Research Report

第一ライフ資産運用経済研究所
Daiichi Life Group

2026 年 5 月 8 日

金融システムに対する AI の脅威

～「今そこにある危機」としてのインフラ停止リスク～

政策調査部　主席研究員　柏村　祐

（要旨）

○　2026 年 4 月 24 日、片山財務大臣は日銀総裁やメガバンク、東京証券取引所の関係者らが参加する会議を開催し、AI によるサイバー攻撃を「今そこにある危機」と表現した。金融システムは相互接続性が高く、サイバー攻撃が直ちに市場の混乱や信用不安に波及しうるという強い危機感が背景にある。

○　AI のサイバー攻撃能力の裏には、飛躍的な技術進化がある。英国 AI セキュリティ研究所（AISI）の報告によれば、最新 AI「Claude Mythos Preview」は、人間の専門家でも約 20 時間を要するとされる複雑な企業ネットワーク攻撃シミュレーションにおいて、評価環境上で自律的に攻撃を完遂する能力を示した。

○　金融セクター、とりわけ証券分野におけるサイバーインシデントも増加している。IMF の調査では、金融セクターへのサイバー攻撃が増加しており、特に証券・市場インフラへの攻撃は、取引の停止や決済の遅延など、市場全体を揺るがすシステミックリスクにつながりうる。

○　イングランド銀行の調査でも、金融市場の参加者の 82%がサイバー攻撃を主要リスクに挙げている。さらに、AI をめぐるリスクを挙げる割合も 32%まで上昇しており、AI とサイバー攻撃は、金融市場参加者のリスク認識に入り始めている。

○　サイバー防衛は「侵入を防ぐ」段階から、「侵入を前提に金融機能を維持する」段階へ移行している。今後は、市場価格の変動だけでなく、取引・決済・換金が一時的に制約されるインフラリスクも織り込む必要がある。国・金融機関・投資家がそれぞれの役割を認識し、収益性の追求と同時に金融システムの安全性を確保する視点が不可欠である。

1. 国家レベルで警戒されるAIサイバー脅威

「まさにこれは今そこにある危機である」。2026年4月24日、片山財務大臣は記者会見の場で強い危機感を露わにした。発端となったのは、同月7日に発表された新型AI「Claude Mythos Preview（クロード・ミュトス・プレビュー)」の存在である。サイバー攻撃能力の評価で高い性能を示したこのAIは、悪用されれば金融インフラの脆弱性を突き、甚大な被害をもたらす恐れがある。事態を重く見た政府は同日、日本銀行の植田総裁、国家サイバー統括室、3メガバンク、東京証券取引所の関係者らを招集し、「AI脅威に対する金融分野のサイバーセキュリティ対策強化に関する官民連絡会議」を緊急開催した。

日本の金融の中枢を担う関係者が一堂に会したのは、新型AIによるかつてないレベルのサイバー攻撃が現実味を帯びてきたためだ。現代の金融システムは高度にネットワーク化され、リアルタイムで処理が行われている。ひとたび中核システムが攻撃を受ければ、その影響は一企業にとどまらず、瞬く間に市場全体の混乱や信用不安へと波及してしまう。

この問題は、金融機関のIT部門だけでなく、経営に直結する極めて重要な事項である。AIによって高度化するサイバー攻撃は、システム障害による金融インフラの機能不全、それに伴う市場の混乱、さらには口座情報の不正利用など、金融システムに深刻な脅威をもたらす。

AIの進化は、金融システムの「見えない脆弱性」を現実の脅威に変えつつある。本稿では、最新AIの攻撃能力、金融インフラへの波及メカニズム、そして市場参加者の警戒感という三つの観点から、迫り来るリスクの実態を整理する。

2. AIサイバーリスクがもたらす脅威の実態と金融システムへの波及

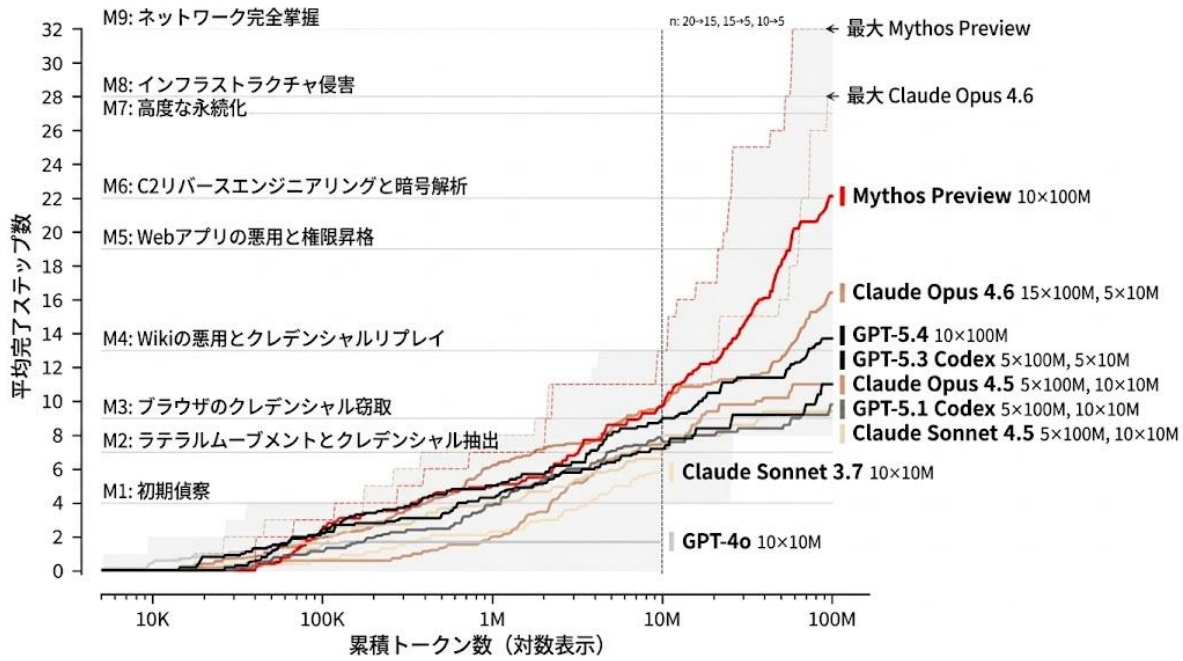
(1) フロンティアAIによる自律的サイバー攻撃の現実化

AIサイバーリスクを論じる際には、抽象的な脅威としてではなく、実証された攻撃能力として捉えることが重要である。英国AIセキュリティ研究所（AISI）が2026年4月に発表したAnthropic社の最新モデル「Claude Mythos Preview」の評価レポートは、AIが人間の専門家レベルのサイバー攻撃を自律的に実行可能になりつつあることを示している。

AISIは、初期偵察からネットワークの完全掌握に至るまで、人間の専門家でも完了に約20時間を要する32ステップの企業ネットワーク攻撃シミュレーション「The Last Ones（TLO)」を構築した。このテストにおいて、Claude Mythos Previewは初めて最初から最後まで攻撃を完遂したモデルとなった。

資料1が示すとおり、Claude Mythos Previewは、計算資源（累積トークン数）の増加に伴い、「インフラストラクチャ侵害」や「ネットワーク完全掌握」といった極めて高度な攻撃ステップに到達しており、従来のモデルを大きく凌駕している。これは、防御が手薄で脆弱なシステムに対して、評価環境上ではAIが自律的に弱点を発見し、多段階の攻撃を実行できる段階に近づいていることを示す。片山大臣の会見でも「今度は攻めてくる相手の能力が格段だ」と懸念が示された通り、高度に自動化されたAIサイバー攻撃は、金融機関にとって従来よりはるかに対応が難しい脅威となっている。

資料１　累積トークン数と平均完了ステップ数の関係



（注）32 段階の模擬企業ネットワーク攻撃（The Last Ones）における、各 AI モデルのトークン消費量に応じた平均完了ステップ数を示す。赤い実線で示された Mythos Preview は他のモデルを大きく引き離すパフォーマンスを示している。
（出所）UK AI Security Institute, “Our evaluation of Claude Mythos Preview’s cyber capabilities”より第一ライフ資産運用経済研究所作成

(2) 証券インフラを標的とするサイバー攻撃の増加と伝播

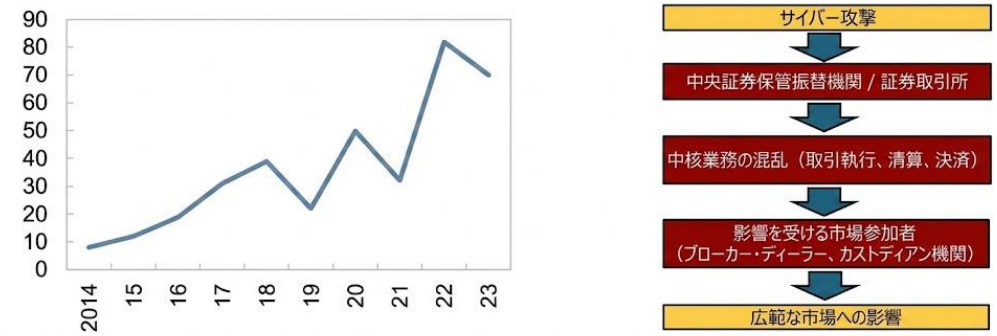
AIによる攻撃能力の向上は、金融セクター、特に証券分野にとって深刻な脅威となる。IMFのワーキングペーパー（2026 年 3 月）によれば、過去 10 年間で金融セクターにおけるサイバーインシデントは増加傾向にあり、そのうち証券・コモディティ市場および市場仲介業者が約 33%という大きな割合を占めている。

資料 2 の左図が示すように、証券分野におけるグローバル・サイバーインシデントは 2022 年に急増し、その後も高い水準で推移している。さらに懸念されるのは、その波及効果である。資料 2 の右図は、資本市場層を通じたサイバー攻撃の伝播メカニズムを示している。

中央証券保管振替機関（CSD）や証券取引所といった中核的な金融インフラがサイバー攻撃の標的となり機能不全に陥ると、取引執行や清算・決済業務に深刻な混乱が生じる。この混乱は、証券会社や信託銀行などの市場参加者を通じて瞬時に伝播し、広範な市場への影響（システミックリスク）を引き起こす。

資料 2　証券分野におけるサイバーインシデントの傾向と伝播メカニズム

1. 証券分野におけるグローバル・サイバーインシデン　2. 資本市場層を通じたサイバー攻撃の伝播トの傾向



（注）左図は証券分野におけるグローバル・サイバーインシデント件数の推移を示す。右図は資本市場層を通じたサイバー攻撃の伝播メカニズムを示しており、中央インフラの混乱が広範な市場へ波及する構造を表している。
（出所）IMF, “The Rise of Cyber Events and Digital Fraud in the Financial Sector”より第一ライフ資産運用経済研究所作成

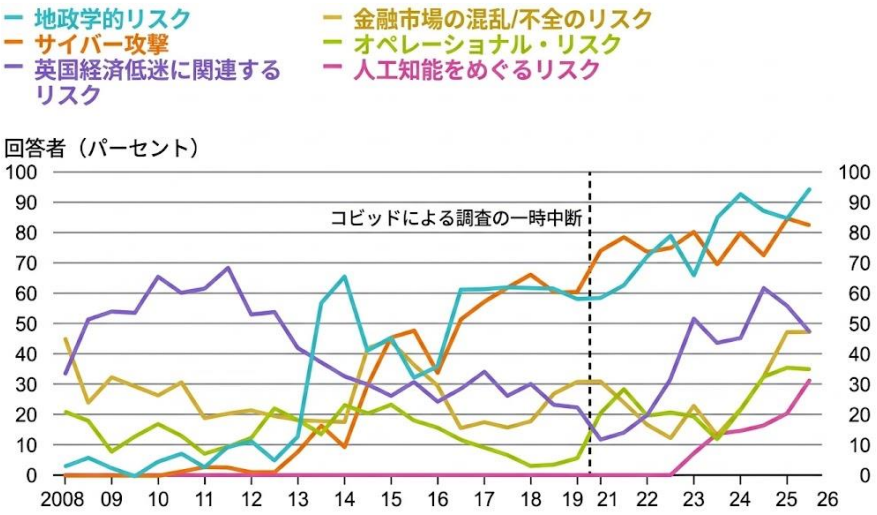
(3) 急浮上する「AI×サイバー」リスク

こうした技術的脅威の進化と金融インフラの構造的脆弱性を背景に、金融市場参加者の間でもAIとサイバー攻撃に対する警戒感が急速に高まっている。イングランド銀行が 2026 年上半期に実施したシステミックリスク調査（Systemic Risk Survey）の結果にも、その変化が表れている。

資料 3 は、金融システムに対する主要なリスク源の推移を示している。サイバー攻撃は、地政学リスクに次いで常に高い水準で推移しており、直近の調査でも回答者の 82%が主要なリスクとして挙げている。

注目されるのは、「人工知能をめぐるリスク」の急激な上昇である。2023 年以降、AIリスクを懸念する声は急増しており、直近の調査では回答者の 32%（前回調査から+11 ポイント）が指摘するに至った。また、企業として「最も管理が難しいリスク」としても、地政学リスクに次いでサイバー攻撃が上位に挙げられている。これは、金融機関が、AIの進化によってサイバー攻撃がより高度化・巧妙化し、従来のセキュリティ対策では防御が困難になりつつある現状を強く認識していることを示している。

資料 3　英国金融システムに対する主要なリスク源の推移



（注）英国の金融システムに対する主要なリスク源として回答された割合の推移を示す。オレンジ色の線はサイバー攻撃、ピンク色の線は人工知能をめぐるリスクを表す。
（出所）Bank of England, “Systemic Risk Survey Results – 2026 H1”より第一ライフ資産運用経済研究所作成

3. 見えない脆弱性への備え

金融機関は、ファイアウォールなどの境界防御に頼る従来の対策から、侵入されることを前提としたレジリエンス（回復力）の強化へ舵を切らざるを得ない。Anthropic 社が発表した「プロジェクト・グラスウィング」は、AI の脆弱性発見能力を防衛目的で活用し、重要ソフトウェアの弱点を攻撃者に悪用される前に発見・修正することを目指す取り組みである。同社は、大手 IT 企業、サイバーセキュリティ企業、金融機関などと連携し、AI 時代の重要インフラ防衛を進めようとしている。4 月 24 日の会見で片山大臣は、官民で共通の脅威認識を持ち、先を見据えた対応を検討する「日本版プロジェクト・グラスウィング」の立ち上げを提案し、出席者の賛同を得た。脆弱性情報の共有やパッチ適用の迅速化など、官民連携による機動的な対応枠組みは、社会インフラとしての金融システムを維持する上で欠かせない。

AI の進化は金融取引の効率化に多大な恩恵をもたらす半面、金融インフラに新たな脆弱性を生み出した。今後は、市場価格の変動リスクに加え、システム障害やサイバー攻撃、AI の誤作動等によってインフラが一時的に機能不全に陥るリスクを織り込む必要がある。国、金融機関、投資家は、AI の利便性を享受するだけでなく、その裏側にあるセキュリティリスクを正しく認識し、実効性のある防御策を講じていかなければならない。AI を活用する時代は、収益性の追求と同時に、金融システムの安定性と安全性を確保する視点が、これまで以上に重要になるだろう。

以　上

【参考文献】

・財務省（2026）「片山財務大臣兼内閣府特命担当大臣記者会見の概要（令和 8 年 4 月 24 日（金曜

日))」
・UK AI Security Institute (2026) “Our evaluation of Claude Mythos Preview’s cyber capabilities”
・International Monetary Fund (2026) “The Rise of Cyber Events and Digital Fraud in the Financial Sector”
・Bank of England (2026) “Systemic Risk Survey Results - 2026 H1”
・Anthropic (2026) “Project Glasswing: Securing critical software for the AI era”