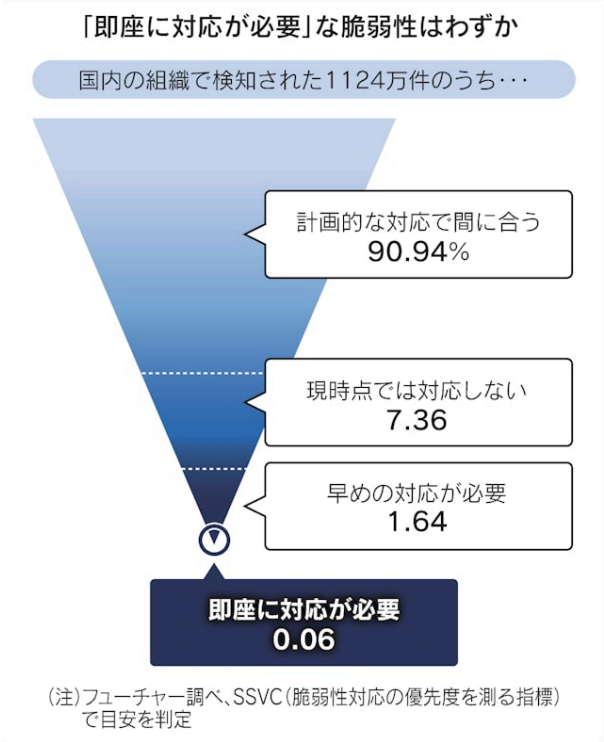
AI「アラート地獄」に落ちるな　ミュトスで見えた新サイバー防衛－編集委員　須藤龍也

2026/05/12 05:00 日本経済新聞電子版　2183文字

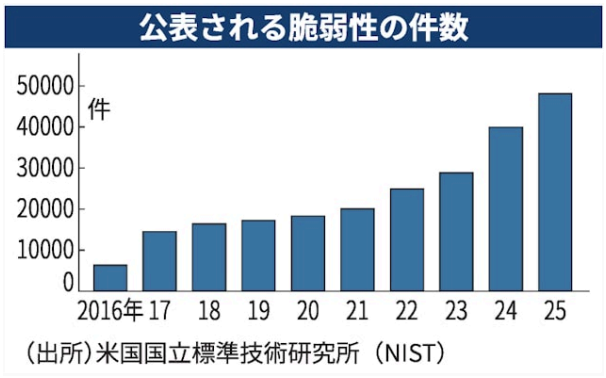
米国アンソロピックの新型ＡＩ（人工知能）モデル「Claude Mythos（クロード・ミュトス）」の登場が、世界に動揺を与えている。ソフトウエアに埋もれた弱点（脆弱性）を見つける能力の高さから、社会システムに甚大な影響を与えるサイバー攻撃リスクへの警戒感が高まる。一方で見えない実態が脅威をあおっている印象も拭えない。限られた時間の中、企業や組織は冷静に受け止め、現実的な守りへの備えを固める必要がある。



サイバー攻撃、守る側もＡＩ必須

4月7日に発表されたミュトスの実力は、断片的な情報にとどまる。アンソロピックのブログによると、数千件に及ぶ深刻度の高い脆弱性を特定し、サイバー攻撃を仕掛ける手順を自律的に作成したとする。影響を考慮した同社は公開先を一部に限定。そこに選ばれた米モジラはウェブブラウザー「Firefox（ファイアフォックス）」について、ミュトスの力を借りて271件の脆弱性を修正した最新版を4月21日に公開した。

AIの進歩はミュトスに限らない。英国の研究機関「AI安全研究所」が4月30日、米オープンAIの新モデル「GPT-5.5」のサイバー能力がミュトスと同等の水準に達したとする調査結果を公表した。アンソロピックのダリオ・アモデイ最高経営責任者（CEO）は5日に開かれた金融業界向けのイベントで、中国のAIモデルが6～12カ月程度でミュトスと同等の能力に追いつく可能性があると示唆した。



AIの日進月歩でサイバー脅威が高まるなか、守りへの活用は必須だ。

大手IT企業に勤めるエンジニアの男性は2025年5月、AIに社内システムの脆弱性を探させたところ、3日で10件見つけた。人の手なら１件あたり少なくとも数週間かかる。「攻撃者は必ずAIを使ってくる。守る側もAIで武装し、攻撃者よりも早く脆弱性を見つけ、対抗するしかない」と話す。

ＡＩに対するサイバー防衛の課題も見えてきた。進化し続けるＡＩが脆弱性を次々と見つけ出すため、対処する人手が足りなくなる問題だ。多くのシステムが稼働する大企業の情報システム部門が、大量の脆弱性の警告に追われて何もできなくなる「アラート地獄」に陥る恐れがある。

サイバー攻撃、組織全体のリスクベースで対応を

ITコンサルティング、フューチャーの神戸（かんべ）康多シニアアーキテクトは、AIが見つけ出す大量の脆弱性を前に、組織が対応できず破綻する未来に危機感を募らせる。「即座に対処しなければならない脆弱性を洗い出し、優先順位をつける。実効性のあるトリアージが不可欠だ」と指摘する。

そこで用いられるのが、米カーネギーメロン大学が考案した「SSVC」と呼ばれる脆弱性対応の優先度を測る指標だ。組織全体のリスクベースで判定するため、あるIT機器で脆弱性が見つかっても、ビジネスへの影響など自社の環境に応じて必要な脅威を見極めることができる。

フューチャーの顧客が所有するIT機器やシステムから2025年に見つかった脆弱性1124万件をSSVCで判定したところ、即時対応しなければならないものは0.06%だった。90%超は定期的なメンテナンス時に対応すればよいと判断。対応する必要がない脆弱性も7%に上った。

脆弱性対策のための指標としては、サイバーセキュリティーの国際団体が管理する「CVSS」が事実上の標準だ。IT機器やシステムの脆弱性の深刻度をスコアリングし、その上で「緊急・重要・警告・注意・なし」と分類する。

SSVCの判定結果を、IT機器やシステムの脆弱性の中身のみで測るCVSSに置き換えると、多くの組織がすぐに対応が必要との目安にする「緊急・重要」が54%に該当した。これではアラート地獄に巻き込まれる可能性が高い。

調査を担当したフューチャーの高橋真哉マネジャーは「（CVSSで該当する）脆弱性が必ずしも組織に大規模な被害を引き起こすわけではない。そのためにもSSVCによる判定は有効だ」と語る。「サイバー核兵器並み」などとミュトスをめぐる臆測が飛び交う今こそ、冷静な判断と対応が求められる。

ＡＩ時代のサイバー対策、時間との勝負

脆弱性の対応を迅速に進めるためには、組織の体制も変わる必要がある。フューチャーによると、SSVCで即時対応が必要と判断されたケースの対応状況を調べると、平均で88日、中央値で45日かかったことが分かった。

AI時代のサイバー攻撃対策は、時間との勝負だ。平均88日の対応では遅い。深刻な脆弱性に一刻も早く対処し、攻撃を受けた前提で次の対策に乗り出す必要がある。

フューチャーの神戸氏は「脆弱性を自動的に修正したり、不具合が起きたら元に戻したりする仕組みを採り入れるなどして即時対応できる体制をつくらなければこの先、システムを守ることができなくなる」と危機感を募らせる。

ただ現実問題として、24時間稼働し続ける大企業の基幹システムや大手通販サイトなどは、修正のためにシステムを簡単に停止させられないのが実情だ。

同社の高橋氏は「24時間365日稼働が当たり前の社会そのものを見直す時期に来ているのではないか」と話す。「ランサムウエア攻撃などで業務が止まることを考えれば、ITシステムを止めてメンテナンス時間をしっかり確保することで、本質的なシステムの可用性が高まり、結果として機会損失が少なくなる」と指摘する。