TRENDS

AIとセキュリティ
DXの進展に向けた四つの観点とアプローチ

東京電機大学 名誉教授 兼
研究推進社会連携センター 顧問・客員教授
佐々木 良一

はじめに

多くの企業が新たなビジネス価値を創出するためデジタルトランスフォーメーション(DX)に取り組んでいる。このDXの成功の主要な課題が人工知能（AI：Artificial Intelligence）の利用であり，阻害要因となり得るのがセキュリティ問題である。したがってDXの成功のためにはAI自体の適用を高度化するとともに，AIとセキュリティの組み合わせにおいてどのような問題があるかを知る必要がある。そこで，次の四つの観点から問題点と対応策の検討を行った。

(1) Attack using AI（AIを利用した攻撃）
(2) Attack by AI（AI自身による攻撃）
(3) Attack to AI（AIへの攻撃）
(4) Measure using AI（AIを利用したセキュリティ対策）

AIとセキュリティの組み合わせに関する四つの観点と必要な研究・開発

(1) AIを利用した攻撃

今後，AIを利用した不正者によるサイバー攻撃は増加してくると考えられる。特に，AI機能付きのマルウェアは近い将来，確実に誕生するだろう。今後は小さな種々のAI機能付きマルウェアが侵入し，協力しながら環境に最も適した攻撃をするようになっていくのではないかと考えている。少なくとも研究レベルでは，このような動きを考慮して今後の対策を考えておくことが大切となる。

(2) AI自身による攻撃

AIが人間に及ぼす悪影響で最も大きなものは，人間を上回る能力を有するAIが誕生し，将来的に人間が絶滅させられるのではないかという問題である。しかし，現在のAI研究は「汎用AI」ではなく「専用AI」の研究が中心であり，多くのAI研究者は人間に対する反乱が起きる可能性は無視できるほど低いと考えている。これに対し著者は次のように考えている。すなわちAIが反乱を起こす可能性は極めて低いが，心理学でリスク認知バイアスなどが指摘されている通り，リスクに対する人間の知覚能力は極めて低い。また，専用AIでもAI兵器などの分野で反乱が起きると取り返しのつかないことになる可能性が強い。したがってAIが異常行動を起こしたとき，正しく検知・停止できるかどうかの実験などが大切になっていく。

(3) AIへの攻撃

AIシステムへの攻撃による問題を考えておくことも必要である。主な攻撃方法として次のようなものが知られている。

(a) 訓練済みモデルの誤分類を誘発するノイズ付加攻撃：判定・予測用データにノイズなどが加えられると，判定・予測の精度が低下し，誤判定などが誘発され得る。
(b) 機械学習に対する偏った訓練データを意図的に与えることなどが原因で，不適切な判断をさせてしまう攻撃：米国IT企業がAIを用いて自動的に会話するチャットボットを開発し，クラウドソーシングを利用して学習さ

1971年東京大学卒業。同年日立製作所入社。システム開発研究所（当時）にてセキュリティ技術，ネットワーク管理システムなどの研究開発に従事。2001年東京電機大学教授。研究推進社会連携センター総合研究所特別専任教授，サイバーセキュリティ研究所所長などを経て，2020年から現職。日本セキュリティ・マネジメント学会会長，内閣官房サイバーセキュリティ補佐官などを歴任。工学博士（東京大学）。著書に『ITリスクの考え方』（岩波新書），『ITリスク学：「情報セキュリティ」を超えて』（共著，共立出版），『「つながる世界」のサイバーリスク・マネジメント：「Society 5.0」時代のサプライチェーン戦略』（監修，東洋経済新報社）ほか多数。

せた。ところが悪意を持ったユーザーが協力して差別的な意見を繰り返し入力した結果，一日も経たぬうちに同様の発言を繰り返すようになってしまったという。

これらの攻撃方法や対策の研究はすでに行われているが，さらに研究を高度化することが期待される。

(4) AIを利用したセキュリティ対策

セキュリティ対策にAIを用いるアプローチである。学術文献の検索サービスによる論文調査やWEB上の製品紹介の調査から，AIは「マルウェアの検出」,「ログの監視・解析」,「トラフィックの監視・解析」,「セキュリティ診断」,「スパムの検知」,「情報流出」など多くのセキュリティ対策に適用されていることが分かる。

筆者らも (a)「機械学習を利用した標的型攻撃用C&C (Command and Control) サーバの自動判別システム」や，(b)「ルールベースシステムやベイジアンネットワークを利用した知的ネットワークフォレンジックシステム」などにAIを適用してきた。これらの適用を通じてセキュリティ対策にAIを応用するのは有効なアプローチであると考えている。しかし，サイバー攻撃は時間とともに特性が変化することが多く，それぞれの期間における十分なデータの入手が必要であるなどの課題にも対応する必要がある。

AIとセキュリティの組み合わせに関する今後のアプローチ法

セキュリティ研究は常に攻撃者に対する後追い研究であった。何らかの形で攻撃者を先回りし，「Beyond Attackers」を実現できないかという見果てぬ夢がある。この実現のためには前述の四つの観点からの研究の深化はそれぞれ重要であるが，(4)「AIを利用したセキュリティ対策」と他の三つの観点からのアプローチをうまく結合することが重要になるのではないかと考えている。例えば，前述の (1) と (4) とをうまく結合することにより，AIを使った攻撃方法を検討して防御方法を事前実装すれば，新しい攻撃が出てきても事前に対応できる可能性がある。このため，AIによるマルウェアの隠蔽と検出をコンペ形式で行うためのテストベッドを実現し，模擬的な攻防によって，今後出てくる可能性のある新しい攻撃法をあらかじめ把握する研究なども考えられる。

おわりに

より良いDXの進展のためには，AI自体の適用の強化とともに，AIとセキュリティをうまく組み合わせた統合的研究・開発が重要である。日立においてもより良い統合的研究・開発が進めばよいと考えている。また，そこで生まれた技術などを生かし，Lumadaを用いて顧客との新たな協創が可能になっていくことを期待したい。