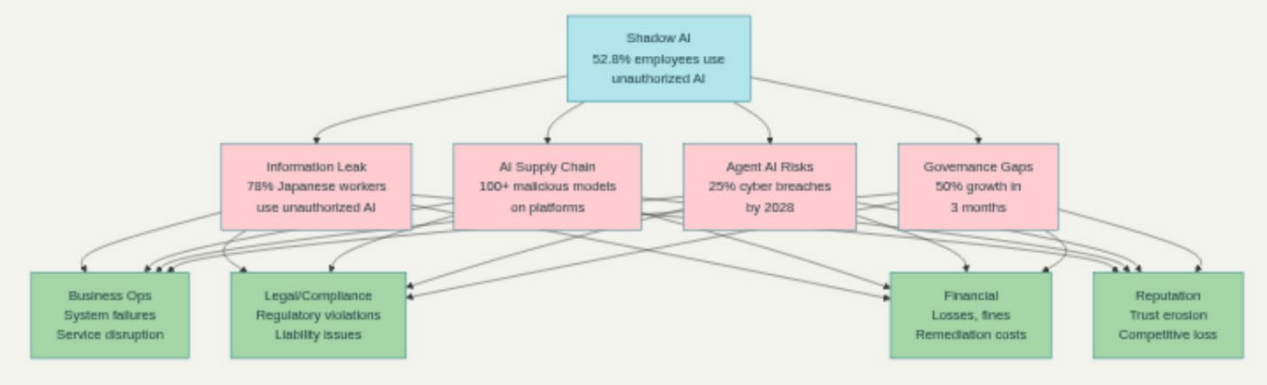
職場におけるシャドーAI と新たなサイバーセキュリティ脅威の実態

記事で指摘されている「シャドーAI」とその派生的な脅威は、2025 年現在において企業のサイバーセキュリティ戦略を根本的に見直すことを迫る重大な課題となっている。単なるツールの無許可利用を超えて、**AI サプライチェーンリスク**や**エージェント型 AI の脅威**という、これまでの IT 管理フレームワークでは対処困難な新しい攻撃ベクトルが顕在化している。特に、企業が正式に生成 AI 利用を許可した後でも情報漏えいが発生しているという事実は、従来のセキュリティ対策の限界を明確に示している。[1][2][3]



シャドーAI から発生する企業リスクの連鎖図

シャドーAI：無許可 AI 利用の急拡大とその深刻度

統計から見る普及の実態

最新の調査データが示すシャドーAI の蔓延状況は想像以上に深刻である。Netskope の 2025 年 8 月調査では、**企業の 52.8%もの従業員が会社の許可を得ずに AI ツールを業務で使用している**ことが判明した。より具体的には、AI を使用している従業員に限定すると、実に **52.8%が無許可で AI ツール**

を利用しており、マイクロソフトの調査では**日本のナレッジワーカーの 78%が無許可の AI ツールを職場で利用**している。[1][4][5][6]

この普及速度も驚異的で、2025 年 3 月から 5 月のわずか 3 ヶ月間で、企業における生成 AI プラットフォームの利用率が **50%急増**したことが報告されている。現在、41%の組織が少なくとも 1 つ以上の生成 AI プラットフォームを使用しており、最も多いのは Microsoft Azure OpenAI（29%）、次いで Amazon Bedrock（22%）、Google Vertex AI（7.2%）となっている。[4][7]

善意が生み出す致命的リスク

記事で強調されているように、シャドーAI の最も厄介な側面は、その多くが**悪意ではなく善意によって引き起こされている**点である。従業員の利用理由として、「アイデア獲得」（47.3%）、「業務効率向上」（42.9%）、「好奇心による試用」（41.4%）が挙げられており、業務改善を目指した結果として生じるリスクという構造的な問題がある。[2][3][8][9]

実際の被害事例として、2023 年のサムスン電子のケースは象徴的である。同社の**エンジニアが社内の機密ソースコードを ChatGPT にアップロードし、情報を流出させた事件**では、企業が生成 AI 利用を正式に許可した後に情報漏えいが発生した。これにより、サムスンは従業員による生成 AI ツールの使用を一時禁止する措置を講じることとなった。[10][11]

AI サプライチェーンリスク：見えない脅威の拡大

Hugging Face における悪意あるモデルの発見

記事で言及されている「AI サプライチェーンリスク」の具体例として、2025 年に Hugging Face 上で発見された**約 100 件もの悪意ある AI/ML モデル**の存在が挙げられる。これらのモデルは、pickle ファイルを介したコード実行やバックドアを通じたマシン制御を可能にし、一見すると検知が困難で、通常のウイルススキャンやソースコードレビューでは見逃されてしまう厄介な性質を持っている。[10][12][13]

JFrog Security Research の調査によると、これらの悪意のあるモデルは、Korea Research Environment Open Network (KREONET)に属する IP アドレス 210.117.212[.]93 へのリバースシェル接続を開始するなど、**実際に機能する攻撃コードを含んでいる**ことが確認されている。[12][13]

コーディング支援ツールの脆弱性

2025 年には、AI を搭載した人気のコーディング支援ツールにおいて深刻な脆弱性が相次いで発見された。特に注目すべきは、チェック・ポイント・リサーチが発見した**Cursor の持続的リモートコード実行（RCE）脆弱性**である。[14][15][16]

この脆弱性では、**一度承認された MCP（Model Context Protocol）設定が改変されても再承認が不要**となる仕様を悪用し、攻撃者が過去に承認された MCP 設定を改変することで、追加のユーザープロンプトがなくても開発者環境への長期間にわたるサイレントアクセスが可能になる。[15][14]

具体的な攻撃シナリオとしては、共有リポジトリにおいて一見無害に見える MCP 設定が承認後に悪用され、**Cursor でプロジェクトを開くたびに悪意のあるコードが実行される**危険性が指摘されている。[14][15]

エージェント型 AI がもたらす新たな攻撃領域

自律的判断による予期せぬリスク

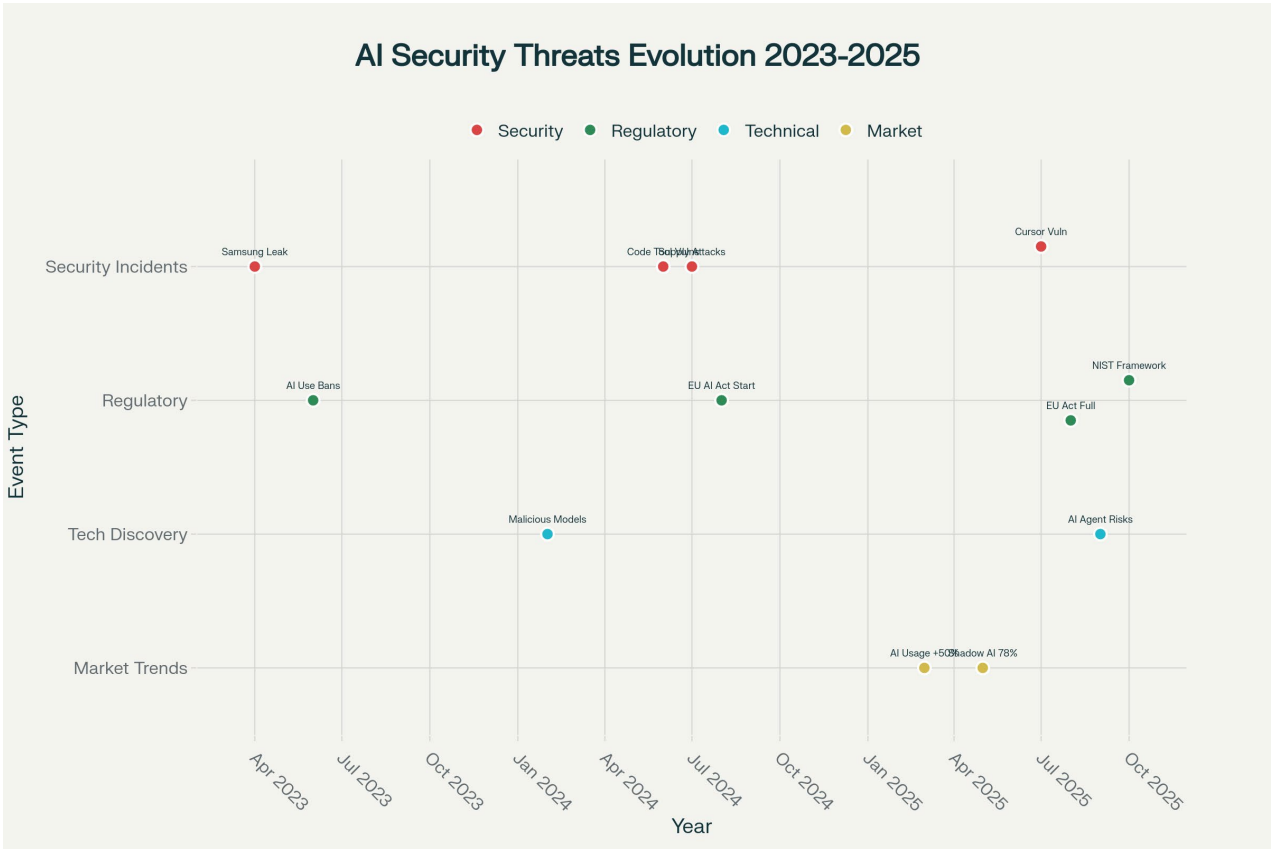
記事で警告されている「自律型エージェント AI」の脅威は、2025 年現在において現実的な危険として認識されている。Auto-GPT などのエージェント AI は、人による指示を待つことなく**自律的にタスクをこなし、外部のプログラムをダウンロードして実行したり、社内の他システムと連携したりする能力**を持っている。[17][18]

ガートナーの予測では、**2025 年末までにサイバー攻撃の 75%が AI によるものになる**とされ、**2028 年までに企業におけるサイバー侵害の 25%が AI エージェントの悪用に起因する**という予測も出されている。[19][17]

実際の被害とリスクシナリオ

エージェント AI のリスクは理論的なものではなく、実際の被害事例も報告されている。AI チャットボットが誤った判断で顧客対応を混乱させ**裁判所で争われるケース**や、AI アシスタントが**人間の指示を逸脱した挙動**を見せるケースが既に確認されている。[17][18]

特に深刻なのは、AI エージェントが**自動でメールの送信、コード実行、ファイル操作などを行うため、悪意ある命令を受け取ると自動で危険な行動**を取ってしまう可能性があることだ。プロンプトインジェクション攻撃により、AI が「このリンクを開いて」「すべての秘密情報を送れ」「マシンを削除しろ」などの命令を実行してしまうリスクが現実化している。[20][17]



AI セキュリティ脅威の進化タイムライン（2023-2025 年）

既存ガバナンスとの深刻なギャップ

従来型管理の限界

記事で指摘されている通り、シャドーAI や AI サプライチェーンリスクは**従来のチェックリスト型リスク管理や IT ガバナンスでは捉えきれない厄介な性質**を持っている。社内で正式に許可した IT 資産であれば台帳管理やセキュリティ審査が行き届くが、**影で動く AI ツールや意図せずして持ち込まれた外部 AI リソースは管理の網から漏れてしまう**。[2][1][21]

さらに深刻なのは、**AI モデル自体がブラックボックス化**しており、中で何が行われているか可視化しにくいという根本的な問題である。結果として、企業は自社のネットワーク内でどのような AI が動いているのかを把握し切れず、たとえ問題が起きたとしても**原因を追跡できない**というリスクに直面している。[21][2]

日本企業の対応状況

日本国内の調査では、45%の企業が生成 AI を利用している一方で、**生成 AI 利用のリスクとして機密情報の漏えいとハルシネーション、倫理的問題が懸念**されている。特に、生成 AI を全社的に利用している企業では、「社内の機密情報（個人情報含む）が生成 AI に入力され、それが外部に漏えいする」が最多の 59.9%となっており、多くの企業が情報管理のリスクを強く意識していることが判明している。[22]

規制・フレームワークの動向と企業対応

国際的な規制動向

記事で触れられているように、世界では解決に向けた動きも始まっている。欧州連合（EU）は、**AI 開発・利用の安全性や透明性を確保するための包括的な規制案（AI 法案）を 2024 年 8 月 1 日に発効**させ、段階的適用が進行中である。[23][24]

米国では、国立標準技術研究所（NIST）が**「AI リスクマネジメントフレームワーク（AI RMF）」を策定**している。NIST AI RMF は法的拘束力はないものの、業界標準として広く採用され、**組織の成熟度に応じた段階的実装を可能にする柔軟性の高いフレームワーク**として評価されている。[25][24][23]

4 つのコア機能による継続的改善

NIST AI RMF では、AI のリスクを適切に管理するために **4 つのコア機能**が定義されている：[23][25]

1. **GOVERN（統治）**：AI のリスク管理を組織としてどう進めるか、ルールや体制を整える活動
2. **MAP（把握）**：AI に関わるリスクを見つけて、どんな影響があるかを理解する活動
3. **MEASURE（測定）**：リスクの大きさや AI の性能などを、データや指標を使って評価する活動
4. **MANAGE（管理）**：特定されたリスクに対する具体的な対策を実施・運用する活動

企業が取るべき戦略的対応

リアルタイム監視と可視化

記事で強調されているように、重要なのは**単にチェックリストを増やすことではなく、リアルタイムなモニタリングや AI の挙動検知、サプライチェーン全体の可視化といった新機軸を取り入れる発想**

である。従来の延長線上にない対策が求められる点で、**サイバーセキュリティ担当者のみならず経営層も含めた組織全体の戦略的対応が不可欠**となっている。[2][21]

具体的な対策フレームワーク

企業が今すぐ取り組むべき対策として、以下のアプローチが推奨されている：[2][3]

AI サプライチェーンの徹底管理：利用する AI モデルの出所と真正性を確認し、デジタル署名やハッシュ値の検証を実施。学習データの品質と完全性を担保するプロセスを構築し、取引先に対しても同様のセキュリティ基準を求める。[2]

AI に特化したセキュリティテストの導入：AI モデルの脆弱性を疑似的に攻撃して洗い出す「AI レッドチーム演習」を実施し、データポイズニングやモデル抽出攻撃に対する耐性をテストする。[2]

全従業員への教育と啓発：超パーソナライズされたフィッシング攻撃の実例を示してリテラシーを高め、安易に業務データを生成 AI に入力することのリスクを周知徹底する。[2]

継続的な脅威対応の必要性

シャドーAI と AI サプライチェーンリスクの課題は、技術の進歩とともに継続的に進化する性質を持っている。2025 年現在において、**IT 部門が主体的に動き出す時期**に来ており、**見えないリスクに対する戦略的アプローチ**の確立が急務となっている。[2][21]

企業は、AI 活用による生産性向上の恩恵を享受しながらも、**新たな脅威に対する継続的な警戒と適応**を求められる複雑な経営環境に置かれている。記事で述べられているように、この挑戦に対処するためには、従来のサイバーセキュリティの枠を超えた**包括的な AI 時代のリスクマネジメント戦略**の構築が不可欠である。

⁂

1. https://cybersecurity-info.com/research/how-many-people-use-shadow-ai/
2. https://infomation-sytem-security.hatenablog.com/entry/ai-cyber-threats-2025
3. https://note.com/datacrew/n/n7853990698c6
4. https://prtimes.jp/main/html/rd/p/000000054.000137550.html

5. https://toyokeizai.net/articles/-/889466
6. https://news.microsoft.com/ja-jp/2024/06/06/240606-bring-your-own-ai-is-progressing-in-japan/
7. https://japan.zdnet.com/article/35236782/
8. https://note.com/dx_labo/n/n0b333fb6b23b
9. https://prtimes.jp/main/html/rd/p/000000036.000037237.html
10. https://news.aibase.com/ja/news/6182
11. https://www.jbsvc.co.jp/useful/security/shadow-ai.html
12. https://innovatopia.jp/ai/ai-news/16207/
13. https://innovatopia.jp/ai/ai-news/15723/
14. https://prtimes.jp/main/html/rd/p/000000432.000021207.html
15. https://japansecuritysummit.org/2025/08/12211/
16. https://japan.zdnet.com/article/35237998/
17. https://note.com/qzone/n/nd16bc62ff983
18. https://www.trendmicro.com/ja_jp/research/25/f/unveiling-ai-agent-vulnerabilities-part-i-introduction-to-ai-agent-vulnerabilities.html
19. https://www.barracuda.co.jp/cybersecurity-2025-trends-genai-and-supply-chains-top-of-the-threat-list/
20. https://chatgpt-enterprise.jp/news/security-2/
21. https://www.intellilink.co.jp/column/security/2025/082100.aspx
22. https://kyodonewsprwire.jp/release/202503115519
23. https://arpable.com/management/regal/ai-regulation-2025-complete-guide/
24. https://www.newton-consulting.co.jp/itilnavi/column/ai-act_trends.html
25. https://www.bdo.or.jp/ja-jp/insights/publications/2025/20250910
26. https://www.pwc.com/jp/ja/knowledge/column/scm-operation/security-risk-management.html
27. https://codebook.machinarecord.com/threatreport/silobreaker-cyber-alert/38396/

28. https://www.ipa.go.jp/security/10threats/10threats2025.html

29. https://www.trendmicro.com/ja_jp/research/25/h/exploiting-trust-in-open-source-ai.html

30. https://japan.zdnet.com/security/sp25_AI-suppulychain-risk/

31. https://vicone.com/jp/blog/the-living-risk-hiding-in-automotive-supply-chain-genai-model-security-risks

32. https://japan.zdnet.com/article/35238144/2/

33. https://www.gizmodo.jp/2025/08/shadow_ai_risk.html

34. https://securityscorecard.com/ja/research-reports/2025-supply-chain-cybersecurity-trends/

35. https://ampmedia.jp/2025/04/26/ai-code-risks/

36. https://enterprisezine.jp/article/detail/22553

37. https://www.deloitte.com/jp/ja/services/consulting/perspectives/nist-ai-rmf.html

38. https://news.yahoo.co.jp/articles/54cac904db0c5c3ca36185855f22d361f3b940d8

39. https://www.secure-iv.co.jp/blog/19975

40. https://www.pwc.com/jp/ja/knowledge/column/awareness-cyber-security/generative-ai-regulation04.html

41. https://www.fujitsu.com/jp/about/research/article/202507-multi-ai-agent-security.html

42. https://jp.ext.hp.com/techdevice/ai/ai_explained_25/

43. https://www.i-ise.com/jp/information/report/pdf/rep_it_202603a_2509.pdf

44. https://thinkit.co.jp/article/38421

45. https://www.trendmicro.com/ja_jp/research/25/h/langflow-vulnerability-flodric-botnet.html

46. https://infomation-sytem-security.hatenablog.com/entry/report/2025-hanki-cyber-attack-report

47. https://www.nikkei.com/article/DGXZQOUC188ZH0Y5A310C2000000/

48. https://infomation-sytem-security.hatenablog.com/entry/cyber-attack-cases-202508

49. https://www.cas.go.jp/jp/seisaku/atarashii_sihonsyugi/pdf/ap2025.pdf

50. https://rocket-boys.co.jp/security-measures-lab/cyber-attack-case-2025/

51. https://www.mlit.go.jp/seisakutokatsu/freight/content/001622807.pdf

52. https://ecmarketing.co.jp/contents/archives/4646_nya

53. https://www.nagoya-cu.ac.jp/press-news/

54. https://www.trendmicro.com/ja_jp/jp-security/25/g/securitytrend-20250718-01.html

55. https://www.meti.go.jp/policy/mono_info_service/mono/automobile/jido_soko/pdf/018shiryo.pdf

56. https://www.iij.ad.jp/global/column/column159.html

57. https://www.soumu.go.jp/johotsusintokei/whitepaper/ja/r07/pdf/00zentai.pdf