特集

AI時代のサイバーセキュリティ

編集にあたって
～新たな時代のレジリエントで持続可能なデジタル経済社会の構築に向けて～

石黒正揮 | （株）三菱総合研究所
佐々木良一 | 東京電機大学
佐々木貴之 | 横浜国立大学

デジタル経済社会において人工知能（AI）の活用は目覚ましく，社会課題を解決するための中核技術として，AIは世の中を劇的に変革している．これに伴い，AIを活用したシステムにおける潜在的な脅威やリスクが拡大している．特に，AIにおける機械学習に固有の脆弱性に対する攻撃，人間の能力を超えるAIが攻撃に悪用される脅威が現実化しつつある．たとえば，自動運転の標識認識システムや顔認識システムにおいて，学習済みAIに微小な細工を加えたデータを入力させることで誤認識を生じさせる敵対的サンプルや，ネットワークスキャンデータを学習し効率的な攻撃コードを生成し，侵入を行うようなマルウェアが出現している．さらには，AI兵器の暴走などのリスクについて，開発者が守るべき原則や倫理についても議論が高まっている．今後，AIにかかわる脅威は人間の理解を超えて進化する可能性があり，従来のサイバーセキュリティの延長線上では防御は困難となり，攻撃と防御の様相は一変していくと考えられる．

AIは，人類にとって脅威でもあり，恩恵をもたらすものでもあるが，攻撃者がAIを悪用することを止めることができない以上，防御側においてもAIを活用したサイバーセキュリティの高度化，自律化を図らなければ，これからのAI時代の安全・安心は確保できない．

AIのサイバーセキュリティについては，国内外の多くの組織により取り組み課題が挙げられている．国内においては，岸田政権において閣議決定された「新しい資本主義」においては，AI実装におけるプライバシー保護のため秘匿化したデータをそのまま

機械学習する技術の開発推進や，デジタルトランスフォーメーション(DX)におけるサイバーセキュリティの強化が挙げられている．DXの実現と推進のためには，サイバーセキュリティの確保は不可欠であり，DXとサイバーセキュリティは一体的に進めなければならない．また，サイバーセキュリティ戦略本部が決定した「サイバーセキュリティ研究開発戦略（改訂）」においては，AIを活用したサイバーセキュリティとAIそのもののサイバーセキュリティの強化が掲げられている．

信頼できるAIにかかわる取り組みもサイバーセキュリティを確保する上で重要となる．EUでは，2021年提出されたARTIFICIAL INTELLIGENCE ACT（Proposal）においてAI利用リスクへの対処に関する義務や禁止事項を規定しており，法的拘束力を持つ初の規制として注目される．IEEE ETHICALLY ALIGNED DESIGNでは，AIシステムが人間の幸福を優先することを目的とした概念を提唱し，説明責任（Accountability），透明性（Transparency）の確保の必要性を挙げている．2021年開催されたG20ローマ・サミットでは，AIシステムに関する透明性と責任ある開示，AIシステムのセキュリティ，説明責任の重要性を掲げているG20 AI Principle（AI原則）を推進することが確認された．

総務省「AIとセキュリティに関連した技術研究開発課題に係る調査」の有識者検討会においては，AIセキュリティの研究開発のためにAI検証テストベッドの構築とデータ整備の重要性が示されている．特に，AI検証テストベッドを活用して，AIを用いた攻撃と防御の攻防戦を通じて高度なセキュリティを実現することが可能となると考えられる．

このような動向を踏まえて，本特集では，AI時代のサイバーセキュリティの脅威・リスクとそれらに対する技術・取り組状況および今後の課題の展望についてまとめることとした．AIにかかわるセキュリティは，(1) Security for AI（AIのためのセキュリティ），(2) AI for Security（セキュリティのためのAI），(3) Attack using/by AI（AIを悪用した攻撃または自律的AI自身による攻撃）に分けられる．AIセキュリティは，多様なステークホルダによる取り組みが不可欠であるため，本特集では，産官学におけるサイバーセキュリティの最前線で活躍する第一人者の知見を結集して関連分野の取り組み動向，課題，今後の展望についてまとめた．各テーマの概要は次ページにまとめている．

本特集で紹介したサイバーセキュリティに関する取り組みと今後の課題への対応を推進することにより，レジリエントで持続可能なデジタル経済社会の基盤を構築し，新たな時代のサイバーセキュリティを切り開いていくことが期待される．

(2022年7月31日)

概要

1 AI サイバーセキュリティのリスクと対策および今後の課題～社会的に受容される AI の実現に向けて～

応 般

石黒正揮 | （株）三菱総合研究所

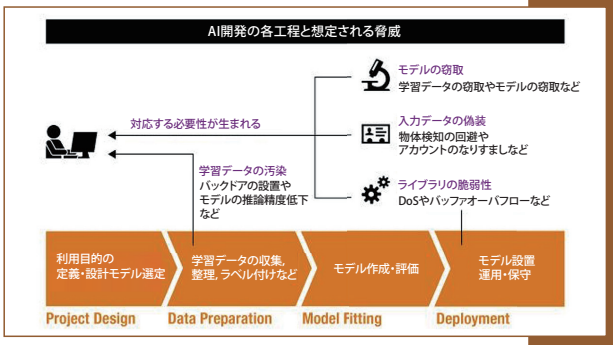
DX や SDGs の実現に向けた社会課題の解決において人工知能は中核技術として世の中を劇的に変革している．このような中で，AI にかかわるセキュリティ事故は急激に増加している．AI 時代のサイバーセキュリティは，AI 固有の脆弱性，人間を超える AI を用いた脅威，AI 対 AI の攻防を前提とした進化への対応のため，従来のサイバーセキュリティとは様相が一変する．本稿では，AI サイバーセキュリティの対策，技術課題，今後の展望についてまとめる．

2 開発者のための AI セキュリティ入門
― AI に対する攻撃手法と防御手法を俯瞰する―

応 般

高江洲勲 | 三井物産セキュアディレクション（株）

さまざまな分野で AI の社会実装が進む一方で，AI に対する攻撃手法も数多く生まれている．AI への攻撃が横行する時代が到来する前に，AI の防御技術を確立することは急務であるが，AI への攻撃手法は既存手法とは原理が異なり，従来の防御技術のみで AI を守ることは困難である．そこで本稿では，開発者の皆様に AI セキュリティの理解を深めてもらうために，AI に対する攻撃手法と防御手法を幅広く取り上げる．

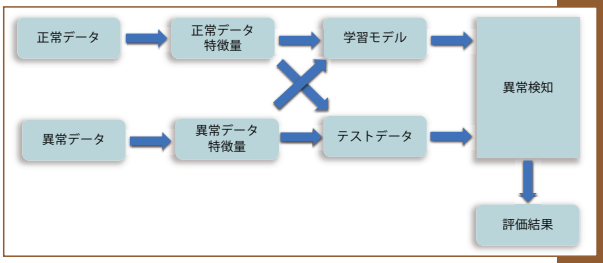


3 AI によるサイバーセキュリティ防御
― AI を活用したセキュリティ対策研究の最前線―

応 般

清本晋作・中原正隆・成定真太郎・長谷川健人 | （株）KDDI 総合研究所

本稿では，AI のセキュリティ対策への活用，特に AI を活用した異常検知について，その概要，ならびに，ネットワーク，ソフトウェア，ハードウェア，それぞれのセキュリティ対策への定期用事例を述べる．また，本稿の最後で，AI のセキュリティ対策への活用における課題と，今後の研究開発の方向性について論じる．



概要

4 サイバーセキュリティ DX を促進する自動化技術の発展

高橋健志 | 国立研究開発法人　情報通信研究機構

各所で人材不足が叫ばれる中，セキュリティオペレーションを着実かつ継続的に実施すべく，AI 技術を活用したセキュリティオペレーションの自動化が期待を集めている．AI 技術は，人が実施していたオペレーションを部分的に代替するほか，AI に技術やノウハウを継承することで，効果的かつ効率的なセキュリティ対策を実現可能にする．本稿では，マルウェア解析や広域攻撃観測などのさまざまな分野にて，セキュリティオペレーションを自動化・省力化するための AI 技術の研究開発の現状について俯瞰する．そして，さらなる AI 技術の利活用促進に向けた課題について議論する．

5 AI を活用したシステムへの攻撃と防御に関する最新セキュリティ研究動向

森　達哉 | 早稲田大学／理研 AIP

本稿は，自動運転における環境認識など，AI を活用するシステムをターゲットとした敵対的入力攻撃とその防御方法について，いくつかの事例をまじえて解説する．また，AI システムの敵対的入力を評価する上で重要な観点となる，アナログ領域の敵対的入力，リアルタイムな攻撃を実現する上で鍵となる環境の変化に対する頑健性，機械学習に依存しないチェック機構に基づく対策方法の有用性について概説する．

6 安心安全なデータ利活用に向けた取り組み
―ソサエティ DX におけるデータ・AI モデルの保護を実現するサイバーセキュリティ技術―

山中啓之 | （株）NTT データ　　諸橋玄武・森山敏行 | NTT 社会情報研究所

機械学習技術の発展やデジタル化の普及に伴い，データが持つ価値は一層高まっており，今後は企業や業界を超えたデータ活用を行う社会となっていく．一方，プライバシー保護やデータ主権に関する機運も高まっている中，機微なデータを保護しつつ安全に利活用できる仕組み・技術が求められている．企業や業界の壁を超えてデータを統合・分析するための技術として「秘密計算」「連合学習」「差分プライバシー」について概説する．