05

AIイノベーション

生成AIで変容するデジタルクライム

生成AIの台頭により、デジタルクライムが変質している。攻撃側が精巧な偽造によるeKYC突破やフィッシングサイトの高速量産を仕掛ける中、正規仕様の隙を突く攻撃への対策が急務である。サービス設計段階での対応や動的な不正検知、サイバーBCPを連動させた多層防御モデルが必要だ。

生成AIの台頭と「正規仕様」に潜む隙

生成AIの登場により、デジタルクライム（デジタル犯罪）の防衛環境が今まさに転換点を迎えている。ここで示す事例としてのデジタルクライムとは、金融関連サービスが提供する正規の機能・仕様そのものの組み合わせや、仕様の隙を悪用し、不正送金やポイントの不正取得、アカウントの乗っ取りなどを実行する行為のことで、プログラムそのものを改ざん、機能不全に陥らせることを目的としたものではない。

ハッキングに代表されるようなこれまでのサイバー犯罪の多くはプログラムの脆弱性を突く攻撃が主流であったが、現在、これに加えて上記のような正規の機能や仕様の隙をつく手口が増えつつある。（なお、ここでの生成AIはミュトスなどのフロンティアAI登場以前に公開されたAIを悪用したものである。）

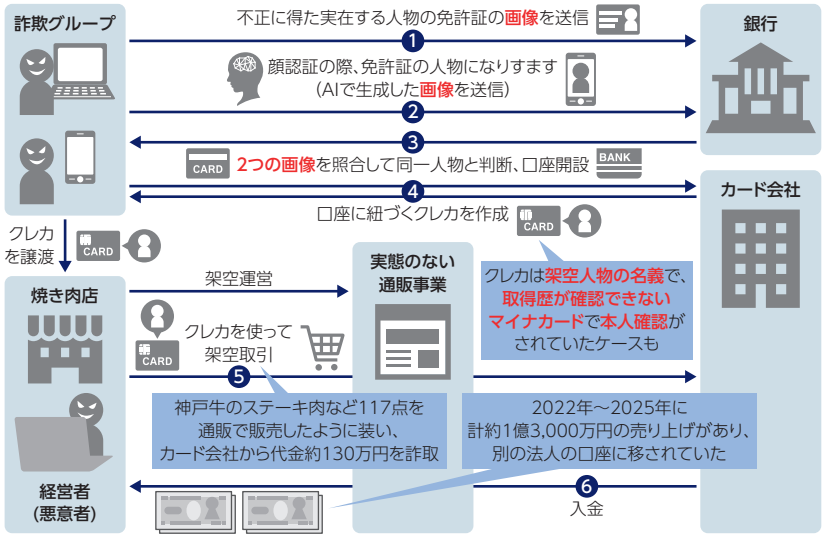
デジタルクライムにおける生成AIの悪用は、防御側の想定を超えるスピードで進行している。その背景にはディープフェイク等に代表される技術の「高度化」と、AIのコード・文章生成能力による攻撃プロセスの「効率化」がある。以下、高度化と効率化を駆使した顕著な具体的な手口と、正規のサービス仕様の隙のメカニズムを紹介する。

まず、生成AIによるなりすまし技術の「高度化」によって、オンライン本人確認のeKYC方式（本人確認書類の撮影＋容貌確認）の脆弱性を突いた不正が確認された。

具体的には、悪意者が偽造・盗難した本人確認書類と生成AIで生成した顔画像を悪用して口座を開設。不正取引の期間が2022年4月から2025年9月に及ぶことから、eKYCの判定ロジックにおいて容貌確認が静止画のみであった時期の仕様の隙を突き、開設した口座や、架空名義の信用情報を悪用して利用限度額を確保していたと考えられる。さらに、これらを用いて架空のEC取引を捏造し、決済代金を詐取する巧妙な資金化が常態化している。この詳細な手口は図表を参照いただきたい。

また、複数のプラットフォームを跨いで情報を転用・資金化していた事例も確認された。フィッシングサイトで詐取された運転免許証の画像を元に、生成AIで動的な容貌画像（セルフィー）を作成し、eKYCを突破して口座を不正開設する手口であった。

図表　生成AI悪用の詐欺手口



（出所）野村総合研究所

本人確認書類の画像を標的としたフィッシングは、2024年1月頃から継続的に観測されている。プラットフォームを跨いだ情報の転用がこれまた常態化しており、手口の巧妙化が顕著である。

フィッシングサイト構築の領域においては、生成AIの高度なコード生成能力により攻撃の「効率化」が進んでいる。文章やUIの最適化によって、不審なポイント（日本語の乱れ、デザインの不自然さなど）が徹底的に排除され従来型のセキュリティ意識を持つユーザーであっても見破ることが困難となっている。

生成AI時代に求められる対策

生成AI時代のデジタルクライムに対抗するためには、防御側も人手判断に依存せずAIを活用した自動対応を主軸に据え、組織的な防衛水準の刷新と具体的なシステム対策を一体化させた多層防御へと転換しなければならない。

第一に、攻撃の「効率化」への対抗である。成功した手口は瞬時にパッケージ化され、業界全体へ横展開される。後手の運用では莫大なコストと信頼失墜を招きかねない。したがって、サービス仕様の設計段階からAIによる自動化・大量実行を想定したリスク分析を徹底し、正規仕様の悪用可能性を考慮しなければならない。さらに業界全体の最新の攻撃動向を常時チェックし、自社の対策水準を業界標準に引き上げるための組織的な情報連携が不可欠となる。

第二に、技術の「高度化」への対抗である。防御側も生成AI・機械学習ベースの技術を主軸に据え、検知ロジックを常に拡充しなければならない。

具体的には、リスクベース認証や行動バイオメトリクスの導入が挙げられる。生成AIを活用して、ユーザーの行動の連続性、スマートフォンの持ち方やタイピング速度といった操作デバイスの癖、取引の内容から総合的にリスクを判定し、不正操作の疑いがある場合はリアルタイムで追加認証やアカウントロックを自動実行する。また、犯罪収益移転防止法改正により現行のホ方式[1)]は原則廃止され、公的個人認証サービス（JPKI）やICチップ読み取りは必須要件となる。

第三に、突破される可能性を前提に、有事におけるサービス停止時の対応を定義したサイバーBCPの策定である。どの機能をどのタイミングで一時停止すべきか明確な判断基準をマニュアル化し、迅速な機能制限やパッチ適用を迷わず実行できる体制を組むことが不可欠である。

金融庁から26年5月にフロンティアAIの脅威に関する要請が発出された。AIの進化に伴い、攻撃側が優位に立ちやすい「いたちごっこ」が激化すると考えられる。前述した正規仕様の悪用もフロンティアAIにより高度化することが想定される。利便性最優先のサービス設計は、結果としてAIによる高速かつ大量の不正リクエストを許容する隙を生み得る。そのため、防御側も高度なAIを武器に据えねばならない。脅威の変容に合わせて自動診断や検知ロジックを迅速に自律更新し続ける、高度なAI防御体制の構築が必要となる。

NOTE

1）「ホ方式」とは、犯罪収益移転防止法施行規則に規定されている本人確認方法。利用者に本人確認書類（運転免許証やマイナンバーカード等）をスマホ等で撮影・送信させ、加えて本人の容貌確認（まばたきや首振りをするセルフィー撮影）を行って双方のデータを照合するもの。

Writer's Profile

上田 なみ Nami Ueda
セキュリティソリューション事業企画部
シニアアソシエイト
専門はデジタルクライム分析
focus@nri.co.jp