SCIS 2022 2022 Symposium on
Cryptography and Information Security
Osaka, Japan & Online, Jan. 18 — 21, 2022
The Institute of Electronics,
Information and Communication Engineers

AI Safety とSecurity の研究動向:国際と国内・産学官
Research Trend of Ai Safety and Security

櫻井　幸一*　　溝口　誠一郎 †
Kouichi SAKUAI　Seiichiro MIZOGUCHI

キーワード　人工知能、機能安全、サイバーセキュリティ、安全論証

あらまし

2005 年に人工知能学会はセミナー「コンピュータセキュリティと AI」を主催している[1]。このセミナーの目的である「人工知能とセキュリティ分野の研究者間の交流，セキュリティにおける技術動向の収集，人工知能研究者の新たなる応用研究の場の探求等へのに貢献」“は、現在も期待されている。しかし、今日との違い（というよりも進化・変化）は、コンピュータからサイバー空間/社会へ、また AI 操縦自動車の登場による機能安全対セキュリティを議論すべき時代になり、産学のより一層の交流が期待される。
国際研究集会では、主要 AI 系国際会議/併設ワークショップ IJCAI/AIsafety や AAAI/SafeAI が対象としている人工知能と安全性に加えて、セキュリティ系国際会議/併設ワークショップ ACM-CCS/AI-SECURITY や IEEE Security&Privacy/Deep-Learning-Security などが活性化してきている。

近年、AI 技術を利用した多くの製品やサービスが出現し、その長所ばかりが注目される一方、人が介在しないシステム主導の自律型意思決定の短所も浮き彫りになってきた。特に、安全性に影響を与えうる意思決定を担う AI 技術利用のシステムも存在し、その意思決定が引き起こす機能不全や、意思決定に対するサイバー攻撃の危険性も考慮する必要がある。

ISO/IEC ガイド 51:2014 においても、「安全」を「許容できないリスクが無いこと」と定義しており、潜在的な故障の存在を認めた上で、故障による事故の発生やその影響を最小化する取り組みをするとしている。例えば、自律運転車両などでは AI 技術の積極的な利用が想定される一方、従来までの伝統的な「安全性論証アプローチ」では課題が多く、AI 専門家の知見を生かした新たな「安全性論証」が必要となる。

＊ 九州大学（sakurai[at]inf.kyushu-u.ac.jp）
† DNV ビジネスアシュアランスジャパン株式会社（seiichiro.mizoguchi[at]dnv.com）

このような背景のもと、産学の垣根を超え、これらの分野に関連する研究者・技術者間における幅広い意見交換を通じて、より充実した AI 技術利用システムの安全性論証の議論を進めるため、AI とセキュリティを積極的・明示的に扱う研究協議会[3]を設立した。（代表：九大・櫻井 と主幹事：DNV ビジネス・アシュアランス・ジャパン　溝口 ）

また、AI のセーフティとセキュリティに関する誤動作・攻撃・防御・追跡・分析を含む新しいアイデアを広く模索し、研究を深めることを目的とし、人工知能学会・合同研究会・萌芽研究会セッションが企画された[4]。

本発表では、国際と国内、さらに産学官における人工知能 Safety and/vs　Security の現状と課題を論じる。

謝辞:
本研究は　DNV ビジネス・ジャパンと Open RDG-AI/SS 事務局の支援を受けています。

参考文献

[1] 人工知能学会セミナー「コンピュータセキュリティと AI」2005
[2] 櫻井幸一 “AI のセキュリティとトラストに関する国際学術動向” DNV GL Safety & Security フォーラム 2021.02.10
[3] 人工知能の安全性とセキュリティに関する研究協議会 2021.11.05
(https://sites.google.com/view/openrdg-aiss/Home)
[4] AI Safety & Security 人工知能学会合同研究会・萌芽研究会セッション 2021.11.26
[5] 溝口・櫻井 （SCIS2022/本シンポジむ）
[6] Simen Elsdevik “AI+Safety” DNV-GL 2018.08.18
(https://ai-and-safety.dnvgl.com/#sec-al-risk-ex)
[7] ISO/IEC ガイド 51:2014