サイバーセキュリティにおける AIシステムの倫理的使用の原則

kaspersky bring on the future

kaspersky.com

機械学習に対するKasperskyのアプローチ：

https://www.kaspersky.com/enterprise-security/wiki-section/products/machine-learning-in-cybersecurity

高度なテクノロジーがますます重要な役割を果たすようになり、世界は急速に変化しています。特に、急速な発展を遂げている人工知能（AI）はすでに世界に数多くのメリットをもたらしていますが、メリットのひとつがサイバーセキュリティの向上です。毎日新たな脅威が多数出現しており、これらの脅威を全て手動で検知することは不可能です。何年もの間、サイバーセキュリティでは、AIアルゴリズムにより脅威検知プロセスを自動化して高速化し異常を検知して、マルウェア識別の精度を向上させてきました。当社はそのソリューションにおいて、AIの一部であると考えられている機械学習（ML）を採用してきました。

一方で、AIを使用することはリスクを伴うため、関与している全ての関係者による責任あるアプローチが必要になります。従って、当社では、全てのステークホルダーのメリットとなるイノベーションをリードするために、サイバーセキュリティにおけるAI/MLの開発と使用について、以下のような倫理的原則を規定します。また、ほかのサイバーセキュリティ企業に対しても、この原則への賛同を呼びかけたいと考えています。

#1 透明性

当社は、企業が製品やサービスにおいてAI/MLテクノロジーを使用することについて、お客様が情報を知る権利があるという信念を持っています。そのため、**当社のソリューションがどのように動作しAI/MLテクノロジーを利用するのかという原則について、説明することに全力を注いでいます。**透明性への取り組み（Global Transparency Initiative）では、世界各地で多くの「トランスペアレンシーセンター」を開設し、お客様やそのほかのステークホルダーが、AI/MLテクノロジーの活用など当社の開発プロセスを確認し、カスペルスキー製品やソリューションの完全性と信頼性を検証できるようになっています。また、透明性の原則に従って、**可能な限り最大限に解釈可能なAI/MLシステムを開発し、これらのシステムが提供する成果の妥当性を保証するために必要な安全対策を導入することに注力しています。**

#2 安全性

AI/MLアルゴリズムが実際に導入されると、このようなシステムに意図的なエラーを発生させるように設計されたさまざまな形態の攻撃に対して、脆弱（ぜいじゃく）になる可能性があります。サイバーセキュリティにおいて脅威検知のミスによる損失は甚大で、そのため、潜在的な脅威に関して安全性と耐性が非常に重要になります。当社は、**AI/MLシステムの開発と使用において、安全性を最優先に取り組んでいます。**これは、全てのAI/MLシステムの品質を保証するための厳格な対策を実装することで実現しています。このような対策の主な柱として、AI/ML固有のセキュリティ監査の実施と「レッドチーム化」、トレーニングプロセスにおけるサードパーティのデータセットへの依存の最小化、多層型保護設計ベースのアンサンブル運用、クライアントのマシンにインストールされたモデルではなく、必要な安全対策を備えたクラウドベースのMLテクノロジーの採用などがあります。

セキュリティシステムにおいて機械学習をセキュアにする方法

https://content.kaspersky-labs.com/se/media/en/business-security/enterprise/machine-learning-cybersecurity-whitepaper.pdf

マルウェア対策で使用するニューラルネットワークの脆弱性を検証する － 敵対的攻撃からの保護：

https://securelist.com/how-to-confuse-antimalware-neural-networks-adversarial-attacks-and-protection/102949/

#3
人間による制御

マルウェアは、コードの難読化、パッケージ化、暗号化、動的コード生成などの高度な手口を通して変異するため、特にAPT（Advanced Persistent Threat、持続的標的型）攻撃やそのほかの複雑な脅威の分析には専門家の意見が必要になります。最高の保護を実現するために、当社の全AI/MLシステムに**不可欠な要素として、人間による制御を維持することに全力を注いでいます。**当社のAI/MLシステムは、自己完結的かつ自律的なモードで動作するように設計されていますが、その動作は当社の専門家が継続的に監視しています。当社の専門家は、実際に侵入してくるセキュリティ脅威に関する情報にリアルタイムでアクセスし、必要に応じて専門知識を活用してAI/MLシステムの動作を修正し、新たに出現するサイバー脅威に対抗できるよう順応させることができます。当社は、進化し続けるサイバー脅威に対して総合的な保護を実現するために、脅威インテリジェンスのビッグデータに裏打ちされたMLアルゴリズムと人間の専門知識を組み合わせています。

#4
プライバシー

AI/MLシステムの実装では、ビッグデータが極めて重要な役割を果たしますが、使用されるデータの一部は個人情報として認識される場合があります。従って、このようなデータを使用する際の倫理的アプローチでは、全体的に個人のプライバシーを考慮する必要があります。そのため、**当社はプライバシーに対する個人の権利を尊重することに尽力しています。**サイバーセキュリティの観点から具体的には、データ処理の制限、データ構成の削減、可能な限りの仮名化または匿名化、データ完全性の確保、およびデータとシステムを保護し、権利の有意義な行使を確保するためのほかの技術的および組織的対策の適用などに幅広く取り組んでおり、いずれも個人のプライバシーを保護することを目的としています。

データ処理に対するKasperskyのアプローチ

https://www.kaspersky.com/about/data-protection

#5
サイバーセキュリティのための開発

サイバーセキュリティコミュニティ内およびユーザー間で信頼を構築し、維持することは何よりも重要です。個人と組織の保護、および安全な世界の構築を中心とする当社の基本的な価値観に沿って、**当社はAI/MLシステムを防御目的にのみ利用することを約束します。**当社にとって、企業の評判と完全性は非常に重要な資産です。特に防御テクノロジーに特化することで当社の使命を遂行し、ユーザーとそのデータを保護するという取り組みを示し、責任あるサイバーセキュリティプロバイダーとしての評判を高めています。当社は、私たちの生活の全てがテクノロジーによって向上する未来を信じています。世界中の誰もが、テクノロジーがもたらす無限の機会の恩恵を得られるよう、当社はテクノロジーを守ります。

#6
開かれた対話

当社はAIの倫理的使用におけるベストプラクティスを共有するために、全てのステークホルダーとの**対話を推進することに尽力しています。**この点に関して当社は、国連（グローバルデジタルコンパクト、オープンエンド作業部会、インターネットガバナンスフォーラムなど）やそのほかの主要なグローバルプラットフォームを含む、全ての関係者と協議する準備ができています。当社の立場は、全てのステークホルダー間での継続的な協力によってのみ、障害を乗り越え、イノベーションを推進し、新たな展望を切り開くことができるというものです。

ICTセキュリティに関する国連オープンエンド作業部会の支援の下での非公式対話への貢献：

https://docs-library.unoda.org/Open-Ended_Working_Group_on_Information_and_Communication_Technologies_-_(2021)/Kaspersky_SUBMISSION_OEWG_MAY_22.pdf

Kasperskyでは、サイバーセキュリティにおけるAIの使用について、皆様のご意見をお待ちしています。上記の原則に関してご質問やご意見がございましたら、メールでお問い合わせください。TransparencyCenter@kaspersky.com

Cyber Threats News: securelist.com
IT Security News: business.kaspersky.com
IT Security for SMB: kaspersky.com/business
IT Security for Enterprise: kaspersky.com/enterprise

www.kaspersky.co.jp

PR-1064-202310
*記載の情報は2023年10月時点のものです。

株式会社カスペルスキー