情報セキュリティ白書

Information Security White Paper

2025

一変する日常：支える仕組みを共に築こう

独立行政法人情報処理推進機構
Information-technology Promotion Agency, Japan

「情報セキュリティ白書2025」の刊行にあたって

「情報セキュリティ白書」は、2008 年以来、サイバーセキュリティ分野における、政策や脅威の動向、インシデントや被害の実態等をまとめ、皆様のセキュリティ対策の推進、学習・研鑽等にお役立ていただくという趣旨で発刊し、産業界、学界、一般の方に広く愛読されてきました。

サイバー空間を巡る脅威は年を追うごとに質・量ともに増大しております。2024 年も国内国外を問わず、ランサムウェア攻撃、標的型攻撃、DDoS 攻撃等、様々なサイバー攻撃による脅威に晒されました。また、今般の厳しい国際情勢下において、影響工作を始めとした地政学的背景に起因するサイバー空間のリスクも顕在化しております。サイバー攻撃の手口も、取引先や委託先等のサプライチェーン上でセキュリティ対策が不十分な部分を入口とするものや、複雑なソフトウェアのサプライチェーンの脆弱性を狙ったもの、更には、生成 AI を悪用したもの等、一層高度化・巧妙化しております。

他方、データ駆動型の便利で豊かな社会、Society 5.0 の実現を目指し、サイバー空間とフィジカル空間が融合していく中で、セキュリティ面でのリスクが顕在化してきております。これまでのフィジカル空間での経済社会行動が IoT 機器やロボット等、様々なデバイスとつながることによりデータ化され、ネット上のサイバー空間に集積し、そのビッグデータが生成 AI により解析、最適化されるサイクルの中で、サイバー攻撃を許す隙が増えるとともに、一度インシデントが起きるとその影響が瞬時に広範に伝播し、大規模な情報漏えいやインフラの機能不全をもたらすリスクがますます高まってきております。

こうした中で、国内では、2022 年 12 月に閣議決定された国家安全保障戦略において「サイバー安全保障分野での対応能力を欧米主要国と同等以上に向上させる」との目標が掲げられ、2025 年 5 月にはサイバー対処能力強化法及び同整備法が成立し、「国民生活や経済活動の基盤」と「国家及び国民の安全」をサイバー攻撃から守るための能動的なサイバー防御を実施する体制の整備が進められています。

また、経済社会インフラが直面するサイバーリスクへの耐性を確保する観点から、システムの設計段階、すなわち、アーキテクチャーレベルでセキュリティを組み込んでいく、「セキュア・バイ・デザイン」の視点に立った様々な制度整備や取り組み、これらを推進していくための人材や技術等、サイバーセキュリティ供給能力の強化に向けた取り組み等も新たに動き出しております。

本白書が、2024 年度の情勢を踏まえた脅威分析と政策動向の総括を通じ、関係者の皆さまの日々の対策検討や実践に資するものであること、そしてより安全で信頼されるデジタル社会の確立に寄与する一助となることを、心より願っております。

2025年9月

独立行政法人情報処理推進機構（IPA）

理事長 齊藤 裕

目次

コラム

2025

情報セキュリティ白書

序章

2024年度の情報セキュリティの概況

近年、情報セキュリティの脅威は一層深刻化しており、サイバー攻撃の手法も高度化している。2024 年においては、ランサムウェア攻撃や、DDoS 攻撃等のインシデントが相次ぎ、重要インフラや企業の運営に影響を与えた。国内では 2024 年 6 月に、総合エンターテインメント企業がランサムウェア攻撃を受け、動画配信サービスやオンラインショップの障害、出荷遅延等の被害が生じた。また印刷会社に対するランサムウェア攻撃では、約 60 の委託元に影響が及んだ。これらのインシデントは、サービス停止や情報漏えいにより多数の企業・組織及び利用者に被害をもたらし、情報セキュリティ対策の重要性を改めて認識させた。国外では、鉄道、空港、水処理施設等の重要インフラに対してランサムウェア攻撃被害が発生し、安全保障の観点からも対策が急務となっている。

2024 年には、政治的なイベントに関連した DDoS 攻撃が増加し、公共の安全や秩序が脅かされる事態も発生した。2024 年 7 月、8 月にはオリンピック関連のスポンサー、パートナーの Web サイトを標的とした DDoS 攻撃が観測された。また 2024 年は世界各国で重要な選挙が行われ、選挙運動、政党、選挙インフラを対象とした DDoS 攻撃が観測された。米国では、大統領選挙を狙った DDoS 攻撃が 11 月に発生した。日本でも、2024 年 7 月と 10 月に安全保障イベントに関連した DDoS 攻撃が発生した。また、2024 年末から 2025 年初頭にかけて、航空会社、金融機関、携帯通信会社が相次いで DDoS 攻撃を受け被害が発生した。これらの攻撃には IoT ボットネットが利用されている。

2025 年 1 月、警察庁と NISC（現 NCO）は、2019 年ごろから継続していた複数の攻撃キャンペーンについて、国家に支援されたサイバー攻撃グループによるものとして注意喚起を行った。これらの攻撃は、日本の安全保障の棄損や先端技術情報の窃取を目的としており、攻撃手法の公表を通じて被害の拡大防止が呼びかけられた。

国際的には、国家を背景としたサイバー攻撃の激化による被害が発生した。「Salt Typhoon」と呼ばれる攻撃グループによる攻撃では、米国通信事業者 9 社を含む世界中の企業数十社のシステムへの侵入が観測され、広範なスパイ活動及び情報収集が行われたことが確認された。国家を背景とした攻撃グループに対しては複数の国、組織が連携し、情報共有や摘発を行っている。

2024 年は AI の悪用による被害も報告された。前述の選挙妨害においては生成 AI が偽情報の生成に多用されたという。偽情報の流布を利用した情報操作型サイバー攻撃は、社会の混乱や分断、政府機関の信頼失墜等、サイバー領域と認知領域の双方にわたる攻撃手段として，国家の安全保障上の脅威ともとらえられる。今後も警戒が必要である。

このような状況を踏まえ、日本国内においてもサイバーセキュリティ政策の強化が進められた。ランサムウェア攻撃の被害拡大や DDoS 攻撃における IoT 機器の悪用に対して、政府は 2024 年度のサイバーセキュリティ戦略において、サプライチェーン・リスクへの対応と DX 推進・支援の強化を掲げた。経済産業省は「ソフトウェア管理に向けた SBOM（Software Bill of Materials）の導入に関する手引」「セキュア・ソフトウェア開発フレームワーク（SSDF）導入ガイダンス」の発行等で、設計段階からセキュリティを考慮するセキュア・バイ・デザインの施策を推進した。また、2025 年 3 月には IoT 製品のセキュリティ評価認証制度として「セキュリティ要件適合評価及びラベリング制度（JC-STAR）」の運用が開始された。更に、サプライチェーン強化に向けたセキュリティ対策評価制度の検討等にも取り組んでいる。

サイバー安全保障分野では、「外部からのサイバー攻撃について、被害が発生する前の段階から、その兆候に係る情報その他の情報の収集を通じて探知し、その主体を特定するとともに、その排除のための措置を講ずることにより、国家及び国民の安全を損なうおそれのあるサイバー攻撃の発生並びにこれによる被害の発生及び拡大の防止」を図る「能動的サイバー防御」の実現に向けた検討が進められた。その結果、2025 年 5 月には「重要電子計算機に対する不正な行為による被害の防止に関する法律」及び「重要電子計算機に対する不正な行為による被害の防止に関する法律の施行に伴う関係法律の整備等に関する法律」が成立した。今後、官民連携の強化、通信情報の利用、攻撃サーバーの無害化等の実践を通じ、サイバー安全保障分野での対応能力向上が期待される。

	主な情報セキュリティインシデント・事件	主な情報セキュリティ政策・イベント
2024年 4月	● 米国のセキュリティベンダーが提供するファイアウォール用OSに対するゼロデイ攻撃を確認（1.2.4） ● 米国のマルチクラウドデータウェアハウスプラットフォームを利用している複数の組織を標的としたデータ侵害が発生（1.1.1）	● 米国「外国敵対勢力が管理するアプリから米国人を保護する法」成立（4.1.1）
5月	● 国家の支援が疑われるサイバー攻撃グループが、国内の暗号資産関連事業者から約482億円相当の暗号資産を窃取（1.2.2） ● 行政機関等から通知書等の印刷と発送を請け負っていた印刷会社でランサムウェア被害が発生（1.2.1）	● 「重要経済安保情報保護活用法」成立（3.1.1） ● NISCと警察庁が、米国CISAの作成したサイバー脅威緩和に関する国際ガイダンスに共同署名（4.1.1） ● 「AIソウル・サミット」開催（2.1.3）
6月	● 総合エンタメ企業が展開する動画共有サービス等がランサムウェア攻撃を受け、サービス停止（1.2.1）	● 「G7プーリア・サミット」開催（3.1.1）
7月	● 日本・NATOの活動に抗議するDDoS攻撃が発生（1.2.3） ● 米国サイバーセキュリティ会社のシステム障害により世界約850万台のWindowsデバイスに影響が発生（1.1.1） ● パリオリンピック関連のスポンサー、パートナーを標的としたDDoS攻撃が発生（1.1.1）	● NISCと警察庁は、オーストラリアのACSCが作成したAPT40に関する国際アドバイザリーに共同署名（4.1.1） ● NISC「サイバーセキュリティ2024」公表（3.1.1） ● NISTは、生成AIのセキュア開発のためのプロファイルである「SP 800-218A」公開（4.1.2）
8月	● 不動産仲介業の従業員が同業他社に転職する際、不動産登記簿に基づく社内資料を不正に持ち出し（1.2.7） ● 米国の国際空港がランサムウェア攻撃を受け、フライト情報表示等の重要な機能に影響が発生（1.2.5）	● EU「AI Act」発効（2.1.1、2.1.3） ● 経済産業省「ソフトウェア管理に向けたSBOM（Software Bill of Materials）の導入に関する手引ver 2.0」公表（3.1.3）
9月	● 米国司法省は、国家の支援が疑われる攻撃グループに侵害された20万台超の消費者向け機器からなるボットネットを無害化したと発表（1.2.2） ● 米国の水処理施設にランサムウェア攻撃（1.2.5）	
10月	● ランサムウェア開発者らを欧州刑事警察機構等による共同捜査により逮捕（4.1.1） ● 日米共同統合演習に抗議するDDoS攻撃が発生（1.2.3）	● オーストラリアのACSCは、重要インフラ事業者に向けて策定した「OTサイバーセキュリティの原則」公開（4.1.5）
11月	● 米国大統領選挙で、複数の国家が関与すると見られる影響工作を確認（2.2.3） ● 米国大統領選挙期間中に大規模なDDoS攻撃が数日にわたって発生（1.1.1） ● 国家の支援が疑われる攻撃グループが9社の米国通信事業者、及び世界中の企業数十社を侵害していたことをFBI等が公表（1.1.1、1.2.5）	● IPAとAJCCBCは、オランダのNCSCと協働し、タイで重要情報インフラ保護に関する人材育成プログラムを提供（4.1.1） ● 経済産業省とIPAは、米国政府・EU政府と連携し、「インド太平洋地域向け日米EU産業制御システムサイバーセキュリティウィーク」開催（4.1.1）
12月	● 米国の地域交通局がランサムウェア攻撃を受け、鉄道の遅延等の一時的な混乱が発生（1.2.5） ● 年末から年始にかけて国内の重要インフラ企業等へ大規模なDDoS攻撃が発生（1.2.3）	● EU「サイバーレジリエンス法」発効（4.1.3） ● 国連総会にて、サイバー犯罪に関する包括的な国際条約である「国連サイバー犯罪条約」採択（4.1.1） ● EUのサイバーセキュリティ能力を強化する「サイバー連帯法」及び「改正サイバーセキュリティ法（CSA）」が成立（4.1.3）
2025年 1月	● 警察庁及びNISCは、安全保障や先端技術に係る情報窃取を目的とした攻撃キャンペーンについて、国家の関与が疑われる組織的なサイバー攻撃活動であるとして注意喚起（1.2.2）	● 「U.S. Cyber Trust Mark」運用開始（4.1.2） ● 米国大統領令14144、ソフトウェアサプライチェーンセキュリティ強化策等を指示（4.1.2） ● EU「デジタルオペレーショナルレジリエンス法」全面適用開始（4.1.3） ● 米国大統領令14179、Biden政権のAI統制施策を棄却（4.1.2）
2月	● 営業秘密にあたる研究データを外国企業に漏えいしたとして国立研究開発法人の元研究員に有罪判決（1.2.7）	● 「AIアクションサミット」開催（2.1.3） ● 「サイバー対処能力強化法案」及び「同整備法案」が閣議決定（3.1.1） ● 米国DHS、CISA等所管機関の活動縮小（4.1.2）
3月	● 地方銀行をかたる自動音声を含む電話による大規模なボイスフィッシング被害が発生（1.1.2）	● 経済産業省「セキュア・ソフトウェア開発フレームワーク（SSDF）導入ガイダンス案（中間整理）」公開（3.1.3） ● IPA「セキュリティ要件適合評価及びラベリング制度（JC-STAR）」運用開始（3.3.1）

※表には、2024年度の主な情報セキュリティインシデント・事件、及び主な情報セキュリティ政策・イベントを示している。表中の数字は本白書中に掲載している項目番号である。他のインシデント・事件や、政策・イベント等については本文を参照いただきたい。

第2章
最近のサイバー空間を巡る注目事象

最近のサイバー空間には AI の普及により急速な変化がもたらされている一方、その AI を悪用した偽情報の拡散が公共への大きな脅威となっている。

そこで本章では注目事象として、AI セーフティ実現に向けた取り組みと、偽情報・誤情報の脅威と対策を取り上げる。

2.1 AIセーフティ実現に向けた取り組み

2025 年現在、AI（Artificial Intelligence:人工知能）という言葉は連日のようにニュース等で取り上げられている。その大きな契機となったのは、2022 年 11 月に登場した米国 OpenAI, Inc.（以下、OpenAI 社）の対話形式（チャット形式）で利用できる ChatGPT と、その後に続く、高度な汎用性を備えた AI の急速な発展と普及であろう。最先端の汎用的 AI（General-purpose AI）は、あたかも人間の知能を再現あるいは凌駕するような振る舞いを見せるが、実際にそれらの AI が何を実現するのか、その利用にどのような危険性があり、具体的な対策として何が講じられているのかについて、様々な議論が展開されている。本節では、AI の技術的概要、想定されているリスク、AI の安全な利用に関わる取り組み、AI とサイバーセキュリティにまたがる議論を概観する。

2.1.1 AIの急速な発展

AI 技術の研究史はコンピューターの登場から間もない 1950 年代にまでさかのぼり、代表的な技術も度々変化を重ねている。技術の発展とともに AI のできることが広がる一方で、AI の詳細な動作に対する理解は困難なものとなり、最先端の AI は事実上のブラックボックスと化している。本項では、AI 技術の発展を大まかに振り返る。

(1) 20 世紀中の AI 技術

1950 年代からインターネットの本格普及期となる 2000 年ごろにかけて、AI 技術や関連技術が数多く開発された。チェスや将棋を指すプログラムも最初期のころから存在している。自動計画や論理型プログラミングといった分野の研究も広がった。人間の脳の神経網にヒントを得て考案されたニューラルネットワークも古くより研究されている。例えば、1989 年に米国では、ニューラルネットワークを用いた文字認識により、手書きの郵便番号を機械に判読させる技術が実用化された※1。

(2) 機械学習とディープラーニングで発展した AI

2000 年ごろを境に世界中でインターネットの利用が爆発的に広まり、デジタル化されたデータの量も飛躍的に増大した。しかもこれらのデータはインターネットを通じて全世界に公開され、AI の技術開発にも利用できるデータが多く含まれていた。更に、コンピューター自身の高性能化によって、膨大な量のデータを高速に処理できる環境が整った。特に、AI が必要とする種類の計算処理に有効な、GPU（Graphics Processing Unit）と呼ばれる半導体の開発と普及が進んだ。GPU は当初はゲーム用の描画処理高速化のために開発されたものだが、その計算能力が AI 分野でも有益であると広く認識されるようになり、AI の計算処理に使用されるようになった。このような背景のもとに、「機械学習（machine learning）※2」と呼ばれる技術が 2000 年代に急速に発達した。機械学習の発達によって、従来の AI では難しいとされていた水準の性能が、様々な用途において実現された。例えば、電子メールのスパムフィルターの高性能化※3 や、EC サイトにおける高精度なレコメンデーション等※4 である。

更に機械学習の発達には、ニューラルネットワーク技術の大幅な進歩があった。大きな転換点となったのは、ディープラーニング（深層学習）※5 の登場である。ニューラルネットワークは人間の神経細胞を模した基本部品（人工ニューロン）が連なってできるネットワークであるが、これを大規模化する技術を実用化したものがディープニューラルネットワークである。ディープニューラルネットワークの

関連技術を総称してディープラーニングと呼ぶ。

写真を見てその写真に写っているものが何かを高精度に判定できるAIは、ディープラーニングによって2012年ごろに実現した。ディープラーニングはその後長足の進歩を遂げ、人間の発話を聞き取って文字に変換したり、テキストで与えた指示に基づいて絵を描いたり、自然言語のクイズに答えたり等、従来であれば困難であった様々な機能を、しばしば人間よりも優れた正答率で実現する水準に達した※6。他方で、ディープニューラルネットワークが人間をも凌駕する程の性能をどのように発揮するのか、動作メカニズムの詳細を明確に説明することは困難になっていった。

(3)汎用的AIの登場・発展

「文章、画像、プログラム等を生成できるAI」を総称して「生成AI(Generative AI)」と呼ぶ※7。生成AIの特徴は、自然言語文での入力が可能である等、入力データの自由度が非常に高く、それらの入力に対して適切に見える結果を生成することである。生成AIサービスの代名詞となったOpenAI社のChatGPTは2022年11月に登場した※8。ChatGPTは、対話形式(チャット形式)で、人間が入力した自然言語文に対し、人間のような自然な応答を返す。自由度の高い入力を受け付けるという性質は、入力内容を変えるだけで、一つのAIを多種多様な用途に転用できる可能性を示唆する。欧州連合(EU:European Union)の「AI Act※9」では、著しい汎用性を示し、異なるタスクを広範囲にわたって適切に実行でき、多種多様なシステムやアプリケーションに統合可能なAIを「汎用的AI(General-purpose AI)」と呼んでいる。生成AIは汎用的AIの中核的実装技術となっている。

汎用的AIの開発にあたっては、従来とは桁違いに大規模なディープニューラルネットワークを対象に、膨大な量の機械学習を実行する必要がある。そのためには、高度な技術・知見を有する専門家を集め、学習に必要な大量のデータを収集し、膨大な計算を実行するための大規模なコンピューター資源を確保することが必要とされる※10。代表的な汎用的AIとしては、OpenAI社のGPT、Meta Platforms, Inc.(以下、Meta社)のLLaMA、Google LLC(以下、Google社)のGemini、Anthropic PBCのClaude等がよく知られている。これらの開発元企業はいずれも高度な技術力と莫大な資金を有するグローバル企業である。経済的体力が必要な汎用的AI開発競争には高い参入障壁が築かれている現状があり※11、この状況は経済的・技術的な「AI軍拡競争(AI arms race)」とも呼ばれる※12。

汎用的AIの技術の発達の勢いは衰えを見せず、去年できなかったことが今年はでき、今できないことが来年あるいは3ヵ月後にも可能になるかもしれないというスピード感で開発競争は進んでいる。この背景には、ディープニューラルネットワークの規模や駆動計算量を拡大すればそれに応じてAIの性能が向上し、この傾向に上限がないという経験則(スケーリング則)※13が発見されたことがある。実際、汎用的AIの性能を測るベンチマークテストにおいて、汎用的AIの正答率は人間を超える水準に順次到達している※14。遠くない将来には、あらゆる面で人間と伍する「AGI(Artificial General Intelligence:汎用人工知能)」や人間を超える「ASI(Artificial Super Intelligence:人工超知能)」と呼ばれるAIが登場するという議論もあるが、実現可能性は定かでない。その反面、汎用的AIの動作メカニズムは解明されておらず、AIの研究開発者の間でもブラックボックスとなっている。AIのブラックボックス性は、後述するAIセーフティをどう保証するかという課題にも密接に関わっている。

AIの発展を振り返ってみると、機能の多様化、性能の向上と引き換えに、AIの内部動作はブラックボックス化が進んできたと言える。ブラックボックスであるということは、AIが何かの拍子に人間の期待に反する危険な挙動を見せるおそれがあることも意味する。高度な機能・性能による利益を求めてブラックボックス化した汎用的AIを利用するのか、リスクを回避し、汎用性には乏しいが振る舞いがよく解明されている従来のAIを利用するのか、あるいは、リスク対策を徹底した上で汎用的AIを積極的に採用するのかという、リスクを意識したトレードオフ判断がAI利活用に求められている。

2.1.2 AIリスクとは何か

AIの利用の拡大や社会への浸透に伴って何らかの危害が人間や社会に生じるリスクをAIリスクと呼ぶ。2025年1月末に英国科学技術イノベーション省(DSIT:Department for Science, Innovation & Technology)が発表した「International AI Safety Report 2025※15」は、AIリスクに関する議論を進める調査報告書であり、その中では、汎用的AIを念頭に置きつつ、AIリスクを次のように整理している。

- AIの悪用がもたらすリスク
- AIの不適切動作によるリスク

- システミックリスク

次項ではこれらのリスクについて、関連する実事例等を紹介しながらより詳細な説明を行う。

(1) AI の悪用がもたらすリスク

サイバー犯罪者等が AI を悪用することで生じるリスクである。生成 AI によって作られた合成ポルノ画像の流布による個人攻撃や人権侵害、架空のニュース画像等による世論操作、サイバー攻撃の半自動化、化学兵器・生物兵器開発の支援等がこれに該当する。

(a) フェイク画像等がもたらす個人への被害

生成 AI が生み出す、本物と見分けがつかない偽の画像や動画、音声、及びその生成技術を「ディープフェイク※16」と呼ぶ。人物の写真を与えてその顔を入れ替えたり表情を変えたりといった加工が行えるだけでなく、まったく架空の画像を作り出すこともできる。ディープフェイクはディープラーニングの登場から程なくして発達し、ディープフェイクを悪用した詐欺や恐喝、個人や組織の評判を傷つけるための偽情報の拡散、心理的な虐待といった悪用が FBI により警告されている※17。特に、女性や子供が被害者となるケースが多く、深刻な人権侵害となる可能性がある※18。英国の調査※19 では、女性は男性よりもディープフェイクによる被害をより恐れているという。日本においても、子供達が性的虐待コンテンツを AI で生成した・生成された事例が報道※20 されており、問題は深刻化の一途をたどっている。

(b) 世論操作

ディープフェイクの増加は、偽情報による世論操作のリスクも深刻化させている。悪意のある人物や団体が AI で生成された偽情報を用いることで、大規模かつ巧妙な世論操作が可能になるのではないかという議論があり、懸念が高まっている。近年は、国家を背景とする脅威アクターがこのような活動を試みる場合、これを「影響工作（Influence Operations）」と呼ぶことが多い※21。

ロシアによるウクライナ侵攻が起きた 2022 年にはウクライナのゼレンスキー大統領が降伏を宣言するディープフェイク動画が流布された※22。報道によれば、ロシアの関与があるとされる「Doppelgänger（ドッペルゲンガー）」という脅威アクターは、偽ニュースやディープフェイク動画を駆使して、ウクライナや西側諸国を否定的に、ロシアを肯定的に描いたコンテンツを広めようと試みたとされる※23。更に、2024 年の米国大統領選挙においても偽情報の拡散活動を行ったとされ、米国司法省による取り締りが行われている※24。日本に対しても、東京電力ホールディングス株式会社福島第一原子力発電所の処理水の海洋放出に合わせるように、「処理水ではなく核汚染水だ」という言説とともに不安を煽る偽情報が SNS を中心に広がった事例を OpenAI 社が報告している※25。ディープフェイク等の偽情報の脅威については「2.2 偽・誤情報の脅威の動向」を参照されたい。

(c) AI を悪用したサイバー攻撃

標的型攻撃のような高度なサイバー攻撃が汎用的 AI によって完全に自動化されるのではないかという懸念がある。幸いなことに、2024 年夏の時点では AI の能力はその水準に達していないと評価されている※26。他方で、ChatGPT のような AI サービスを活用することで、手動で行うサイバー攻撃を効率化する試みが、OpenAI 社と Microsoft Corporation（以下、Microsoft 社）によって確認・摘発されている※27。汎用的 AI を利用することで、高度な技能を持たない人物であってもサイバー攻撃を実践できると指摘されており、サイバー犯罪ビジネスへの参入障壁を引き下げてしまうことが、目下の大きな懸念となっている※28。日本においても、中高生 3 人が汎用的 AI を悪用してハッキング行為を行い、携帯電話会社の契約システムに不正ログインを繰り返し、利益を上げていたという事例が 2025 年 2 月に報道された※29。

(d) CBRN 兵器の開発支援

化学兵器や核兵器等の大量殺戮のおそれのある兵器は、化学（Chemical）、生物（Biological）、放射性物質（Radiological）、核（Nuclear）それぞれの頭文字を取って CBRN 兵器と呼ばれる。開発に必要な知見の多くは国家機密として管理され、材料となる物資の流通が厳格に規制されているため、国家以外のアクターが CBRN 兵器を開発することは容易ではない。しかし、インターネット上に存在する膨大な知識を学習している汎用的 AI や、化学物質や合成生物学に特化した AI であれば、少なくとも知見については十分に補えるのではないかという懸念がある。本項執筆時点で、これが机上の空論ではないことを示す研究※30 や事例※31 の報告がある。報告された事例では、汎用的 AI に生物兵器の開発に関する助言を求め、その助言に基づき民間サービスを駆使して合成した DNA をホワイトハウスに持ち込むというデモンストレーションが行われた。2025 年 1 月時

点※32で、汎用的AIは仮想的なエキスパートとして振る舞える可能性があるとの指摘や、チャットを通じて利用できる汎用的AIよりも合成生物学等に特化したAIの方が有力であるとの指摘もあり、議論は定まっていない。

(2) AIの不適切動作によるリスク

AIによる利用者の意図しない動作によって生じるリスクである。事業管理を任せていたAIが誤作動を起こし損失を被る、AIに履歴書の審査を任せていたところ特定の集団を差別するような偏った審査結果が発覚する、AIが人間の意図と大きく異なる振る舞いを示しAIを含むITシステムが暴走する等が該当する。

悪用は、誰かに危害を与えるような結果を意図して招く行為であるのに対し、不適切動作は、その背後に誰かの悪意がなくても、利用者の意図に沿わない振る舞いをAIが示した結果として事故が起こり、危害が生じる可能性のある動作を指す。以下に示すのはこの種の不適切動作に由来するリスクである。

(a) 信頼性の問題

OpenAI社は、同社のAIであるGPT-4が模擬的な司法試験※33で受験者の上位10%に入る成績を収める性能を発揮したと2023年3月に報告した※34。同時期に、当時のChatGPTが米国の医師国家試験に対して合格水準の成績を獲得できたとの報告もあった※35。このような結果は、汎用的AIが高度な専門家を凌駕する知性を獲得したかのような印象を与える。しかし、前者の模擬司法試験の成績は、最終的な司法試験合格者の中位以下の成績に相当し、必ずしも優れた結果とは言えないことが判明した※36。更に、医療における汎用的AIの性能評価を行った結果、ChatGPTも含め、人種間で答えに差異のない医療上の質問に対して、あたかも人種間で違いがあるような示唆を含む返答が得られた※37。

これらの事例は、AIの振る舞いの信頼性は限られた尺度や切り口だけでは十分に評価できない可能性を示しており、AIに何かの仕事を任せる際には、その結果を任せた人間がしっかりとチェックする必要があると言える。2023年には、米国ニューヨーク州の民事訴訟に必要な資料を作成するのに弁護士が生成AIを利用したところ、実在しない判例6件が参照されていたことが発覚している※38。2025年3月には、日本のこども家庭庁において、虐待が疑われる子供を児童相談所で一時保護するかどうかの判定を支援するAIを開発しようとしたところ、AIによる判定の正確性が疑われるケースが全体の約6割を占め、システムの導入が見送られた※39。

(b) バイアス

与えられた情報を適切に分類することはAIに期待される重要な機能の一つである。例えば、人事部門に寄せられた就職希望者の履歴書を見てAIに採用の適否を判定させることが考えられる。この取り組みは汎用的AIが登場する前の2018年ごろに実際にAmazon.com, Inc.（以下、Amazon社）において試みられたが、技術職への応募履歴書を評価する際に、対象者が女性であるというだけで評価が下がるという偏り（バイアス）が生じることが判明し、当該プロジェクトは中止となった※40。これは過去に採用されていた技術職が男性ばかりであり、機械学習のデータとしてそれらの偏りのある履歴書を用いた結果とされる。同様のバイアスは、顔認識※41や再犯予測※42の傾向に人種別で偏りが出る等の、人権侵害につながる事例が報告されている。また、汎用的AIにおいても、特定職業の従事者の画像を描かせると性別に関するバイアスが生じること※43や、テキストの出力であっても類似するバイアスが生じること※44が知られている。

バイアスを生じさせる要因の中で代表的なものは、学習データ自身の偏りである。上記のAmazon社の事例のように、AIが参考にするデータ自身に偏りがある場合に、これをAI自身が是正することは基本的にできない。このため、学習データの準備の際に入念な配慮を行う必要があるが、学習データの母体となる実社会自体に偏りがある場合には、もはやデータ自体の問題とは言えず、容易には解決できない問題となっている。

(c) コントロールの喪失

例えば、工場内の機器に対するAI制御が人間の手を離れてしまい、非常停止等の介入ができなくなる事態は、コントロールの喪失と見なされる。AIの動作が人間の期待から逸脱した状態でコントロールの喪失が発生すると、大きな被害がもたらされるおそれがある。

Sakana AI株式会社は、学術論文の執筆をアイデアの考案から検証、結果のまとめに至るまで自動化するAI Scientistという研究プロジェクトを2024年8月に発表し注目された※45。この発表は学術研究の自動化を試みておおむね成功したという点だけでなく、自動化された研究プロセスの中で、研究の際の実験時間制限を不正に書き換える等のAIの不適切動作が確認されたという点

でも議論を呼んだ。同社は2025年2月にも、AIの動作の高速化をAI自身に行わせるAI CUDA Engineerという手法を発表し、大幅な高速化を達成したとうたった[※46]。しかし、AIがシステム内の評価コードの抜け穴を見つけ、精度の検証等のチェックを回避していることが判明した。同社は後日、見かけの高速化を行う抜け道をAI自身が発見し悪用したことと、修正に取り組むことを発表した[※47]。AIがいわば人間を出し抜くような振る舞いを見せる傾向は、高度な推論能力を持つとされる最先端の汎用的AIでも見られたとの実験結果が報告されている[※48]。

(3) システミックリスク

AI自体は正常に動作していても、社会にAIが浸透する過程で社会の中の様々なバランスが崩れて、何らかの被害を生じるリスクがある。例えば、労働者の業務がAIに置き換えられることで失業者が増える、一部の大企業や先進国にAIの研究開発能力等が集中し格差が広がる、一部の大手AIベンダーがAI市場を独占することでそれらのAIの不具合が世界に影響する、AIの動作に必要な電力や水資源の多大な消費が著しい環境負荷となる、分散している断片的な個人情報をAIが統合・分析することでプライバシーが侵害される、AIの生成した画像や音声が著作権を侵害する等が該当する。AIが組み込まれた社会システムにおいて発現するこれらのリスクは、AIシステムが社会全体に対して広範かつ深刻な影響を及ぼす可能性を持つものであり、単一の不具合や誤作動が連鎖的に波及し、社会的・経済的に大規模な影響を引き起こすリスクを指してシステミックリスクと呼ばれ、以下のように既に複数の懸念が指摘されている。

(a) 労働市場リスク

汎用的AIは多種多様な業務の効率化を支援する。例えば、プログラミング支援AIツールであるGitHub Copilotを使用したプログラマーの生産性が向上した事例[※49]、中程度の専門的なライティングを主体とする業務において4割の時間短縮と質の向上が達成できた事例[※50]等が報告されている。このような生産性の向上は省力化につながり、結果として、労働市場における求人数の減少が起こると考えられる。世界有数のフリーランス事業者向けマッチングサイトである米国Upworkでは、ChatGPTが登場した2022年11月から2024年2月にかけて、ライティング、カスタマーサービス、翻訳といった、汎用的AIによる支援が有効な職種において16～33%の求人数の減少が確認された[※51]。依頼に伴う時給も下落しており、翻訳業務については20%以上の賃下げが見られた。これらの事象は失業率の上昇を示すものではないが、汎用的AIの浸透がもたらす社会的な負の側面を示す事例である。米国において生じた1980年から2016年にかけての賃金格差の拡大の半分以上は、高度な技術による自動化の影響を受けて、定型業務に特化した労働者の賃金が低下したことで説明されるとの指摘[※52]もあり、AIによる更なる自動化が労働市場に小さくない影響を及ぼす可能性が懸念されている。

(b) 世界的なAI研究開発格差

2024年を起点に過去8年を振り返ると、最先端のAIによる学習にかかるコストは毎年2～3倍のペースで増加しており、2027年までに10億ドルを超える可能性があるとされる[※53]。Google社は2022年に米国ネブラスカ州に6億ドルをかけてデータセンターを建設[※54]し、2024年にはミズーリ州に10億ドル規模のデータセンターを建設する計画を発表した[※55]。2024年12月にMeta社は、AIの進歩を支えるためとして、100億ドル以上を投資して米国ルイジアナ州に新たなデータセンターを建設することを発表した[※56]。これらの費用を負担できる一部の企業と、そうした企業が所在する一部の国はAIの技術革新・サービスについて更に優位な立場に立ち、そうでない国や企業は相対的にその地位を低下させることになる。結果として、AIにまつわる経済格差が世界規模で拡大し、それに伴う社会不安等の問題をもたらすことが懸念される。

(c) 市場における一部企業への集中と単一障害点

上記で指摘した一部の企業によるAI投資の突出は、限られた企業に社会が強く依存するという集中と表裏一体である。2024年第4四半期の時点で、Amazon社、Microsoft社、Google社3社によるクラウドコンピューティング市場の寡占率は3分の2を超えている[※57]。AIを活用するITシステムの構築や運用には大規模なクラウドデータセンターが欠かせないため、クラウド基盤の集中はAI基盤の集中と実質的に同義である。このことに伴う問題として、これらAI基盤及びその上で稼働するAIの故障や動作不良が、全世界に影響を及ぼし得る点が挙げられる。システム全体の安全性が一部の構成要素に左右されるときに、その構成要素を単一障害点[※58]と呼ぶ。特定企業への世界中の利用者からの依存は、そ

れらの企業を社会的な単一障害点にしてしまうおそれがある。

(d)環境へのリスク

データセンターとデータ送受信で消費されるエネルギーは、汎用的 AI 登場前の 2020 年の時点で、世界のエネルギー使用に関連する温室効果ガス排出量の約 0.7%を占めた※59。国際エネルギー機関(IEA : International Energy Agency)の推定によれば、2024 年から 2026 年にかけて世界の電力需要は 3.4% の増大が見込まれ、データセンターの急拡大の影響が大きいと指摘されている※60。同様に IEA では、2023 年から 2030 年にかけて、データセンターは世界の電力需要の増加の 10% 弱を占める可能性があると予測している※61。データセンターにおける電力消費のすべてが AI に由来するわけではないが、例えば Google 社においては、同社のデータセンターの 2023 年のエネルギー消費量が前年と比較して 17% 増加し、同社全体での温室効果ガス排出量も 37% 増加した※62。同社では AI が全製品に深く結び付いているため AI の環境負荷だけを切り分けることは無意味かもしれないとしているが、近年の積極的なデータセンターへの投資の状況は既に述べたとおりである。同社は必要な電力を賄うために、次世代原子炉の開発企業と契約を締結した※63。Microsoft 社も、米国ペンシルベニア州のスリーマイル島原子力発電所を再開する契約を 2024 年に結び、20 年間、同発電所の発電能力をすべて購入することに合意した※64。これは約 80 万世帯の消費電力に相当する。日本においても、2025 年 2 月に決定されたエネルギー基本計画において、電力需要が今後増えていくことを念頭に置きつつ、再生可能エネルギーに加え原発の活用がうたわれることとなった※65。AI 利用の増大については地球環境への負荷というリスクも無視できなくなりつつある。

(e)プライバシーへのリスク

生成 AI は学習した内容に基づき何らかの応答を返す。学習データにプライバシー情報が含まれる場合には、AI の利用時に応答を通じてプライバシー情報の漏えいが生じる可能性がある※66。また、インターネット等に分散しているデータと、他のプライバシー情報を AI によりまとめて処理することで、予期せぬ目的で利用されてしまうことが懸念されている※67。進歩した AI の能力によって、膨大なデータの中から、それまでは見つけることが難しかった個人情報を特定したり推論したりできるのではないかとの指摘もある。その一方で、例えばヘルスケア産業※68や監視カメラ※69への AI 導入を通じて、今まで以上にプライバシー情報が IT システムに引き渡される機会が増している。AI を活用するネットサービス企業においてプライバシー保護が不十分であった事例※70や、家庭用防犯カメラのサイバーセキュリティ対策が不十分であった事例※71等が報告されている。リスクの度合いは判然としないものの、汎用的 AI の関わりによってプライバシー保護に問題が生じる可能性はあると見られている。

(f)著作権侵害のリスク

生成 AI は学習データに基づき様々な応答を合成するが、結果的に、学習データに含まれる著作物とほとんど変わらない内容を再現したり、あるいは、酷似する表現を用いた応答を合成したりする場合がある。学習データが原著者の許諾を得ていない著作物を含んでいれば、生成 AI による著作権侵害が起こる可能性がある。生成 AI の学習に用いられるデータセットの中に実際に著作者の許諾がないデータを含むものがあるとして、著作権者が権利侵害を訴える事例も既にある。2023 年 2 月に米国 Getty Images Holdings, Inc. は画像生成 AI Stable Diffusion の開発元である英国 Stability AI Ltd を訴え※72、同年の暮れには New York Times 紙が Microsoft 社を提訴した※73。世界中の Web サイトは学習データの重要な供給源であるが、この流れを受けて、それらの運営元において、AI 学習用途での閲覧やデータ取得を制限する動きもある※74。高性能な AI の開発に必要な学習データの量はますます増大しており、著作権侵害へのリスク対応と AI の発展がもたらす利益の間で適切に平衡を保つ議論の重要性が高まっている。

2.1.3 AIセーフティに関する取り組み

2023 年 11 月に、英国の公的機関として AI セーフティ・インスティテュート(AISI : AI Safety Institute)が設立され、日本を含む各国でも AISI あるいは相当する機関を設置する動きが広がった。英国 AISI の設立趣意書では AI セーフティ(AI 安全性)を「AI に由来する危害を理解し、予防し、緩和すること」と定義している※75。AI セーフティを具体化するための取り組みは多岐に渡るが、その中でも中心となっているのは、AI ガバナンスの実践である。以下では AI セーフティと AI ガバナンスの基本的な考え方を概観する。

(1) AI セーフティとは何か

冒頭の英国 AISI による定義は一つの例に過ぎず、我が国においては「人間中心の考え方をもとに、AI 活用に伴う社会的リスクを低減させるための安全性・公平性、個人情報の不適正な利用等を防止するためのプライバシー保護、AI システムの脆弱性等や外部からの攻撃等のリスクに対応するためのセキュリティ確保、システムの検証可能性を確保し適切な情報提供を行うための透明性が保たれた状態」を AI セーフティとしている※76。各国の AISI や、AI 開発を手掛ける民間企業におけるとらえ方は、次のように整理できる。

- 狭義の AI セーフティ：AI リスクを抑制する上で個々の AI 及び AI を内包する IT システムが満たすべき性質、及び実際にそれを満たしていること。具体的には、AI 出力のバイアスが十分に小さく公平であること、AI の判定の過程が事後的に精査可能・説明可能であること、サイバー攻撃に対して適切に防御されていること等。
- 広義の AI セーフティ：AI リスクを抑制・制御し、社会に対する害悪を低減するあらゆる取り組み及びその実践を通じて AI リスクが受け入れ可能な水準に抑制されていること。後述する AI ガバナンスの浸透、高いリスクを伴う用途への AI の導入に対する規制、AI リスクに対抗するための防御における AI 利用の推進等。

狭義の AI セーフティは主に技術的対策で担保されるのに対し、広義の AI セーフティはマネジメントやガバナンスといった社会的対策で担保される。技術的対策には、AI の動作を混乱させる入力を学習段階で取り入れて耐性を付ける敵対的学習※77 の実施や、学習データの品質の十分な精査、AI の応答を別の AI で監視すること等が含まれる。本項では技術的な議論には立ち入らず、後者の社会的対策に焦点を当てる。

(2) AI ガバナンスの概要

AI の利用にあたってはリスクベースアプローチが国際的に広く共有されている。AI の利用におけるリスクベースアプローチとは、「予め事前に当該利用分野における利用形態に伴って生じうるリスクの大きさ（危害の大きさ及びその蓋然性）を把握したうえで、その対策の程度をリスクの大きさに対応させる」方法を指す※7。言い換えれば、リスクマネジメントの一種として AI の利用を管理することにほかならない※78。リスクマネジメントを大きな柱として AI 利用を進める取り組みは「AI ガバナンス」と呼ばれている。以下では組織的なプロセスである AI ガバナンスと、AI ガバナンスの基礎となる AI セーフティ評価について概説する。

(a) AI ガバナンス

前述のとおり、AI ガバナンスの大きな柱は AI 利活用を対象領域としたリスクマネジメントである。リスクマネジメントとは次の四つのステップの反復を中核的プロセスとする PDCA 型のマネジメント手法であり、リスクを抑制しつつメリットを最大化することを目的とする※79。

- リスク特定：組織の目的達成に影響し得るリスク事象を洗い出す。マネジメントの対象範囲内で生じる可能性のある事件や事故を書き出すことに相当する。
- リスク分析：個々のリスク事象について、その発生や進展のメカニズム、発生の見込み・改善性、影響の度合い・詳細（インパクト）等を明らかにする。
- リスク評価：リスク分析の結果に基づき、現在のリスク水準が受け入れ可能なものかどうかを判断する。
- リスク対応：受け入れ不可能な水準のリスクに対し、事件・事故の予防策や影響の低減策等の対策を講じる。

AI ガバナンスのプロセスには、AI 利活用に関連したリスクの把握が欠かせない。リスク特定・分析・評価にわたる 3 ステップをまとめてリスクアセスメントとも呼ぶ。リスクアセスメントは、AI 利活用に伴うリスクを認識し、対応を決定するための準備作業にあたる。

代表的なリスクアセスメント手法の多くは IEC/ISO 31010※80 で紹介されているが、AI システムの利用を念頭に置いたリスクアセスメント手法には定番と言えるものがまだ存在しない。AI システムを対象とした数少ないリスクアセスメントの手引きとしては、一般社団法人日本品質管理学会がまとめた「AI リスクアセスメントガイドブック※81」がある。現状では、リスクアセスメント手法として実績のあるものを用いつつ、AI に特有の事情を勘案しながら、各自で工夫して取り組みを進めていくケースが多いと考えられる。

AI ガバナンスの全体像を整理した手引きとしては、米国国立標準技術研究所（NIST：National Institute of Standards and Technology）のまとめた「AI Risk Management Framework※82」（以下、AI RMF）が著名である。AI RMF は多様な AI 全般を対象としているが、生成 AI に特化した注意事項をまとめた追加文書「生成 AI プロファイル※83」もよく知られる。AI RMF に相当する日本の手引きとしては、経済産業省・総務省によ

る「AI 事業者ガイドライン※84」がある。EU で AI 利用を規制する法律として制定された「AI Act」も、AI ガバナンスの手引きとしての側面を持つ。なお、日本 AISI と米国 NIST は、AI 事業者ガイドラインと AI RMF の比較を行い、相互運用性を確認するクロスウォーク活動※85を 2024 年度に実施している。

(b) AI セーフティ評価

AI を内包する IT システムを「AI システム」と呼び、AI システムの心臓部にあたる AI は「AI モデル※86」と呼ぶことが多い。AI モデル単体がソフトウェア部品として提供されることも多いため、両者の区別は重要である。

AI モデルと AI システムの区別は AI セーフティの観点でも重要である。狭義の AI セーフティにおいて満たされるべき性質は、それが AI モデルにおいて満たされるべきなのか、AI システムにおいて満たされるべきなのかという観点の違いにより、検証方法や安全性確保の方法が異なってくる。

AI システムが AI セーフティの観点で適切であるかどうか見定めることを「AI セーフティ評価」という。AI セーフティ評価は基本的には狭義の AI セーフティに関する評価であり、AI システムとその内部の AI モデルが評価対象となる。AI セーフティ評価の切り口を整理した資料としては、日本 AISI がまとめた「AI セーフティに関する評価観点ガイド※87」がある。

「2.1.1 AI の急速な発展」で述べたように、AI の実装技術には複数のアプローチがあり、評価手法はそれぞれに異なる。昨今注目を集めているのは汎用的 AI として提供されることも多い LLM（Large Language Model：大規模言語モデル）※88であるが、LLM を対象としたセーフティ評価の基本は、一般的には機械的なベンチマークテストの実施である※89。具体的には、用意した膨大な量の問題を解かせ、正答率を調べるというものである。クイズの回答により、民族や人種に対する差別的なバイアスの有無や、マルウェア作成依頼を拒否するといった安全対策が機能しているかどうかの確認を行うものであり、自動的に実行できる特徴と相まって、ベンチマークテストは AI セーフティ評価の主流となっている。

ベンチマークテストは主に AI モデルの評価手法であって、AI システム全体の評価とは異なる。例えば、内包する AI モデルには問題がないが、AI システム内で動くデータベース部分にサイバー攻撃に対する脆弱性がある場合、AI モデルのセーフティ評価だけではこれを見落とすことになる※90。また、大半のベンチマークテストは一問一答形式であり、複数回の応答を経た後に AI の動作がおかしくなるようなケースの特定にも適さない。

AI セーフティ評価の焦点を AI モデルから AI システムや利用者等に広げ、実利用に近い状況下でのテストを攻撃者の目線で個別に組み上げて実施する手法は、「レッドチーミングテスト」と呼ばれる。この呼び名は、サイバーセキュリティ分野で行われるサイバー攻撃模擬演習において攻撃側をレッドチーム、防御側をブルーチームと呼ぶ慣習にならったものである。レッドチームは AI システムや利用業務、サイバーセキュリティ等のエキスパートで編成し、それらエキスパートの専門的知見を活用することで、状況に則した高度な攻撃シナリオを組み立てる。そのシナリオに沿って模擬的に AI システムのセーフティが損なわれる攻撃を行い、リスク要因や改善点を洗い出す。レッドチーミング※91 テストの進め方については、日本 AISI のまとめた「AI セーフティに関するレッドチーミング手法ガイド※92」が詳細な手引きとなっている。

AI セーフティ評価を通じて問題が見つかった場合には、何らかの対処を施すことになる。技術的な対処については、NIST や日本の AI プロダクト品質保証コンソーシアム、国立研究開発法人産業技術総合研究所が、AI モデルへの攻撃と防御の手法を文書にまとめている※93。加えて、AI の動作を大きく左右する学習データの品質確保の手引きとして、日本 AISI から「データ品質マネジメントガイドブック※94」が公開されている。

現行の AI セーフティ評価の手法は万全ではなく、今知られている手法の範囲で問題が見つからなかったとしても、未知のリスクが残る可能性がある。また、汎用的 AI の動作が解明されていないことや、応答が状況に依存することに加え、AI セーフティの評価観点を決定する際に、多様なステークホルダーが関与できておらず、観点の洗い出しが行き届いていないという批判もある※95。多様なステークホルダーをどのように参加させるかというプロセスにも定まった議論はまだない。AI の技術が毎月のように進歩し続けていることも念頭に置き、今後の AI セーフティ評価手法の進展を注視しつつ、今できることを疎かにしない姿勢が重要と言える。

(3) AI セーフティに関する国際連携

汎用的 AI は、大手ベンダーによりクラウドサービスの一種として提供されることが多い。AI サービスは国境を越え、一つのサービスに複数の国のルールや事情が関わってくる。このため、AI セーフティに関しても国際的な連携や協調が望まれ、各国の AISI も深く関わる形で取

り組みが進みつつある。以下では AISI を含む国際会議である AI セーフティサミットについて述べる。

(a) AI セーフティ・インスティテュート

前述のとおり、2023 年 11 月に英国において設立された、AI セーフティを担う公的機関が AI Safety Institute（AISI）[※96] である。発足当時、英国 AISI は以下の任務にあたるものとされた[※75]。

- 先進的な AI システムに関する評価手法の開発・実施：AI セーフティに関連する AI の能力の特徴付け、システムのセーフティとセキュリティの理解、社会的影響の評価を目指す。
- AI セーフティに関する基礎研究の推進：様々な探索的研究プロジェクトを立ち上げ、外部の研究者を招集する。
- 情報交換の促進：任意参加と既存のプライバシー及びデータ規制遵守を前提として、政策立案者、国際パートナー、民間企業、学術界、市民社会等との明確な情報共有チャネルを確立する。

上記の内容からは、科学的な裏付けを重視しつつ AI セーフティ評価を基盤として内外の連携を図るという全体像が見える。英国 AISI は、AI モデルのセーフティ評価ツールである Inspect[※97] の開発・公開、Inspect を用いた AI セーフティ評価の結果報告[※98] 等、技術的な道具の整備と知見の展開に取り組んでいる。

他方で、AI セーフティへの関心を持つのは AISI だけではなく、考え方も一つではない。EU の「AI Act」は人命への危害や人権侵害につながる特定用途の AI 導入を禁止し、また、ハイリスク AI システムという区分による規制強化等、リスクベースアプローチの徹底を求めている。人権侵害を問題視する点は、2021 年に UNESCO（United Nations Educational, Scientific and Cultural Organization：国際連合教育科学文化機関）で採択された「Recommendation on the Ethics of Artificial Intelligence（人工知能の倫理に関する勧告）[※99]」、2023 年に発表された米国「AI 権利章典[※100]」も同様である。また Biden 政権下で 2023 年 10 月にまとめられた大統領令 EO 14110[※101] は、サイバー攻撃や CBRN 兵器開発支援への悪用等の国家安全保障上のリスク対応に言及している。加えて、国際連合（United Nations 以下、国連）がまとめた提言[※102] においては、AI の利用が先進国に集中することに伴うグローバル格差の拡大を懸念してもいる。英国 AISI が示した、技術的な観点を主とする AI セーフティ評価がこれらの基礎になることは確かであるが、議論の裾野は幅広く、AI セーフティ確保のために必要な取り組みも現時点で定まったものはない。

英国に続き、各国において AISI またはそれに相当する公的機関を設立する動きが見られる。米国は英国とほぼ同時に、NIST 内に AISI を設立した[※103]。2024 年 2 月には日本 AISI が設置された[※104]。その他、シンガポール[※105]、カナダ[※106]、韓国[※107] 等が AISI を設けている。フランスは AISI に相当する機関として INESIA（Institut national pour l'évaluation et la sécurité de l'intelligence artificielle）を 2025 年 2 月に設置した[※108]。EU には「AI Act」に関連する実務機関として European AI Office[※109] が設置され、各国の AISI のカウンターパートとして活動している。

これら AISI の動きに対しては大手 AI ベンダーらも協調姿勢を示している。米国では、AI 関係の民間企業等を集めた AISI コンソーシアムを招集し、大手 AI ベンダーを含む 200 以上の関係組織が参加している[※110]。米国政府は自発的コミットメント[※111] という形で AI ベンダーらの協力姿勢を取り付けている。各国の AISI を含む国際会議である AI セーフティサミットにはこれらの企業も参加し、歩調を合わせている（「2.1.3 (3) (b) AI セーフティサミット」参照）。

AI セーフティに関するこれら公的機関及び AI ベンダーらが国際的な交流・連携を深め、世界で調和の取れた AI セーフティの取り組みが進展することが期待される。

(b) AI セーフティサミット

英国 AISI の設立は、2023 年 11 月初頭に英国 Bletchley Park で開催された AI セーフティサミット（AI Safety Summit）[※112] にて発表された。当時の Rishi Sunak 政権の主導によるこの国際会議には、米国、イタリアをはじめとした G7、EU 等が参加したほか、OpenAI 社 CEO、DeepMind 社共同創立者、米国の Elon Musk 氏等、AI 産業における世界的なキーパーソンも集まった。同 AI セーフティサミットでは締めくくりに、AI リスクに対応するための国際連携を呼びかける「ブレッチリー宣言[※113]」を参加 28 ヵ国と EU が採択した。

AI セーフティサミットはおおむね半年ごとに開催する予定となっており、2024 年 5 月下旬には第 2 回にあたる AI ソウル・サミット[※114] が韓国にて開催された。初回同様に G7 に EU、韓国、オーストラリアを加えて採択されたソウル宣言[※115] では、① 2023 年の広島 AI プロセス

の流れを汲む形で国際的に相互運用可能な AI ガバナンスの枠組みを整備すること、② AISI 及びそれに相当する機関の各国における設立を支援すること、③多種多様な利害関係者を包括していくこと等がうたわれた。更に、民間企業が誓約主体となる「フロンティア AI のセーフティに関するコミットメント※116」も採択され、OpenAI 社や Google 社のような AI 分野を率いる IT 企業自身がソウル宣言の趣旨に賛同し、積極的に AI セーフティ及び AI ガバナンスに取り組むことが示された。広島 AI プロセスから AI ソウル・サミットに至る討議の中で浮上した AI セーフティに関する論点については、日本 AISI が「AI セーフティに関する活動マップ（AMAIS）※117」として整理している。

(c) 2025 年冒頭の国際動向

2024 年 11 月、米国大統領選挙において当時の現職副大統領であった民主党の Kamala Devi Harris 候補を破り、共和党の Donald Trump 大統領が誕生した。共和党は同時実施となった上下院選挙のマニフェスト※118 において、AI のイノベーションを阻害するものであるとして、Biden 政権下でまとめられた EO 14110 を廃止することを確約しており、実際に 2025 年 1 月に同大統領令は廃止となった※119。これに代わるものとして「AI における米国のリーダーシップへの障壁を除く」大統領令 EO 14179※120 が署名されたが、詳細な行動計画は 180 日以内に決定するとしており、米国の AI 政策の動向には今後大きな変化が生じるものと見られる。

政権交代に伴う AI 政策の変化の兆候は英国にも見られる。2024 年 7 月の英国総選挙で保守党は敗北し、党首である Sunak 首相に代わって労働党の Keir Starmer 首相が誕生した。2025 年 1 月には Starmer 政権における AI 政策にあたる「AI 機会行動計画※121」が発表され、社会における AI 利活用を積極的に推進することがうたわれた。更に、2025 年 2 月に、英国 AISI は AI Safety Institute から AI Security Institute へと改称し、国家安全保障や犯罪に関連する AI セキュリティに取り組むことが発表された※122。AI セーフティの分野では重要視されるバイアスや言論の自由といったテーマについては、今後の取り組みの対象外になるとしている。プレスリリースにおいては、国家安全保障はもとより重要なテーマであり、英国 AISI の活動が変化するものではなく明確化するに過ぎないとしつつも、英国のサイバーセキュリティ当局である NCSC（National Cyber Security Centre）や治安当局である内務省（Home Office）との連携強化が盛り込まれる等、重点課題の絞り込みはあったものと見られる※123。

AI ソウル・サミットに続く第 3 回の AI アクションサミット※124 は 2025 年 2 月にパリにて開催された。名前が変化したことからも察せられるように、サミット全体としては、AI の利活用に力点を置きつつ、100 ヵ国超から 1,000 名以上が参加した。国連の António Guterres 事務総長やインドの Narendra Modi 首相も参加する等、AI に関する国際政治の舞台にもなったと言える。他方で、同サミットでは「人類と地球にとって包摂的で持続的な AI※125」と題する共同宣言に対し、世界の AI リーダー国であるはずの英国と米国は署名しなかった。英国政府は国家安全保障とグローバル・ガバナンスに関する懸念から署名できなかったとし、サミットに出席した米国の James David Vance 副大統領は、成長を促進する AI 政策がセーフティよりも優先されるべきだと述べた※126。

これらの動きが AI セーフティに関する取り組みの変化であり、それが他国にも波及するのかはまだ定かでない。その中でもある程度、議論の収れんが見られるのは、AI リスクの議論における、想定リスクとその大まかな特徴の整理である。これを示す成果物が、過去 3 回の AI サミットで公開された、独立した学術調査グループによる 3 編の報告書※127 である。これらの報告書は汎用的 AI を念頭に置き、最先端の AI の能力、学術的な解明状況、AI リスクの現状を整理している。「2.1.2 (1) AI の悪用がもたらすリスク」では、2025 年版の同報告書で取り上げられている AI リスクを紹介した。一方、整理された AI リスクがどのような規模・メカニズム・影響・確度で実現するかは依然として不透明であり、警戒を怠ることはできないと言える。

2.1.4 AIセキュリティの現状

AI セキュリティという言葉には、未だ明確な定義がないのが現状である。しかし、AI の開発や利活用において、サイバーセキュリティに関わる多くの懸念や課題があることは明らかである。以下では、AI が関わるサイバーセキュリティの動向を論じる。

(1) AI セキュリティとは何か

英国 AISI は当初の設立趣意書の中で AI セキュリティを定義し、AI モデルや AI システムを悪意ある攻撃者による攻撃から保護すること、としている※75。AI システムのサイバーセキュリティについては同様の考え方が「AI

Act」の中で説明されており、AIシステムのサイバーセキュリティ対策は、広く受け入れられている狭義のAIセキュリティであると言える。

上記に言うAIセキュリティには、AIの悪用で社会に生じるサイバーセキュリティリスクが含まれない。しかし、既に事例があるように、サイバー攻撃の手間を汎用的AIによって軽減し、攻撃者が優位性を高めることができる。ChatGPTの登場直後からフィッシングメールの劇的な増大も確認されており[※128]、サイバー犯罪の増加の背後に汎用的AIが関わっていることも推認される。悪用目的で安全対策をあらかじめ排除したAIをサイバー犯罪者らが独自に開発することも技術力があれば可能である[※129]。

AIガバナンスは、AIシステムを守ろうとする人々にとって有益であるが、悪意あるアクターがAIを悪用することは防げない[※130]。AIシステムをサイバー攻撃から守るという狭義のAIセキュリティでは網羅されない課題があると言える。Biden政権（当時）で導入されたEO 14110のセクション4ではまさにAIの悪用に対する対策に焦点が当てられ、AI開発に用いられるクラウド環境の利用状況監視や、脆弱性の特定と緩和にAIを活用する取り組みが盛り込まれていた。

本項では、AIシステムのサイバー保護を狭義のAIセキュリティとし、サイバー空間におけるAIの悪用がもたらすリスクから社会を保護する取り組みも加えたものを広義のAIセキュリティとする。

(2) AIセキュリティ脅威の動向

サイバーセキュリティリスクの発現の起点となり得る出来事や状況を「サイバー脅威[※131]」と呼ぶ。本項ではAIセキュリティにまつわるサイバー脅威を「AIセキュリティ脅威」とし、これを大きく三つに分類し、その全体傾向を以下にまとめる。なお、以下のまとめは、ENISA（European Network and Information Security Agency）の年次報告書[※132]や、AIセキュリティに関連する調査を活発に実施している英国Alan Turing Instituteの調査結果[※133]を総合したものである。

(a) AIシステムへのサイバー攻撃

大手ITベンダーAIチャットサービスに脆弱性が発見されたとする事例[※134]はあるが、社会に大きく被害を及ぼす大規模なサイバー攻撃事例は2025年3月末時点ではない。しかし、AIモデルの実行や学習基盤として広く利用されるPyTorchというライブラリにサプライチェーン攻撃を可能にする脆弱性が見つかった事例[※135]や、様々なオープンソースAIモデルの共有サイトHugging Faceに侵入可能な脆弱性が見つかった事例[※136]、中国発の高性能AIとして話題を呼んだDeepSeekのセキュリティに問題があり、背後のデータベースに侵入できることが判明した事例[※137]等がある。これらの事例は、AIシステムへのサイバー攻撃の被害は顕在化していないが机上の空論ではなく、AIに特化した攻撃手法よりも従来のサイバー攻撃手法に対して防御が不十分であるという現状を示している。

(b) AIを悪用したサイバー攻撃

ChatGPTをサイバー攻撃のアシスタントとして悪用した事例について、OpenAI社とMicrosoft社は2024年2月に詳細な調査報告[※27]を公開した。悪用に関わったアクターの数は300を超え、そのうちの160は国家支援型アクターで、50はランサムウェア攻撃を行うグループであったとされる。これらのアクターは、攻撃対象に関する下調べや、攻撃工程を効率化するプログラムの実装、侵入経路になる脆弱性に関する技術情報の収集等にChatGPTを活用した。ChatGPTは言わばプログラマーとしてのサイバー攻撃者の日常業務を効率化した。AIによるサイバー攻撃の効率化は、低スキルの人員にも攻撃能力を与えてしまうという点で、サイバー犯罪市場への参入障壁を引き下げていることが懸念されている[※28]。

The MITRE Corporation（以下、MITRE）はAIによるサイバー攻撃の自動化能力を測定するベンチマークテストを開発し、状況の継続的な観測に取り組んでいる。MITREの評価[※26]によれば、現時点では、汎用的AIの能力はサイバー攻撃を全面自動化するには不足している。他方で、サイバー防御の自動化については一部限定的な用途における実証試験が始まるという。サイバー攻撃も含め、技術の進展次第で状況が急激に変化する可能性がある。

(c) AIを用いた認知領域への攻撃

サイバー犯罪の領域では汎用的AIの悪用と見られる状況が広く確認されている。ChatGPTの登場以後、フィッシングメールは爆発的な増大を示した[※138]。ディープフェイク画像・音声等を用いた詐欺の事例も複数確認されており、ある会社でCFO（最高財務責任者）になりすまして約40億円を不正送金させ詐取した標的型ソーシャルエンジニアリングとも言うべき事例[※139]がある。更に新たな悪用例として、チャット越しにLLMが詐欺師として

振る舞った国内の事例※140 等が報道されている。

選挙を攪乱させる影響工作が世界中で見られたことは既に「2.1.2 (1) (b) 世論操作」で述べたとおりである。影響工作の背後には国家支援型アクターがいると見られており、中国、ロシア、イラン、イスラエル等が調査報告書※25 において支援国とされている。影響工作の手段としては AI が生成したテキスト、ディープフェイク画像・動画等が用いられ、SNS がフェイク情報の拡散手段になることが多い。現状では SNS 上でのフェイク情報の拡散が進まず、大きな影響を及ぼすには至っていないと評価されている。この原因は明確になっていないが、攻撃側の展開した情報に不備がありフェイクとすぐに分かるようなものになっているといった稚拙さがあったほか、SNS 上で生身の人間が展開するコンテンツとの競合に打ち勝つ必要がある中で、関心の獲得に失敗したという構図が指摘されている。

影響工作の今後を示唆する事例としては、2024 年 11 月に実施されたルーマニア大統領選挙がある。同選挙では、親ロシア極右で無名の Călin Georgescu 候補が第一回投票で首位となったが、未申告の選挙運動資金の提供や影響工作があったとして、憲法裁判所により選挙は無効であるとされた※141。ルーマニアの諜報機関から公開された機密文書※142 によれば、ロシアによる影響工作が続いており、AI も駆使されてきたと指摘されている。ただし、この選挙における影響工作では動画投稿・共有サービス TikTok 上での世論形成が大きな役割を果たしており※143、この種の SNS において広範な影響力を持ついわゆるインフルエンサーの買収も行われている※144。影響工作の中で AI の果たした役割が選挙をどの程度左右したのかは不明だが、影響工作が複数の手段を複合して展開され、全体として一国の民主主義を操作する直前にまで至った現実の事例であると言える（事例の詳細については「2.2.3 (2) (d) ルーマニア大統領選挙」参照）。

(3) AI セキュリティリスクに関する予測

英国 NCSC は 2024 年 1 月に「サイバー脅威に対する AI の短期的影響」と題する報告を公開した※28。要点は次のとおりである。

- AI の影響でほぼ確実に今後 2 年間でサイバー攻撃の量が増加し被害内容も悪化する。
- 国家、非国家、熟練者、そうでない者、あらゆるタイプのアクターが、程度の差こそあれ、既に AI を利用している。
- 偵察とソーシャルエンジニアリングにおいて、AI は両者をより効果的、効率的、かつ発見されにくいものにしている。
- AI は、サイバー犯罪初心者、雇われハッカー、ハクティビストが効果的なアクセスや情報収集活動を行う際の障壁を引き下げる。結果として、今後 2 年間のランサムウェア脅威は助長される可能性が高い。
- 2025 年までに限れば、新しいサイバー脅威が生まれるのではなく、従来のサイバー脅威が悪化する。
- サイバー作戦における AI のより高度な利用は、高度な能力とリソースを有する一部の脅威アクターに限定される可能性が高く、かつ、2025 年までに実現する可能性は低い。
- 2025 年以降、サイバー犯罪市場や犯罪ビジネスにおいて AI を活用した能力が一般化し、より高度な能力が利用可能になることはほぼ間違いない。

その他、Alan Turing Institute 配下の CETaS（Centre for Emerging Technology and Security）が 2024 年 7 月に公表した「悪用時の生成 AI の能力評価※145」では、マルウェア作成の自動化、影響工作による世論の先鋭化、テロにおける兵器開発支援と計画立案という 3 領域について評価を行い、報告時点では危険性は限定的であるとしている。その一方で、AI の進歩がこれらの脅威に及ぼす影響について、技術的及び社会技術的な変曲点を見極めることが重要であるとしている。急激な AI の発展がより一層高速化するようなことがあれば、対策が手遅れにならないようその機をとらえるべきとの趣旨である。

(4) AI セキュリティの具体的な対策

AI システムをサイバー攻撃から保護するという狭義の AI セキュリティは AI セーフティに包括される。このため、NIST の AI RMF や我が国の「AI 事業者ガイドライン」に沿った AI ガバナンスの推進を枠組みとして、その中でレッドチーミングテストを含む AI セーフティ評価を行い、適宜対策を講じることが基本となる。加えて、セキュアシステム開発のガイドラインである NIST の「Secure Software Development Framework（SSDF）※146」には生成 AI 利用を想定した追加文書※147 があるほか、英国 NCSC が中心となって発行された「機械学習の原則※148」、更に英国、米国、カナダ、オーストラリア、ニュージーランドのサイバーセキュリティ当局が共同でまとめた「AI システムのセキュア配備※149」、そして「2.1.3 (2) (b)

AI セーフティ評価」で取り上げた手引き類がある。AI モデルの学習工程で考慮すべき対策を除けば、内容の大部分は、一般的なサイバーセキュリティ対策の強度を高めたものとなっている。

狭義の AI セキュリティの確立にあたっては、AI ガバナンスとセキュリティガバナンスの関係をどう整理するかという問題がある。AI ガバナンスの観点からは AI セーフティの一部に狭義の AI セキュリティが含まれる一方で、IT システムの一種でもある AI システムの保護はサイバーセキュリティ上の取り組みでもある。このため、AI システムの保護の主たる責任が AI ガバナンスとセキュリティガバナンスのどちらに置かれるのかという点に何らかの整理を行う必要がある。しかし、それぞれの分野で個別のガイドラインの整備等が進む一方で、両者の統合に関する議論が未成熟であることが指摘されている※150。

AI の悪用を視野に入れた広義の AI セキュリティの対策については、本節の執筆時点では標準的と言えるものがない。AI の開発に必要な基礎的情報や学習データの多くは公開されているため、悪用目的での AI モデル・AI システム開発を防ぐ方法は根本的にない。この点では、AI をセキュリティ強化の目的で積極活用する AI for Security の考え方が有望であると期待される。加えて、狭義の AI セキュリティ対策を徹底することが、通常の目的で利用される AI システムの乗っ取りと犯罪等への転用を防ぐという意味でも不可欠と言える。全体として広義の AI セキュリティの議論は煮詰まっておらず、今後の推移を注視していく必要がある。

COLUMN

サイバーセキュリティとデジタルトランスフォーメーション ～WISDOM-DXと生成AIによる「情報セキュリティ白書」の分析～

福岡伸一氏は、著書「世界は分けてもわからない[i]」の中で、鼻の移植手術のために鼻を取り出すとしたらどこまで深くえぐりとればよいかという思考実験を紹介しています。鼻と呼ばれる突起物だけでなく、鼻の奥の嗅上皮、嗅上皮のレセプターと神経線維、脳の嗅球、神経細胞群、連動する筋肉や骨や関節の仕組みというように嗅覚という機能を切り出すためには、結局、身体全体を取り出すしかない。「部分とは部分という名の幻想である」と述べています。同じように、セキュリティをデジタルトランスフォーメーション（DX）の「部分」として切り出すのも幻想でしょう。

例えば、医療 DX を例に挙げると、システムとして、給食等の供給系、電子カルテ等の基幹系、放射線等の医療機器系、診療予約等の患者サービス系が連携しており、利用者は、供給事業者、医療従事者、病院利用者、それぞれのシステムの保守要員等と多岐にわたります。そこでは利用者によるミスや不正、外部からの攻撃といった様々なインシデントが報告され、一部門のインシデントがシステム全体に影響を及ぼす結果となっています。医療以外の多くの分野においても同様です。そのため、安全管理ガイドラインの策定やサイバーセキュリティ対策の注意喚起、最高情報セキュリティ責任者（CISO）等の任命、ならびに緊急対応体制（CSIRT 等）の整備が進められていますが、担当部門や専門家に任せておけばよいというものではありません。利用者すべてが IT やサイバーセキュリティの基本知識を学ぶ必要があります。例えば、標的型攻撃の手口を知らない利用者が電子メールを利用することは大きなリスクです。

基本的な IT 知識を身につける手段として国家試験[ii]「IT パスポート試験」が注目されています。2024 年度、事務職・営業職等幅広い職種の社会人や学生から 30 万人以上の応募がありました。サイバーセキュリティの基本知識の習得には「情報セキュリティマネジメント試験」も有効です。業務で個人情報を扱う担当者、業務部門・管理部門の情報管理担当者、外部委託先に対する情報セキュリティ評価・確認を行う担当者すべてが、知っておくべき内容です。CISO には、「情報処理安全確保支援士」（登録セキスペ）の資格が望ましいです。少なくとも、登録セキスペの助言を理解し実践できる能力は必須です。DX 推進のためには、各人が業務に関連するセキュリティ対策を実践する能力、いわゆる「プラス・セキュリティ」が求められています。

経済産業省は、DX の促進を図るために、DX を推進する仕組みを社内に構築し、優れたデジタル活用の実績が表れている企業を銘柄や注目企業として選定しています[iii]。2015 年から 2024 年までに DX 銘柄、DX 注目企業、攻めの IT 経営銘柄、IT 経営注目企業等 453 社を「DX 優良企業」として選定しました。また IPA と NICT は、Web 情報を用いて企業の DX 活動を自動評価する「WISDOM-DX[iv]」を開発しています。今回、WISDOM-DX を用いて東京証券取引所の上場企業 1,984 社の DX 活動内容を抽出してベクトル化し、更に「情報セキュリティ白書 2024[v]」の各節の記載内容を生成 AI によってベクトル化しました。これらのベクトル間の距離を算出することによって、白書に記載された事項と各企業の DX 活動の内容の関連性を可視化することができます。横軸を東証上場企業 1,984 社、

縦軸を「情報セキュリティ白書 2024」の「1.1 2023 年度に観測されたインシデント状況」から「4.2 AI のセキュリティ」の各節として図 1 のような関連性ヒートマップを作成しました。

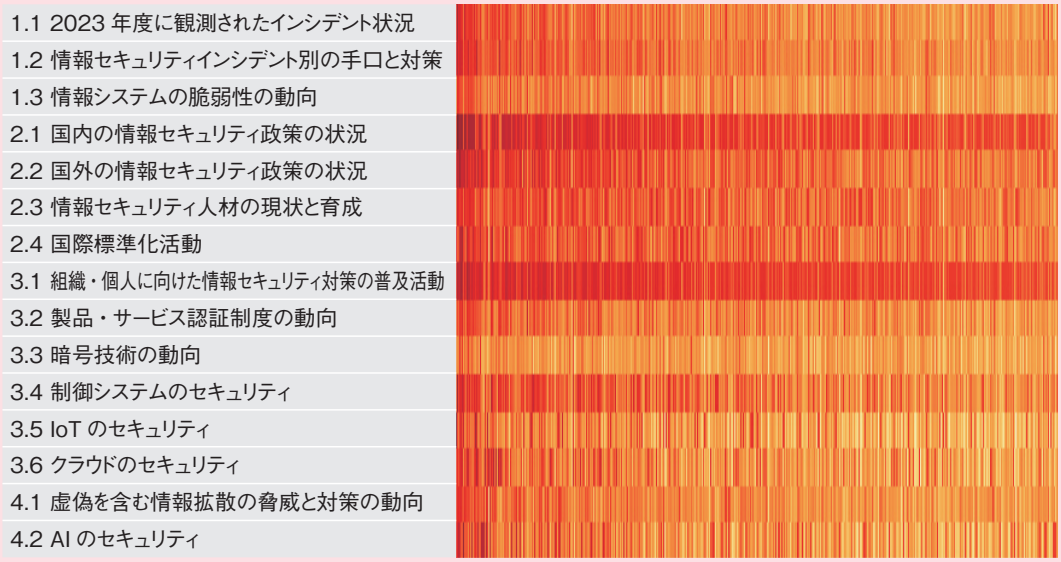


図 1　「情報セキュリティ白書 2024」の内容と企業の DX 活動との関連性

濃い色は、関連性が高いこと、つまり各企業が DX 活動として「情報セキュリティ白書 2024」の記載内容に関連する活動を行っていることを示します[vi]。多くの企業が、「2.1 国内の情報セキュリティ政策の状況」と「3.1 組織・個人に向けた情報セキュリティ対策の普及活動」に記載された内容に関連性のある活動を行っていることが分かります。次に、東証上場企業 1,984 社を「情報セキュリティ白書 2024」の内容と関連性の高い順に左から並べ、DX 優良企業 453 社を縦線としてプロットした分布図を図 2 に示します。

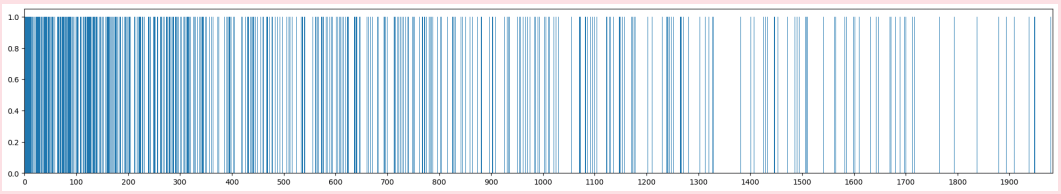


図 2　東証上場企業における DX 優良企業の分布

DX 優良企業の多くが左側に分布している、つまり、DX 優良企業は、「情報セキュリティ白書 2024」の記載内容に関連性の高い活動を行っていることが分かります。DX 優良企業の半数以上が、「情報セキュリティ白書」との関連性上位 20% に含まれています。この「情報セキュリティ白書 2025」も、企業の DX 活動を推進するために積極的にご活用いただければ幸いです。

i　福岡伸一「世界は分けてもわからない」講談社、2009 年
ii　IPA：試験区分一覧　https://www.ipa.go.jp/shiken/kubun/list.html〔2025/7/18 確認〕
iii　経済産業省：デジタルトランスフォーメーション銘柄（DX 銘柄）　https://www.meti.go.jp/policy/it_policy/investment/keiei_meigara/dx_meigara.html〔2025/7/18 確認〕
iv　IPA：WEB データに基づく企業の DX 活動の自動分析・評価システム「WISDOM-DX」を開発　https://www.ipa.go.jp/digital/dx/wisdom-dx/wisdom-dx.html〔2025/7/18 確認〕
奥村明俊、市瀬規善、久寿居大、石川開、鳥澤健太郎、大竹清敬：Web データに基づく企業のデジタルトランスフォーメーション活動評価システム WISDOM-DX の検証　https://proceedings-of-deim.github.io/DEIM2023/4b-8-2.pdf〔2025/7/18 確認〕
v　https://www.ipa.go.jp/publish/wp-security/2024.html〔2025/7/18 確認〕
vi　薄い色は「情報セキュリティ白書 2024」に記載の内容に関連する DX 活動が WEB データから読み取れないことを意味する。

2.2 偽・誤情報の脅威の動向

本節では、虚偽を含んだ情報の拡散の国際的な情勢を概観しつつ、主要な各国事例を取り上げて解説を行う。

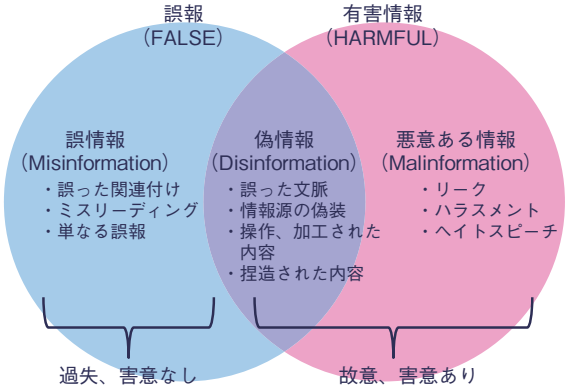
国家が関与するサイバー攻撃は、その手法等から分類すると表 2-2-1 のように整理できる。近年の情報戦・認知戦においては、虚偽を含んだ情報を流布する「情報操作型サイバー攻撃」と、それを含む様々な類型のサイバー攻撃を組み合わせた「ハイブリッド型サイバー攻撃」が大きな脅威となっている。本節では、表 2-2-1 の攻撃類型を使用して以降の解説を行う。なお、誤情報（ミスインフォメーション）と偽情報（ディスインフォメーション）を組み合わせた政府の用法※151 に従い、合わせて「偽・誤情報」と表記する。

2.2.1 虚偽情報の定義

虚偽を含んだ情報の拡散による社会の混乱（情報騒乱）については、2017 年に欧州評議会（CoE：Council of Europe）が用語の整理を行っている（図 2-2-1）。

この整理による各用語の定義は以下のとおりである。

- ミスインフォメーション（Misinformation、誤情報）：事実誤認や過失により誤解を招く文脈で発信される、故意や悪意のない誤情報。
- ディスインフォメーション（Disinformation、偽情報）：社会、公益への危害を目的とした害意のある情報。偽の情報だけでなく、誤った文脈や操作された内容で拡散される真の情報も含まれる。
- マルインフォメーション（Malinformation、悪意ある情報）：リークやハラスメント等、害意をもって広められる真の情報



■図 2-2-1　欧州評議会による情報騒乱（INFORMATION DISORDER）の分類
（出典）Claire Wardle, Hossein Derakhshan「INFORMATION DISORDER : Toward an interdisciplinary framework for research and policy making※153」を基に IPA が作成

ミスインフォメーションとディスインフォメーションの差異は情報の発信者の故意性と害意の有無にあり、ディスインフォメーションとマルインフォメーションの差異は情報自体の真偽性にあり、判断の視点がやや異なる。この分類においては、本来は誤ったニュースを指すに過ぎない「フェイクニュース」は、ミスインフォメーションまたはディスインフォメーションに含まれる。ただし、これらについて確

攻撃類型	攻撃の内容
①情報窃取型	標的型攻撃（マルウェア付きメール、水飲み場攻撃）等により、特定の政府機関、企業、団体、個人のネットワーク、コンピューターに侵入し、機密情報、営業情報、特許、知的財産等を窃取する攻撃。
②機能妨害型	DDoS 攻撃等の手法により、ネットワークの許容量を超える飽和通信要求によって、サーバー、ネットワークを麻痺させる攻撃。
③機能破壊型	標的型攻撃等により、特定の政府機関、企業、団体、個人のネットワークに侵入し、システム破壊・改ざんを行う攻撃。ネットワーク内のデータ消去・改ざんを目的とするものと、制御系システムを標的として物理的破壊を目的とするものがある。
④金銭目的型	標的型攻撃、脆弱性の悪用等により、特定の政府機関、銀行、企業、個人のネットワークに侵入し、不正な送金を行う、またはコンピューター内のデータを暗号化し、解読に身代金を要求する攻撃。
⑤情報操作型	代理主体（Proxy）等を用いて真の発信者を隠匿した上で、SNS 等に偽ニュースを流布させることにより、対象国（主に民主主義国）における世論操作を目的とした攻撃。選挙結果に影響を与えることを企図している攻撃も見られる。
⑥軍事的サイバー攻撃	軍事攻撃と一体的に行われる機能妨害・機能破壊を目的とした攻撃。電子戦の一環として軍隊の指揮統制（C4I）システムを標的とするものと、軍事行動に影響を与える重要インフラを標的としたものがある。
⑦ハイブリッド型	上記①～⑥までの類型を組み合わせた攻撃。近年は①情報窃取型＋⑤情報操作型、②機能妨害型＋⑤情報操作型等の組み合わせが多い。

■表 2-2-1　国家が関与するサイバー攻撃の類型と主な実行主体
（出典）大澤淳「サイバー領域の安全保障政策の方向性※152」を基に IPA が加筆・編集

定的かつ共通した国際的な定義はなく、特にディスインフォメーションについては定義に多少の揺らぎが見られる。日本国内の Disinformation 対策フォーラムでは、「Disinformation」を「あらゆる形態における虚偽の、不正確な、または誤解を招くような情報で、設計・表示・宣伝される等を通して、公共に危害が与えられた、又は、与える可能性が高いもの」と定義している※154。また、欧州対外行動庁（EEAS：European External Action Service）の 2023 年のレポート※155 では、Disinformation について「経済的利益を得るため、または意図的に公衆を欺くために作成、提示、流布され、公共に損害を与える可能性のある、検証可能な虚偽または誤解を招く情報」と説明している。公共の損害とは、「民主的な政治・政策決定プロセスや市民の健康、環境、安全保障等の公共財に対する脅威を指す。」と定義して、その目的の一つとして経済的利益に言及するとともに、その意図として公共への害意を明示している。日本語ではディスインフォメーションに「偽情報」という訳語があてられているが、ディスインフォメーションは単に虚偽の情報を含むだけではなく、相手の誤解を招くために真の情報も混ぜ合わせて加工や情報操作が行われる点に注意が必要である。

2.2.2 偽・誤情報の情勢

本項においては、偽・誤情報をめぐる世界的な情勢を概観し、2024 年度以前の調査等も踏まえつつ、近年の全体動向を解説する。

(1) 偽・誤情報の現在

偽情報（ディスインフォメーション）及び誤情報（ミスインフォメーション）の問題は、2024 年度も引き続き世界的な懸念事項として認識されている。

世界経済フォーラム（WEF：World Economic Forum）が毎年発表しているグローバルリスク報告書の 2024 年版では、偽・誤情報は「今後 2 年間で世界に最も大きなリスクをもたらし得る要因」として、異常気象や武力紛争といったその他の脅威を抑えてトップに記載された（表 2-2-2）。

2024 年に国政選挙が行われる 16 ヵ国を調査対象とした UNESCO による 2023 年の調査※157 では、85% の回答者が「オンライン上の偽情報の影響を懸念している」と答えた。また、87% が「既に偽情報が自国の政治に大きな影響を与えている」と回答しており、偽情報（ディスインフォメーション）はグローバルな民主主義の根幹に対する脅威として浮上している。

順位	グローバルリスク	分類
1	誤報と偽情報	テクノロジー
2	異常気象	環境
3	社会の二極化	社会
4	サイバー犯罪やサイバーセキュリティ対策の低下	テクノロジー
5	国家間武力紛争	地政学
6	不平等または経済的機会の欠如	社会
7	インフレーション	経済
8	非自発的移住	社会
9	景気後退（不況、停滞）	経済
10	汚染（大気、土壌、水）	環境

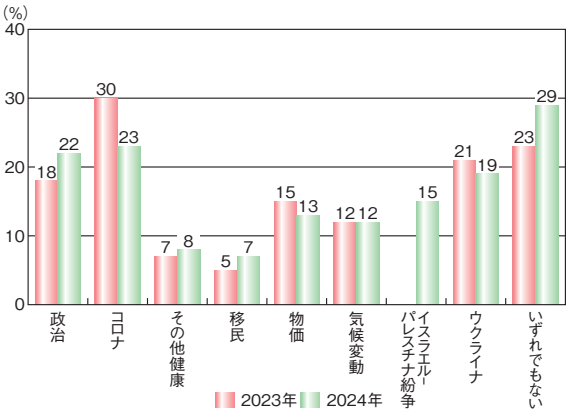
■表 2-2-2　今後 2 年間に想定されるグローバルリスク
（出典）WEF「第 19 回グローバルリスク報告書　2024 年版※156」を基に IPA が編集

SNS が情報源として機能している現代社会では、事実確認されていない情報の SNS 上の拡散が特に深刻である。同調査によると、提供される情報に対する信頼度は、「テレビ」が 66%、「ラジオ」が 63%、「Web メディアやメディアアプリ」が 57% だったのに対し、「SNS」は 50% と比較的低かった。フェイクニュースが最も広まっている場所を尋ねた結果では、回答者の 68% が「SNS」であると答え、「メッセージングアプリ」（38%）を上回り 1 位となった。一方で、ニュースや情報の主要な情報源を尋ねた結果では、「SNS」が 56% と最も高く、「テレビ」（44%）を上回っており、多くの国で 1 位となっていた。

メディアに関する国際比較調査※158 においても、日本を対象とした調査結果では、回答者の 58% が偽・誤情報の拡散に不安を抱いていると回答し、政治問題や新型コロナウイルス感染症（以下、新型コロナウイルス）、ロシア・ウクライナ戦争等に関連した偽・誤情報を日常的に目にしていることが分かっている。2023 年と 2024 年の調査結果を比較すると、政治問題や移民問題のトピックにおいて、偽・誤情報を見かける機会が増えているようである（次ページ図 2-2-2）。

更に生成 AI 技術の発展と一般市民への浸透によって、ディープフェイク等の技術を悪用した偽情報の拡散も、WEF の報告書では主要なリスクとして指摘されている。

G7 諸国、ブラジル、中国、インド、南アフリカで実施された世論調査である「ミュンヘン・セキュリティ・インデックス 2024」でも、「敵国による偽情報作戦」と「AI」に対するリスク認識が急上昇していることが示された※159。ミュンヘン安全保障会議の報告では、特に選挙における AI



■図 2-2-2　虚偽または誤解を招く情報を目にしたことのある話題（日本）
（出典）NHK 放送文化研究所「シリーズ「ロイター・デジタルニュースリポート 2024」（3）～偽情報・誤情報に対する意識は～【研究員の視点】#552[※158]」を基に IPA が編集

悪用の脅威が指摘され、警戒を呼びかけていた[※160]。後述のとおり、主要な各国選挙で AI の悪用も含めた偽情報の拡散が見られたところであり、2024 年は偽・誤情報の脅威が増大するとともに、一般市民にも脅威認識が拡大したといえよう。

(2) 偽・誤情報利用による安全保障上の脅威の拡大

偽・誤情報の中でも、とりわけ偽情報を中心とした悪意ある情報操作が近年の安全保障上の問題となっている。2014 年のクリミア危機以降、ロシアによるハイブリッド戦争を始めとして、サイバー空間を中心とした情報戦における、国家による情報操作が大きな課題となってきた。ハイブリッド戦争は、原義的には、非国家主体がゲリラ戦や生物兵器、サイバー攻撃や情報戦等あらゆる手段を、戦術的優位性を確保するために時と場所を選ばずに実行する新しい戦争形態を指す[※161]。そして IT 技術の進化や Web メディア、SNS の一般社会の浸透等により、ハイブリッド戦争におけるサイバー戦や情報戦のウェイトが増してきた。情報戦は、偽情報を流布することで相手国・地域の社会の混乱や政府機関の信用失墜を企図する「情報操作型サイバー攻撃」によって行われる[※162]。更には、単純に偽情報を拡散するだけではなく、その偽情報の積み重ねによって特定の国家に都合の良い戦略的なナラティブ（物語）を構築し、戦略的ナラティブに基づいた偽情報の拡散によって、一般市民も含めた相手国国民全体の価値観や信念、思考方法まで変容させてしまうことを、偽情報を流布する国々は狙っている。人間の思考方法や推論方法等の認知情報処理フローは、感覚入力された情報と、過去の記憶やイメージによって引き出される記憶系の情報との突き合わせによって反応が生成され、実際の行動が引き起こされるが、現在の IoT 社会では個人の視覚や聴覚等の入力センサーがサイバー空間と接続されたことにより、偽情報は「認知領域」における攻撃手段と見なされることとなった[※163]。その攻防は国家安全保障の文脈で「認知戦」と呼ばれる。例えば、偽情報を用いて選挙干渉を行い、有権者の認知を混乱させ、自国に有利な意思決定を誘導しようとするようなアプローチが代表的な攻撃手法である。特に対ロシアの情報戦に備えてきた欧州各国は、偽情報を用いた影響工作が民主主義制度への信頼を揺るがし、社会を分断することを警戒している。EEAS はこのような工作を「FIMI（Foreign Information Manipulation and Interference：外国による情報操作と干渉）」と呼称し[※164]、増大する安全保障上の脅威であると強調しており、ハイブリッド脅威への対策を加盟国と協力して進めている。ハイブリッド脅威とは、情報操作やサイバー攻撃、経済的威圧等武力行為以外の手段も組み合わせて目標の国家や組織を弱体化させることを目的とする有害な活動を指す[※165]。当然ながら、欧州だけでなく情報戦や影響工作の脅威に晒されている米国や台湾等もこのような情報操作に警戒の動きを強めているところである。EEAS による外国からの情報操作に関する報告書では、第 1 版において 2022 年に収集した事例 100 件の主要アクターについて、88 件がロシア、17 件が中国であったこと（両国の重なりがみられる事例が 5 件のため合計は 105 件となっている）が数量的に示されており[※166]、続く第 2、第 3 版でも中露を主要アクターに挙げている。

こうした脅威拡大に対し、対策を講じているのは国家だけではない。情報戦・認知戦の主戦場となる SNS 等を運営する主要プラットフォーム各社も、国家関与が疑われる影響工作キャンペーンの摘発を強化している。Meta Platforms, Inc.（以下、Meta 社）は 2024 年に 20 件の隠れた世論操作ネットワークを削除したと発表し、ロシア、イラン、中国が主要な発信元であると公表した[※167]。これらのネットワークは、偽の SNS アカウントや偽装ニュースサイトを用いて世論を誘導しようとするものであり、同社はその行動を「協調的不正行為（Coordinated Inauthentic Behavior）」と規定している。一方、Google LLC の Threat Analysis Group は、2024 年第 2 四半期だけで、中国関連の影響工作に関して最大で 3,931 件の YouTube チャンネルを停止し、1,177 件のブログのアカウントを削除したと報告している[※168]。また、同時期にロシア関連の 2,357 件の YouTube チャンネルも削除され

ており、プラットフォーム上での工作の拡大が続いていることがうかがえる。OpenAI, Inc.（以下、OpenAI 社）もまた、ChatGPT 等の自社モデルがロシア・中国・イラン等の国家アクターに利用されていたことを明らかにし、関連アカウントを停止したとする調査報告を公表している※169。2024 年、2025 年に連続して公表された報告書では、イランの「International Union of Virtual Media（IUVM）」に関連する影響工作を行っていた Storm-2035 という ChatGPT のアカウント群の存在を突き止め、その活動の一端として AI を用いた偽情報生成が確認されたことを明らかにした※170。一方で、米国大統領選戦後に大幅な方針転換を行った X（旧 Twitter）では、偽情報対策のラベル付与が縮小され、虚偽投稿の拡散が懸念された。英国サウスポートの暴動時※171（「2.2.3 (2) (e) 英国サウスポート暴動」参照）には、X Corp. のオーナーである Elon Musk 氏自身が偽・誤情報を助長する投稿に反応・拡散する場面も見られ、専門家から「最大のディスインフルエンサー」と批判される事態に至っている※172。一方、X では AI サービスである Grok を利用できるようになっており、それを使って、特定の投稿に対するファクトチェックを指示してリプライすることもできるため、ユーザーの AI 活用時の自主的な偽・誤情報対策の環境を提供している側面もある。

2.2.3 2024年度の注目事象

本項においては、各国における具体的な事例のうち、注目すべき事例を取り上げて解説する。

(1) 米国大統領選挙

2024 年の米国大統領選挙では、ロシア・中国・イランが関与すると見られる影響工作が確認された。米国の主要な情報機関で構成される IC（Intelligence Community）によれば、ロシアは偽のニュースサイト網や米国内の協力者を利用して世論を分断し、自国に有利な主張を広めたという※173。具体的には、ハイチ移民を装って違法投票を示唆する偽動画を作成する等のロシアの一連の情報工作は、選挙の公正さに疑念を生じさせ、信頼性を揺るがす狙いがあったと、米国国家情報長官室（ODNI：Office of the Director of National Intelligence）は FBI 及び米国サイバーセキュリティ・インフラストラクチャセキュリティ庁（CISA：Cybersecurity and Infrastructure Security Agency）との共同声明※174 で指摘している。更にロシアの関連組織はディープフェイク等の生成 AI 技術を駆使し、候補者に関する偽の映像や記事を大量生産して配信していた。「オペレーション・オーバーロード」という作戦では、偽情報の生成や拡散において生成 AI の使用が指摘されている※175。この作戦の特徴は、あえてファクトチェッカーやメディア、ジャーナリストに偽情報のファクトチェック依頼を送ることで彼らの能力に過負荷を与えることに加え、そのファクトチェック過程で偽情報と暴露されたとしても、偽情報や関連するナラティブの露出を高め、結果的に拡散を狙うという手法である。また米国財務省は 2024 年 12 月、ロシア連邦軍参謀本部情報総局と関係する地政学専門センター（CGE：Center for Geopolitical Expertise）が AI を用いて大統領候補者に関するディープフェイク動画を作成し、大規模な偽ニュースサイト網で拡散していたことを明らかにしている※176。CGE は AI 生成コンテンツを保管・配信する独自サーバーを構築し、少なくとも 100 以上の偽装ニュースサイトで偽情報を拡散していた。米国政府はこうした活動に関与したロシア人や組織に対し大統領令 13848 に基づく制裁措置を発動し、関係者を特定・資産凍結する対応を取った。

一方、中国は米国大統領選挙の結果自体には直接介入しない姿勢であると情報機関当局からは評価されているが、SNS を通じて米国社会の対立を煽り、民主主義は混乱していると印象付ける工作が報告された※173。ODNI によれば、中国当局は自国の技術企業と連携して米国向けの偽情報発信能力を強化し、X や TikTok を通じて米国内の左右両陣営に偽装アカウントを使って愛国心を装った分断情報を流していたという※177。例えば、中国発の影響工作「Spamouflage（スパムフラージュ）」では、米国有権者になりすました偽アカウント群を動員し、中絶や人種問題等論争になりやすい論点で相反する主張を大量投稿する手口が確認された。これら偽アカウントは与野党双方の政治家を批判し、特定の党派よりも「米国の愛国者」を名乗ることで信憑性を装っていた。米国当局は 2022 年の中間選挙でも超党派の反中政治家を標的に中国が中傷キャンペーンを行ったと指摘しており、この動向は 2024 年も継続した※173。もっとも、中国の狙いは米国世論の根本的な転向ではなく、米国内に潜む分断を拡大して民主主義の弱点を宣伝することにあったと分析されている。中国政府系の偽情報ネットワークでは、米国の選挙制度そのものよりも、「米国は内部分裂しており、国際的なリーダーシップやガバナンスを失ってグローバルパワーとして衰退しつつある」とのイメージ醸成に注力していた※177。

イランもまた米国大統領選挙に対する攻勢を強めた国の一つである。2024年夏、イランのイスラム革命防衛隊（IRGC：Islamic Revolutionary Guard Corps）系の組織が米国大統領選挙の両陣営関係者へのサイバー攻撃を試み、機密情報の窃取や漏えいによって選挙プロセスに不信を生じさせようとしたことを米国政府は明らかにした※176。実際に2024年9月には、ある大統領候補陣営の幹部らのアカウントがイランのハッカー集団により侵害され、選挙戦略に関わる内部資料が漏えいする事件が起きている※178。米国財務省はこの攻撃に関与したIRGC隊員や民間企業社員を特定し、選挙介入を理由に制裁リストに追加した。またイラン発の偽情報キャンペーンは米国の社会不安を煽る内容が多く、特に2023年10月以降のイスラエル・ハマス紛争に乗じた反イスラエル・反政府デマが米国内の分断拡大に利用された※173。米国情報当局は、イラン指導部が対イラン強硬派である特定候補の当選阻止を意図していた可能性を示唆しており、実際2024年米国大統領選挙でDonald Trump候補の陣営から機密文書が流出した背後にもイランの関与が疑われている※179。選挙後、米国司法省はイランがTrump氏暗殺を企図していたとする告発も行っており※180、イランはサイバー攻撃のみならず物理的脅威も含めたハイブリッド戦術で米国政局に影響を及ぼそうとしたと見られる。

こうした外国勢力の干渉に対し、米国政府とプラットフォーム各社は防御策を講じた。2016年以降の継続する選挙干渉に対し、2024年選挙を見据えて、ODNIに「外国からの悪意ある影響工作対策センター（Foreign Malign Influence Center）」を2022年に設立している。政府側では国土安全保障省（DHS：Department of Homeland Security）傘下のCISAやFBIが候補者や州政府に対し外国の偽情報活動に関する警告を発し、選挙インフラの監視を強化した※181。財務省外国資産管理室（OFAC：Office of Foreign Assets Control）による制裁発動のほか、主要IT企業も偽情報アカウントの削除やディープフェイク検知技術の導入を進めた。現時点では、2024年米国大統領選挙全体を決定的に左右したと考えられるような工作は確認されておらず、外国からの影響工作の影響は大きなものではなかったと分析されている※175。しかし選挙期間中に行われた爆破予告（ロシア発のドメインから発信）※179や大量の偽情報投稿は有権者の不安を煽り、一部投票所の混乱を招いた。

このように、米国大統領選では複数の国からの影響工作を受けており、外国勢力による選挙干渉は引き続き大きな脅威である。そして、これは同盟国を含む民主主義諸国にとっても、深刻な脅威をもたらす情勢となっている。

(2) 欧州での偽情報流布の影響（ロシアを中心とした影響工作と極右台頭）

2024年は欧州各国で重要な選挙が相次ぎ、ロシアを始めとする権威主義国による組織的な偽情報攻勢が各国で顕在化した年でもあった。6月には欧州議会選挙が27ヵ国で実施され、続いて英国やフランス、オーストリア等主要各国でそれぞれの国政選挙が行われた。その結果として、一部では極右政党が一定の勢力伸長を見せたものの、政治的中道勢力が踏みとどまり、外国からの影響工作や選挙干渉が直接選挙結果を左右する事態には至らなかったと当初評価されていた※182。しかしEEASや安全保障専門家らは、外形的に明確なインシデントというべき事案が発生していなくても、ロシアや中国による長年の偽情報攻勢が社会の結束を損ない、政治的分裂を悪化させようとして欧州民主主義に与えた影響を看過すべきでないとして、持続的な脅威として引き続き警戒を強めている※183。

(a) 欧州議会選挙

欧州議会選挙の前後には、ロシアを中心とする外国勢力による選挙干渉疑惑が相次いで浮上した。例えば、2024年3月にチェコ当局はプラハに本部を置くニュースサイト「Voice of Europe」を閉鎖した。このサイトは名称に反し親露派オリガルヒでウクライナ人のViktor Medvedchuk氏が立ち上げた偽装メディアであり、欧州各国の政治家に報酬を支払って親露的な政治宣伝をニュースを装って紛れ込ませていたと報じられた※182。更に5月には、ベルギーとフランスの捜査当局がオランダの極右派欧州議会議員Marcel de Graaff氏の議会補佐官の自宅や事務所を捜索し、ロシアから欧州議会議員への賄賂工作に関与した容疑で捜査を進めた。当局はこの補佐官が「Voice of Europe」を通じてロシア資金を各国の極右議員に仲介し、見返りに親露的発言を欧州議会で行わせていた可能性をつかんだとしている。同補佐官はフランスの「国民連合（RN：Rassemblement National）」（旧、国民戦線）や「ドイツのための選択肢（AfD：Alternative für Deutschland）」の欧州議員にも協力していた過去が判明している。加えて同5月、ドイツ連邦議会議員Peter Bystron氏（AfD所属）がロシア人オリガルヒから2万ユーロの資金提供を受けていた

音声記録が暴露され、議員特権剥奪に至る事案も発生している。このように、ロシア政府は長年かけて欧州各国の極右勢力との間に資金・情報の影響ネットワークを構築しており、その一端が2024年に相次ぎ表面化した形である。欧州議会内では国民連合やAfD、オーストリア自由党（FPÖ：Freiheitliche Partei Österreichs）等親ロシア色の強い右派会派が勢力を増しており、ハンガリーのViktor Orbán首相が組織する欧州右派連合「欧州のための愛国者達（Patriots for Europe）」の旗のもとで第3会派を形成するに至っている。

また中国も、欧州でロシアに倣った影響工作を行っている実態が明らかとなっている。2023年末、ベルギー極右政治家が2019～2022年に中国情報機関と通じて「米欧分断」を図り、香港や新疆ウイグル問題で中国の主張を広めていたとの報道がなされた※182。更に2024年4月にはドイツ当局が欧州議会議員Maximilian Krah氏（AfD所属）の議員補佐官を逮捕した※184。この補佐官は中国の対外工作部門に欧州議会内部情報を繰り返し流し、在独中国人活動家の監視にも協力したスパイ容疑を掛けられている。同氏自身も2024年欧州議会選挙におけるAfDのトップ候補であり、ロシア及び中国から資金提供を受けていた疑惑でドイツ検察当局の捜査下に置かれた。更に欧州議会選挙の直前、オランダでは新党「オランダ計画党（Nederland met een plan）」が中国の統一戦線工作部に関連する団体から寄付を受けていたと報じられた。この新党はEUと中国の関係強化を掲げており、親中的な政策主張の背後に中国当局の影響が指摘されている。

欧州委員会や各国政府もロシア・中国発の情報操作に対抗すべく対策を講じている。EUは2024年2月に発効した「デジタルサービス法（DSA：Digital Services Act）」に基づき、大規模オンラインプラットフォーム等に対し組織的な偽情報操作の迅速な発見・除去を義務付けた。しかし規制の実効性確保には時間を要し、直前の欧州議会選挙期間中もロシア寄りの広告がFacebook上で拡散されたり、TikTokが虚偽広告の承認を許していたりしたとの指摘があった※182。ロシアの偽情報キャンペーングループ「Doppelgänger（ドッペルゲンガー）」により、2023年8月から2024年3月までの間に、3,800万人のフランスとドイツのFacebookユーザーが、影響工作の標的にされた。5月だけでも、275件の反EU・反ウクライナの政治広告が、フランス、ドイツ、ポーランド、イタリアの300万人以上のFacebookユーザーに届いた。また同月、ロシア国営メディアRT（旧Russia Today）のコンテンツをミラーサイトや地方ニュースに偽装して再配信する「マトリョーシカ型」情報操作も欧州各地で確認された。EUはRTやSputnikの公式配信を禁じているが、ロシア側はこのような巧妙な迂回戦術で制裁網を潜り抜け、欧州市民への影響力を維持しようとしている。EEASのEast StratComタスクフォース等は定期的に偽情報のファクトチェックを公開しているものの、プラットフォームの対応には差異がある。例えば、Xでは国家支援型のメディアや政治広告に対するラベル表示を削除してしまっていたり、TikTokでは偽情報を含む広告掲載を承認してしまっていたりする等※182、ある程度ボットネットワークに対処をしていたとしても、RT系記事拡散が可能なプラットフォームとして、EUの規制対応にも関わらず依然として利用されてしまった事例も出ている※185。

（b）フランス総選挙

2024年6月、フランスではEmmanuel Macron大統領が下院を解散し、6月下旬から7月にかけて国民議会下院選挙が実施された。フランスでは極右政党の国民連合がロシアと歴史的に近く、2014年にロシア系銀行から融資を受けた事実も知られる。前述の欧州議会議員補佐官による賄賂工作では、国民連合所属の欧州議員が標的となっており※182、2024年選挙戦でも、ロシアの情報工作ネットワークが国民連合を後押しして、対抗する人民戦線を分断するような偽情報を利用したキャンペーンを行っていたとされる※186。フランス政府は2021年に偽情報対策専門機関VIGINUMを設立しており、2024年2月にはロシア発の工作ネットワークである「Portal Kombat」を摘発・公表した※187。この工作ネットワークはクリミア拠点のIT企業TigerWebが運営に関与し、オリジナルコンテンツを一切作成せず、ロシアまたは親露的なアクターが運営するSNSアカウントやロシアのニュースメディアをリポストして拡散する、ロシア情報工作に典型的な手口であると判断された※188。更に、別のロシアの影響工作ネットワークである「CopyCop」も、特にウクライナ支援に関してMacron政権を誹謗中傷するために、選挙前からその活動を活発化していたことが明らかになっている※189。CopyCopは生成AIも活用した偽サイトとそのネットワークによる情報拡散を特徴とし、一例では、CopyCopのなりすましによる共和国連合（Ensemble pour la République）関連サイトに見せかけた偽サイトで、「現職大統領陣営に投票する代わりに100ユーロの『マクロン・ボーナス』をフランス国民に約束した」といった偽情報を拡散していた。選挙後には、フラ

ンス政府は同じく生成 AI を活用するロシアのネットワークである Storm-1516 の影響工作活動についても明らかにして非難を表明した※190。VIGINUM は 2021 年からロシアを中心とした偽情報の脅威を国民に周知してきており、今回の総選挙の前にも、前述のような報告書発出等により警鐘を鳴らしてきた。その甲斐もあってか、フランス総選挙では極右国民連合が一定議席を維持したものの左派・中道・極右の 3 勢力が拮抗し、ロシアが望むような急進的変化は生じなかった。

(c)オーストリア総選挙

オーストリアでは 2024 年 9 月の国民議会総選挙で、親ロシア志向の強い極右政党 FPÖ が躍進し第 1 党となった。この背景には、ロシアの長年にわたる浸透工作が関係している。2019 年に明るみに出た「イビザゲート事件」※191（FPÖ 党首がロシア富豪の姪と称する女性に便宜供与を約束した録画が公開された事件）以来、ロシアと FPÖ の癒着が長年懸念されていた。選挙後の 2025 年 3 月、オーストリア国内情報局はあるブルガリア人女性スパイの捜査から、ロシアがドイツ語圏向けに大規模な偽情報キャンペーンを展開していた事実を突き止めた※192。この女性は 2024 年末に逮捕されモスクワのために働いていたことを自白したが、最終的に釈放されている。押収された電子機器の分析で、ロシアがウクライナ侵攻後、オーストリアやドイツ等ドイツ語圏諸国を標的にウクライナに関する偽情報を流布し、国内の極右・反政府運動を扇動する工作を行っていたことが明らかになった。キャンペーンの一環として、偽情報の拡散と並行してウクライナ難民を中傷するステッカーの配布やグラフィティ活動を行い、極右のシンボルや民族主義的なメッセージを宣伝していた。こうした事例から、オーストリアがロシア情報工作の欧州拠点の一つとなっていることがうかがえる※193。2024 年選挙で FPÖ が政権入りしたことで、オーストリアの対露姿勢は軟化に向かったが、これにより欧米の対露強硬路線に一定の楔が打たれ、ロシアの欧州分断戦略はオーストリアでは一定の成果を上げたと見ることもできる。

(d)ルーマニア大統領選挙

ルーマニアでは 2024 年 11 月の大統領選挙が、ロシアの情報工作により深刻な混乱に陥った。2024 年の大統領選挙では第 1 回投票で極右ポピュリスト候補 Calin Georgescu 氏が予想外の首位となったが、この結果を受け情報機関当局が「外国勢力の介入による不正の可能性が高い」と指摘する機密解除文書が Klaus Werner Iohannis 大統領（当時）から発表された※194。最終的に憲法裁判所は、11 月に実施された大統領選挙を無効とする判断を示した※195。調査の結果、SNS アルゴリズムの特定候補者への悪用、不透明な生成 AI の利用や不当な資金源による選挙活動等が確認され※196、更に情報機関の調査ではその背後にロシアの存在があり、TikTok 等、SNS 上で Georgescu 候補を支援する大量の偽情報キャンペーンを展開していたことが明らかにされた※197。具体的には、同候補の露出が上がると他の候補者の露出は下がるようなアルゴリズムの適用や、ロシアに紐付くボットネットの活用、TikToker の買収等があったと考えられている※198。この事態を受け、EU 欧州委員会は TikTok Inc. に対し DSA 違反の正式調査を開始し、プラットフォームとしての対応不備を追及した※199。翌 2025 年 5 月に再選挙が実施されることとなった。ロシア政府はこれらの介入疑惑を公式には否定し、「Georgescu 氏は不当に選挙から排除された」と主張し、選挙の正当性に疑問を投げかける発信を行った。また、国営メディアや親露派アカウントを通じてその主張を拡散する宣伝を展開した※200-1。しかしその否定とは裏腹に、TikTok 以外にも、ルーマニア語の Telegram チャンネルのメッセージの 4 分の 1 は、ロシア国営メディアやその他の親露の情報源からの情報を宣伝するものであったことも確認されている※200-2。ルーマニアの事例は、ロシアが東欧でも巧妙に民主主義的プロセス攪乱を図った典型といえる。SNS 利用率の高い若年層を狙い TikTok を主戦場とした点では新しい世代に向けたハイブリッド戦の戦術であり、欧州各国が警戒を強めている。

(e)英国サウスポート暴動

2024 年 7 月 29 日、英国のサウスポートで 3 人の少女が刺殺され、同時に 8 人の子供と 2 人の大人も負傷する事件が起こった。警察は、17 歳のルワンダ系英国人（ルワンダ出身の両親のもとに英国国内で生まれており、移民ではない）の犯行として彼を逮捕した。ところが同日、犯人は 17 歳のイスラム系の亡命希望者である Ali al-Shakati なる人物だとする偽・誤情報がインターネット上に出回った。この「ニュース」を最初に投稿したのは、オンライン・インフルエンサーで陰謀論者の Bernadette Spofforth とされている※201。その数分後、米国の通信社を名乗る Web サイト「Channel3 Now」がこの偽情報をニュース記事のような形で拡散させた。X では、「Channel3 Now」の当該投稿は 2,700 万ビューを超え、

極右系インフルエンサーを中心に移民犯人説が拡散されていった。このような情報拡散で移民への憎悪が高まり、翌7月30日、サウスポートでの追悼集会に乗じて、極右系グループを中心とした暴動が発生した。

英国紙 The Telegraph によると※202、YouTube チャンネル「Channel3 Now」は2012年に開設され、当初はロシアの都市イジェフスクでのカーレースに関するロシア語動画を公開していた。その後しばらくはアカウントが利用されていなかったが、2019年、パキスタンに関する英語の動画を公開し、2023年には「Channel3 Now」の Web サイトがリトアニアのドメインで作成され、同 YouTube チャンネルにリンクされた。英国調査報道局は、「Channel3 Now」が前述のような経緯でロシアのチャンネルとして作られたことを確認し※203、Steven McPartland 元安全保障相も、この暴動を引き起こした SNS 上の偽情報拡散キャンペーンの背後にロシアがいる可能性を示唆している※204。一方で捜査当局は、抗議行動を引き起こしたディスインフォメーションへのロシアの関与を証明するには、まだ十分な証拠がないとしている※201。

この暴動について、「Channel3 Now」のルートにおけるロシア関与の証拠は不十分ではあるものの、移民憎悪の扇動や移民に関係する陰謀論の利用という点でこれまでロシアが英国に行ってきた影響力工作のアプローチやそのナラティブとの類似がある点には注意が必要である。

これまで、ロシアの影響力工作は常に北大西洋条約機構（NATO：North Atlantic Treaty Organization）諸国の内部対立を悪化させようとしてきた。だが英国ではそれらに加えて、ロシアは英国人と移民の間に存在する対立を利用し、民族的・宗教的憎悪を煽って英国社会を弱体化させようとしている。一部の英国人が、移民に対する政府のリベラルすぎる公共政策に不満を抱いているという指摘もあり※201、そのためロシアは対英工作において「なぜ移民が必要なのか？　彼らのホテルや避難所、モスクを破壊する必要がある」と繰り返し扇動している。

RT の宣伝担当者である Vladimir Kornilov 氏は、自身の Telegram チャンネルを使って、この暴動に際し「今こそ英国の国内問題を解決すべきであり、世界のあらゆる紛争に干渉すべき時ではない」という主張を広めていた※201。

更には、国際メディアプラットフォームである openDemocracy によれば※205、こうした移民憎悪の言説を支えるものとして、「グレート・リプレイスメント（大置換）陰謀論」のナラティブが利用されていると考えられるという。この理論では、北半球の白人が南半球からの移民に計画的に「取って代わられ」ており、フェミニスト達が中絶や避妊によって出生率を抑制していると考えられている。このような陰謀論をロシアが利用して相手国の分断を扇動し、社会を不安定化させようとしていることは既に先行研究で指摘されており※206、この暴動でもそのナラティブの背景にあるロシアの影響は看過できない。

なお暴動後、英国政府は「オンライン安全法（Online Safety Act）」の早期施行を表明し※207、SNS 企業に対し有害コンテンツ拡散防止の法的責務を課す姿勢を示した※208。

(3) 太平洋地域での偽情報流布の影響

インド太平洋地域でも、中国とロシアによる偽情報・影響工作が活発化している。とりわけ台湾や南太平洋の小国の政治に干渉する動き、及び地域の安全保障イベントに絡めた情報戦が顕在化している。各国・地域の政府は警戒を強めているものの、偽情報キャンペーンの巧妙化と拡散力の強大化が課題となっている。

(a) 台湾

台湾では2024年1月の総統選挙で与党候補の頼清徳氏が当選（5月に就任）し、これに対し中国が一層の圧力を強めた。中国は軍事演習や外交的威圧に加え、オンライン上での情報戦・認知戦をも仕掛けている。台湾国家安全局の報告によれば、2024年に中国が台湾向けに流布した「物議を醸す情報」は前年比60%増の約216万件に上った※209。その主要テーマは台湾、軍、そして頼清徳総統に対する米国の支援に対する懐疑的な見方を広めることで、「政府に対する国民の信頼を損ない、社会的分裂を高めようとした」と報告書は述べている。拡散の主要なプラットフォームは Facebook であったが、Facebook での件数が前年比40%増なのに対し、X は244%増、PTT や Dcard といった台湾の掲示板フォーラムは644%増と、標的とする SNS の件数増加に偏りが見られる。また、調査対象のプラットフォーム全体で2024年には合計2万8,216件の不正アカウントが確認され、2023年から1万1,661件増加した。主要なプラットフォームはやはり Facebook であり全体の77.85%（2万1,967アカウント）を占めていたが、一方で TikTok の不正アカウント数は1,614%増の4,371に達した。これは、若者世代を対象とした SNS をターゲットとしたものと考えられる。拡散されている偽情報の主要なナラティブは、

「毎年恒例の漢江訓練を含む軍事演習は台湾の指導部が脱走するためのリハーサルである」※210といった政府要人が国を見捨てる類型のものや、「米国はウクライナと同様、台湾にも武器を売って利益を得ようとしている」※211といった「疑米論（米国懐疑論）」が主流であり、毎年の演習や関連する外交イベントの度に繰り返し拡散されている。しかし、2024年7月ごろから、台湾の親日性を攻撃し日台離反を扇動する偽情報の拡散も増えてきている※212。具体的には、「日本人の子孫が台湾に住み、台湾独立を主張している」「台湾の地方政府が、第二次世界大戦中に日本軍によって強制連行された台湾人慰安婦像を撤去し、台湾政府は、台湾人慰安婦は強制連行されたのではなく、自発的に日本軍のために働いたと主張している」「頼清徳・台湾総統が金門島を訪問した際、日本式の旭日旗の前で演説を行った」といった内容である。頼総統の演説については、バックに掲示されていた国防省の紋章が旭日旗に似たデザインだったにすぎないが、演説画像に誤ったキャプションが付されて拡散された※213。これらは微博（Weibo）等中国のSNSにも輸入され、むしろ台湾国内よりも中国で反日感情、反台感情が高まった。台湾有事の懸念が高まる中では、対台湾だけでなく、このように日本を含めた有事に関わる可能性の高い国が関係する偽情報の流布も憂慮されるところである。日本関連の偽情報については、海外の事例であっても政府が鋭敏に把握して反論する体制が今後は求められるだろう。

(b)ツバル総選挙

台湾と公式な外交関係を持つ国の一つであるツバルでは、総選挙前後に中国が偽情報を用いた情報戦を展開し、台湾との外交関係を揺るがそうとする動きが確認されたと報じられた※214。選挙の前日、ある地元メディアは、中国の国営メディアである中国国際電視台（CGTN: China Global Television Network）のドメインcgtn.comを使用したアカウントから電子メールを受け取った。そのメールは、「ツバル放送局の幹部がツバルの選挙と台湾との関係断絶の可能性に焦点を当てた800字のオピニオン記事の執筆依頼に協力を求めるもの」で、記事執筆と執筆依頼の仲介協力に対する謝礼として合計450米ドルを支払うとの記載もあった。このほかにも、選挙直前には「選挙後にツバルが台湾との関係を断つ」との噂が流され、中国が背後で流布した可能性が指摘された。また、直前に隣国ナウルが台湾と断交し中国と国交を樹立したことから、ツバルも追随するとの噂も広まった※215。選挙後に誕生したFeleti Teo新首相は台湾との「揺るぎない」関係継続を表明したが、その後も中国の影響工作は続いた。2024年末には、中国のCGTNがツバル住民に「台湾は中国の一部」と語らせたり、「中国と国交を樹立すればインフラ整備や気候変動対策で利益が得られる」と示唆する複数の映像を放映し、台湾側はこれを両国関係の弱体化を狙った偽情報キャンペーンだとして強く非難し、台湾のツバル駐在大使館も声明で、言論の自由を尊重しつつも中国による世論操作や認知戦を容認しない姿勢を明確にしている※216。

このように、中国は自国の利益のために太平洋島しょ国まで、情報戦・認知戦の戦場を拡大していることがうかがえる。

(c)ソロモン諸島国政選挙における中露の連携

2024年のソロモン諸島総選挙では、ロシアと中国が偽情報を用いた選挙干渉工作を共謀して行ったとみられると報告されている※217。両国の国営メディアは、米国が同国への援助や現地ネットワークを利用して選挙に介入し、都合の悪い結果の場合には暴動を扇動して政権転覆を企図しているとする根拠なき言説を発信したと、オーストラリア戦略政策研究所は指摘している。

具体的には、選挙の約2週間前に匿名寄稿の米国陰謀告発記事がKGB主導で創刊されたとする雑誌CovertAction Magazineに掲載され、程なくロシアのSputnikも匿名情報源に基づき米国による選挙クーデター計画説を報じた※218。中国共産党系紙「環球時報」もこれを引用して米国の介入を糾弾した※219。更に「米国が必要なら暴力的手段による民主的移行を図る」と記された偽造書簡が国際選挙制度財団（IFES：The International Foundation for Electoral Systems）関係者を装って出回り、その内容はSputnik報道で情報源とされる匿名の証言者の言葉遣いと酷似していた。こうした偽情報は一般有権者にはほとんど浸透せず、現地世論を大きく動かすには至らなかったと分析されている。

しかし当時のManasseh Damukana Sogavare首相の与党O.U.R.は問題のSputnik記事をFacebookで共有し、首相自身も外国勢力が選挙に介入し暴動を企図していると公に主張した。この事例は、中国とロシアの制限なき戦略的協力関係が情報戦にも及ぶことを示している。両国の宣伝当局は、報道分野でメディア協力協定を結ぶ等の協力関係構築を進めており※220、欧州や南米でも中露が偽情報の増幅のために連携する事例が増えている※221。国家としては関係を深めつつも情報空

間での連携は機会主義的とみられるという指摘もあり、本件はそうした連携の機会が太平洋地域でも現れた事例であるといえる。今回ロシアが長年唱えてきた「米国国際開発局（USAID：United States Agency for International Development）は政権転覆の道具」とのナラティブが初めて中国の対米プロパガンダに援用された点が特徴的である。このように安全保障上の懸念国同士が連携すると、偽情報の拡散効果は倍増してしまうため、更なる警戒が必要である。

(4) ハイブリッド型サイバー攻撃の事例

2024 年度には、偽情報拡散による「情報操作型サイバー攻撃」と「機能妨害型サイバー攻撃」を組み合わせた「ハイブリッド型サイバー攻撃」が顕在化した。特にウクライナ支援に絡む国際会議や西側諸国が関与する首脳会議の周辺で、安全保障上の懸念国が情報操作とDDoS 攻撃等のサイバー攻撃を同時に仕掛ける事案が発生している。以下、日本が関連する主要な二つの事例を取り上げる。

(a) ウクライナ復興支援会議

日本政府は 2024 年 2 月 19 日、東京で「日・ウクライナ経済復興推進会議」を開催し、官民合わせて約 350 名が参加してウクライナ復興支援策を協議した※222。この重要な国際会議を標的に、親ロシア勢力がサイバー攻撃キャンペーンを実行している。ちょうど会議当日の 2024 年 2 月 19 日、ロシア寄りのハクティビスト集団「NoName057(16)」が「日本を標的にする」と犯行を予告する声明を Telegram 上に発出した※223。声明では、日本が「ロシアと戦闘中のウクライナを支援していること」が標的化の理由に挙げられており、日本のウクライナ支援策への報復としてサイバー攻撃を開始する旨が示された。実際、NoName057(16) は他の親露グループ（Cyber Army of Russia や UserSec 等）とも連携し、日本国内の政府・民間の Web サイトに対する大規模 DDoS 攻撃を仕掛けた。彼らが公開した標的リストには、日本の自動車関連団体や大手エネルギー企業、国際経済団体等のサイトが含まれており、これらに対しアクセス不能にする攻撃を行ったと主張している。更に NoName057(16) 自身は日本の国会の公式サイトや特定政党のサイト、日本税関当局、通信大手、証券業協会サイト等をダウンさせたと主張した。これら一連の攻撃は典型的なハクティビスト型 DDoS 攻撃で人的被害はないが、日本社会に不安を与える心理戦の側面を持つ。併せて、X では岸田首相（当時）が米国政府高官に睨みつけられているように見えるディープフェイク画像※224 や、「ウクライナ復興支援の予算が 50 兆円くらいになる」※225 といったような偽情報が拡散された。こうした影響工作に対しては、外国からの関与があったと外交筋が明らかにしている※226。これはロシアがウクライナ支援国に対し報復的ハイブリッド戦を仕掛けている一例であり、日本のみならず各国で類似の手口が警戒されている。偽情報と「機能妨害型（もしくは機能破壊型）サイバー攻撃」の組み合わせは相乗効果で標的国の混乱を狙うものであり、同様の攻撃に警戒が必要である。

(b) NATO サミットにおけるディスインフォメーションと DDoS 攻撃

ロシアはウクライナ支援の要となる NATO サミットに対しても、サイバー攻撃と偽情報を駆使した情報戦・認知戦を仕掛けている。2024 年 7 月にワシントン D.C. で開催された NATO サミットでは開幕前後の期間を通じて、ロシア系ハッカー集団「People's Cyber Army」や NoName057(16) が、NATO の危機管理・災害対応センター、連合軍特殊作戦部隊司令部、軍需品安全情報分析センターの Web サイトや各加盟国政府及び重要インフラ等の Web サイトに DDoS 攻撃を仕掛けている※227。更に同サミットに合わせ、少なくとも二つのロシア系グループが偽情報キャンペーンを展開していたことが Graphika Technologies, Inc. の調査で明らかになっている※228。一つは NATO 公式サイトを模倣したドメインから偽のプレスリリースを流布する手口で、例えば「NATO が防衛予算を倍増することを決定」「NATO がウクライナ軍をフランスに派遣し、同年夏に発生したフランス国内暴動に投入することを検討」といった偽情報が発信された。もう一つのグループはリトアニア政府から入手したと称する NATO サミットの内部警備資料を公開し、「NATO の機密が漏えいした」と騒ぎ立て拡散した。これらは手口や文体から過去に欧州で活動したロシアのキャンペーングループである、Doppelgänger や「Secondary Infektion」との類似性が指摘されている。また、単純な偽情報だけでなく、軍服を着たキャラクターや食糧とともに武器を配給しているようなイメージを用いて、NATO が戦争を起こしていると思わせるナラティブをベースに、反ユダヤ陰謀論や各国政府高官やセレブリティが小児性愛犯罪に加担しているといったディープステート（闇の政府）陰謀論と関連性が見られる偽情報、その他の様々な陰謀論と組み合わせて誇張したイメージをミュージックビデオにした動

画等も生成 AI によって量産され、拡散されている※229。

更には、日本でも次のような複合的な攻撃が見られた。7 月に「日本・NATO による合同軍事演習の機会を拡大し、パートナーシップを強化していく」との NATO 事務総長による宣言が発表された後、政党や鉄道会社等の Web サイトに対し NoName057(16) 等による DDoS 攻撃が発生した※230。そして同時期に SNS 上で、「NATO や NATO サミットは戦争拡大を目的としている」※231、「NATO はウクライナの加盟を保証していない」※232、「NATO によって日本が戦争に巻き込まれる」※233、「日本は軍需産業発展のため NATO に迎合している」※234、といったロシア寄りの戦略的なナラティブや偽情報が Sputnik を中心に親露派アカウントも利用しつつ拡散された。

こうした一連の行動は、異なるハッカー集団が協調してハードな「機能妨害型サイバー攻撃」とソフトな「情報操作型サイバー攻撃」を並行実行するという、「ハイブリッド型サイバー攻撃」として新たなサイバー攻撃動向として注視すべきである。政府機関や重要インフラ系の Web サイトが機能不全に陥ることで、ウクライナ支援継続に対する嫌がらせと政府への信頼を毀損することを企図し、更にこれらに対する社会の不安感を煽るような偽情報を拡散することで相乗効果を狙うものである。これはまさにハイブリッド戦の典型であり、今後も G7 や NATO といった外交イベントの度に類似の攻撃が試みられると予想される。各国政府と国際機関はこの教訓を踏まえ、技術的防御と戦略的なコミュニケーション、情報発信を組み合わせた包括的な対策を講じていく必要がある。

2.2.4 2024年度以前からの継続事象

本項においては、2024 年度以前から継続している、偽・誤情報の主要な拡散事象を解説する。

(1) 新型コロナウイルス

新型コロナウイルスをめぐっては、パンデミック初期からの偽・誤情報が 2024 年度も根強く残っている。代表的なものはワクチンに関する陰謀論であり、「ワクチン接種を強制される」といった内容が国内では引き続き拡散されたが、実際には WHO（World Health Organization：世界保健機関）主導の「パンデミック条約」草案に強制接種の規定はなく、WHO も何度も否定している※235。また、日本で 2023 年秋に導入された新型コロナウイルスの「レプリコンワクチン」についても、「従来のファイザー製より死亡率が 75 倍」とする誤った情報が 2024 年度に流布された※236。厚生労働省の資料を誤読したもので、ワクチン接種との因果関係が証明されていない数字を用いた主張であり、実際には死亡報告例も取り下げられている。このほか「コロナワクチンで 50 万人が死亡」「日本で人体実験が行われている」等極端な反ワクチン説も散見され、その多くは科学的根拠に欠けることが専門家の検証で明らかになっている※237。こうした新型コロナウイルス関連の偽情報は、度重なる否定やファクトチェックにもかかわらず一部で信じ続けられている。

更に 2024 年 11 月には、日本政府が mRNA ワクチンを「史上最も危険な薬」と分類したと主張するニュースが中国の SNS で拡散されたという事例も発生した※238。このニュース自体は、親露派の偽情報サイトと判明している The People's Voice というサイトから発信されたもので、そのスクリーンショットが Weibo で拡散され X でも広まったものである。2021 年 7 月に中露間でメディア協力に関する二国間協定が締結されて以来、新型コロナウイルスの「バイオラボ陰謀論」等を始めとして両国の偽情報拡散、ナラティブ形成の連携が強まった※220 ことを鑑みると、こうしたワクチンに関する偽情報が日本と関連付けて拡散された事例は中露連携による偽情報拡散の射程が日本に至っているとして警戒すべき一例といえるだろう。

(2) ロシア・ウクライナ戦争

ロシアは 2022 年 2 月のウクライナ侵攻に際し、開戦当初から偽情報を戦略的に利用して情報戦・認知戦を交戦国であるウクライナだけでなく国際世論に対しても展開し、2024 年度においてもその戦闘様相を継続させている。侵略を正当化するため「ウクライナの非ナチ化」や「NATO による脅威」といったナラティブが用いられ※239、ウクライナ政府をネオナチ国家、NATO を陰で操る侵略者、と虚偽の位置付けで描いている。そして国際世論を分断する戦略を有しているため、対ウクライナ支援や NATO 拡大への反対論を煽る影響工作も続いている。

2024 年には米国大統領選挙や NATO サミットにおけるロシア勢力による偽情報の拡散が見られたが、国際または内政上重要なイベント時以外にも多くの偽情報の発信が続いた。一例としては、米国国務省（DOS：U.S. Department of State）報道官がロシア内地への無差別攻撃を容認していることを示唆するかのようなディープフェイク動画がロシアの Telegram 上に出回り、ロシアの国営メディアや政府関係者によって拡散された※240。こうした事例に見られるように、ボットアカウント生成や AI 合

成映像等のデジタル手法が駆使され、国家ぐるみの巨大な宣伝網（SNS、RT や Sputnik 等の国営メディア、偽装ニュースサイト、サイバー工作部隊）によって偽情報が大量生産・拡散されている。サイバー空間での影響工作は、これまで Internet Research Agency という企業がロシア政府機関と協働していたが、これは Yevgeny Viktorovich Prigozhin 氏による創設であったため、同氏が率いていた軍事組織ワグネルの反乱により解体された。その後、現在の工作活動は、ソーシャル・デザイン・エージェンシー（SDA：The Social Design Agency）、インターネット開発研究所（The Institute for Internet Development）、そしてストラクチュラ（Structura）といった三つのロシア企業に帰属するとされており、これらの企業は、「雇われ影響工作企業（influence-for-hire firms）」と呼ばれている※241。

更に、ロシアの偽情報は中国とも連動しており、例えば「ウクライナに米国の生物兵器研究所がある」という根拠なき陰謀論は、後追いで中国政府高官や国営メディアによって増幅された※220。こうした偽情報キャンペーンは単発の誤報ではなく、自国に有利なナラティブを形成するように国家戦略の一環として継続的に構築・流通している点に特徴がある※242。また、紛争が長期化する中で、西側諸国がウクライナに供与した武器の行方をめぐる偽情報も続いている。ロシア発のプロパガンダでは「ウクライナが受け取った武器を第三国やテロ組織に横流ししている」といった主張が度々現れ、2023 年 10 月には「ウクライナ経由の武器がハマスに渡った」との情報が SNS で拡散された※243-1。しかしこれは BBC のニュース映像を装った偽動画によるもので、BBC や調査報道機関がすぐに否定し、実際にはウクライナ当局も供与兵器の管理徹底を強調している。このように、ロシア・ウクライナ戦争に関する偽情報は、ロシアの侵攻を正当化するナラティブをベースに、生物兵器開発の陰謀論※243-2 から新たな情勢を絡めた武器横流し説まで形を変え、ウクライナ支援国に対するハラスメント目的の偽情報も派生して、日本を含む国際社会で流布し続けている。

(3) イスラエル・ハマス紛争

2023 年 10 月に始まったイスラエルとハマスの武力紛争に関する偽情報は、その後も 2024 年度を通じて拡散し続けた。戦場がガザから情報空間へと移る中、戦略的レベルでは「誰が被害者で誰が加害者か」という認識をめぐる物語の争奪となり、偽情報が構築する戦略的ナラティブが重要な役割を果たした※244。両陣営とも自らの正当性を訴える被害者としてのナラティブを掲げ、大量の映像や情報を発信して国際世論の支持獲得を試みている。

この過程で SNS 上には偽・誤情報が氾濫し、過去の映像やゲーム画面があたかも現在の戦闘映像であるかのように使い回されたほか※245、AI で生成・加工された偽の画像・動画が本物として共有される例も相次いだ※246。

開戦時には大量の偽情報が飛び交うこととなり、イスラエルの SNS 分析企業である Cyabra によれば、攻撃開始後の 1 ヵ月で、少なくとも約 4 万件以上の Bot アカウント及び不正なアカウントを確認したという。また、Facebook、Instagram、TikTok、X でこの武力衝突について投稿したアカウントのおよそ 4 個に 1 個が偽のアカウントであることが攻撃後 1 日で判明している。更には、アル・アハリ病院での爆発から 24 時間以内に、X に本件を投稿したアカウントの 3 個に 1 個以上が偽のアカウントであった。2024 年の事例としては、米国で起きた橋梁崩落事故を「イスラエルの仕業」とする荒唐無稽な陰謀論が X 上で拡散される等、紛争と無関係な出来事まで情報戦に利用されている混乱ぶりである。国際的にも、ロシアや中国、イランの国家メディアや偽アカウントがオンライン上でイスラエル・米国を貶めハマスを擁護する情報戦を展開している※247。これらの偽情報も単発の情報発信による社会の混乱（情報騒乱）にとどまらず、各主体の戦略的ナラティブに基づいた情報戦・認知戦の様相を呈している。

2.2.5 状況のまとめと今後の見通し

本項では、これまで解説してきた事例から現況をまとめた上で、今後の見通しを述べる。

(1) 状況のまとめ

本節で取り上げた情勢や事例が示すとおり、2024 年も各分野で偽・誤情報が猛威を振るった。そのうち、主要な傾向として以下の 2 点が挙げられる。第 1 に、生成 AI が偽・誤情報拡散を加速・巧妙化させている点である。2022 年ごろから ChatGPT や Midjourney、Stable Diffusion といった生成 AI がリリースラッシュとなり、おおよそ 1 ～ 2 年かけて一般への浸透が進んできた。また、X では、SNS に付随した AI サービスとして Grok が 2024 年に提供され始め、AI を使った SNS 投稿を格段に容易にした。ユーザーが容易に生成 AI でフェイク

画像・音声・動画を作成できるようになり、災害時に偽の被害写真が出回ったり政治家になりすました偽映像が登場したりする等、ディープフェイクによる新たな混乱が顕在化した。更に、「2.2.2 偽・誤情報の情勢」及び「2.2.3 2024 年度の注目事象」で触れたとおり、国家アクターが情報操作のために戦略的に生成 AI を利用する事態にも至っている。第 2 に、旧来のナラティブが延命し続けている点が挙げられる。新型コロナウイルスやウクライナ戦争に端を発する陰謀論や反権威的なナラティブは、一度否定されても形を変えて生き残り、別の文脈で再利用されている。例えば反ワクチン運動やディープステート陰謀論は、パンデミック後も国際機関の新条約や新たな有事を標的に置き換えながら、その主張が広がり続けている。こうした拡散状況は自然発生的な現象だけではなく、中国やロシアといった安全保障上の懸念国が戦略的なナラティブを拡散することで、自国に有利な認知形成を目指す認知戦の広がりが背景にあることは留意すべきである。総じて 2024 年は、AI と SNS が生み出す拡散構造の中で戦略的ナラティブが延命することで、偽・誤情報が根強く社会に拡散し続けた一年だったともいえよう。

(2) 今後の見通し

2025 年は 2024 年に続き、世界各国で重要な選挙や国際政治行事が実施される予定となっており、それらに絡んだ偽情報の更なる増加が懸念される。具体的には、ノルウェー議会選挙（2025 年 9 月実施予定）、チェコ議会選挙（2025 年 10 月実施予定）、アイルランド大統領選挙（2025 年 10 月実施予定）、シンガポール総選挙（2025 年 11 月実施予定）等、主要各国の国政選挙や、太平洋戦争終結 80 年に伴う各国記念日、G20 サミット等の民主主義プロセスや外交イベントへの介入を狙った情報工作が警戒される。ドイツでは、既に 2025 年 2 月の選挙直前にロシア系グループが関与したと見られる「極右政党への投票用紙が紛失・破棄された」という偽動画が出回り当局が注意喚起する事態が起きた※248。ロシアはこれまでも紛争や国際会議の度に新たな偽情報を投入してきており（「2.2.4 (2) ロシア・ウクライナ戦争」で紹介した武器横流し説等）、ウクライナ復興会議でも「ハイブリッド型サイバー攻撃」による攻勢が懸念される。また、新たな国際情勢の火種として、インド・パキスタン関係が不安定化しており、両国に関連した偽情報の流布が増えていくことが想定される。翻って日本では、国政選挙等で、候補者に関するデマや国外勢力による情報操作が懸念されるところであり、十分な警戒が必要だろう。国際的な場においても日本国内においても、外国からの情報操作と国内の適正な情報環境の維持の双方に注意を払い、適切な対策が求められる。

※1 Y. LeCun et al.：Backpropagation Applied to Handwritten Zip Code Recognition https://doi.org/10.1162/neco.1989.1.4.541〔2025/6/18 確認〕
ニューラルネットを用いない方法での郵便番号の機械認識はもっと早い段階で実用化されている。例えば日本では 1967 年に世界初の手書き文字読み取り試作機（TR-2 型）が完成している。
郵政博物館：最初の郵便番号自動読取区分機 https://www.postalmuseum.jp/column/collection/post_27.html〔2025/6/18 確認〕
※2 電子情報通信学会：「知識ベース」S3 群―3 編―3 章 機械学習 https://ieice-hbkb.org/files/ad_base/view_pdf.html?p=/files/S3/S3gun_03hen_03.pdf#page=1〔2025/6/18 確認〕
※3 Paul Graham：「スパムへの対策 --- A Plan for Spam」（和訳）https://web.archive.org/web/20110724020751/http://practical-scheme.net/trans/spam-j.html〔2025/6/18 確認〕
※4 岡本 一志・藤井 流華：「協調フィルタリング入門」知能と情報（日本知能情報ファジィ学会誌）Vol.31, No.1, pp.5－9（2019） https://www.jstage.jst.go.jp/article/jsoft/31/1/31_5/_pdf〔2025/6/18 確認〕
※5 人工知能学会監修：「深層学習 — Deep Learning」 https://jsai-deeplearning.github.io/support/dlspecial.html〔2025/6/18 確認〕
※6 Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun：Deep Residual Learning for Image Recognition https://arxiv.org/abs/1512.03385〔2025/6/18 確認〕
WIRED：黒 37 手と白 78 手：AlphaGo とイ・セドルが再定義した「未来」 https://wired.jp/special/2016/alphago-vs-sedol/〔2025/6/18 確認〕
※7 総務省、経済産業省：AI 事業者ガイドライン（第 1.1 版） https://www.meti.go.jp/shingikai/mono_info_service/ai_shakai_jisso/pdf/20250328_1.pdf〔2025/6/18 確認〕
※8 CNBC：On ChatGPT's one-year anniversary, it has more than 1.7 billion users—here's what it may do next(2023/11/30) https://www.cnbc.com/2023/11/30/chatgpts-one-year-anniversary-how-the-viral-ai-chatbot-has-changed.html〔2025/6/18 確認〕
※9 European Commission：AI Act https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai〔2025/6/18 確認〕
※10 GOV.UK：Frontier AI: capabilities and risks – discussion paper https://www.gov.uk/government/publications/frontier-ai-capabilities-and-risks-discussion-paper/frontier-ai-capabilities-and-risks-discussion-paper〔2025/6/18 確認〕
※11 TIME：The Billion-Dollar Price Tag of Building AI https://time.com/6984292/cost-artificial-intelligence-compute-epoch-report/〔2025/6/18 確認〕
※12 TIME：The AI Arms Race Is Changing Everything https://time.com/6255952/ai-impact-chatgpt-microsoft-google/〔2025/6/18 確認〕
※13 Jared Kaplan et al.：Scaling Laws for Neural Language Models https://doi.org/10.48550/arXiv.2001.08361〔2025/6/18 確認〕
Tom Henighan et al.：Scaling Laws for Autoregressive Generative Modeling https://doi.org/10.48550/arXiv.2010.14701〔2025/6/18 確認〕
※14 GOV.UK：International Scientific Report on the Safety of Advanced AI: interim report https://www.gov.uk/government/publications/international-scientific-report-on-the-safety-of-advanced-ai〔2025/6/18 確認〕
※15 GOV.UK：International AI Safety Report 2025 https://www.gov.uk/government/publications/international-ai-safety-report-2025〔2025/6/18 確認〕

※ 16 GAO：GAO-20-379SP Science & Tech Spotlight: Deepfakes https://www.gao.gov/products/gao-20-379sp〔2025/6/18 確認〕
※ 17 FBI：Alert Number I-060523-PSA Malicious Actors Manipulating Photos and Videos to Create Explicit Content and Sextortion Schemes https://www.ic3.gov/PSA/2023/PSA230605〔2025/6/18 確認〕
※ 18 Ofcom：A deep dive into deepfakes that demean, defraud and disinform https://www.ofcom.org.uk/online-safety/illegal-and-harmful-content/deepfakes-demean-defraud-disinform/?language=en〔2025/6/18 確認〕
※ 19 Tvesha Sippy et al.：Behind the Deepfake: 8% Create; 90% Concerned. Surveying public exposure to and perceptions of deepfakes in the UK https://doi.org/10.48550/arXiv.2407.05529〔2025/6/18 確認〕
※ 20 NHK：卒業アルバムの同級生を裸に 子どもも加害者?画像加工の実態 https://www3.nhk.or.jp/news/html/20241218/k10014666221000.html〔2025/6/18 確認〕
※ 21 Takamichi Saito：影響力工作についての簡単な整理を通して、「Industrialized Disinformation ー 2020 Global Inventory of Organized Social Media Manipulation」を読み解く https://saitolab-org.medium.com/影響力工作についての簡単な整理を通して-industrialized-disinformation- ー -2020-global-inventory-of-organized-social-319d30d5f377〔2025/6/18 確認〕
※ 22 FNN プライムオンライン：ディープフェイク悪用“ゼレンスキー大統領”が国民に降伏呼びかけるニセ動画 見分ける自信ありますか? https://www.fnn.jp/articles/-/333829〔2025/6/18 確認〕
※ 23 NHK：オープン AI “ロシアなど拠点のグループ 生成 AI で世論操作” https://www3.nhk.or.jp/news/html/20240531/k10014466761000.html〔2025/3/11 確認〕
※ 24 U.S. Department of Justice：Justice Department Disrupts Covert Russian Government-Sponsored Foreign Malign Influence Operation Targeting Audiences in the United States and Elsewhere https://www.justice.gov/archives/opa/pr/justice-department-disrupts-covert-russian-government-sponsored-foreign-malign-influence〔2025/6/18 確認〕
※ 25 OpenAI 社：Disrupting deceptive uses of AI by covert influence operations https://web.archive.org/web/20240530173817/https://openai.com/index/disrupting-deceptive-uses-of-ai-by-covert-influence-operations/〔2025/6/18 確認〕
※ 26 Michael Kouremetis et al：What Lies Beneath the Surface? Evaluating LLMs for Offensive Cyber Capabilities through Prompting, Simulation & Emulation https://www.blackhat.com/us-24/briefings/schedule/#what-lies-beneath-the-surface-evaluating-llms-for-offensive-cyber-capabilities-through-prompting-simulation-38-emulation-40685〔2025/6/18 確認〕
※ 27 Microsoft Threat Intelligence：Staying ahead of threat actors in the age of AI https://www.microsoft.com/en-us/security/blog/2024/02/14/staying-ahead-of-threat-actors-in-the-age-of-ai/〔2025/6/18 確認〕
OpenAI 社：Disrupting malicious uses of AI by state-affiliated threat actors https://openai.com/index/disrupting-malicious-uses-of-ai-by-state-affiliated-threat-actors/〔2025/6/18 確認〕
※ 28 NCSC：The near-term impact of AI on the cyber threat https://www.ncsc.gov.uk/report/impact-of-ai-on-cyber-threat〔2025/6/18 確認〕
※ 29 読売新聞オンライン：中高生の生成AI悪用事件、「指示役」中3男子「本人確認が甘い楽天狙った」…カード不正利用も判明 https://www.yomiuri.co.jp/national/20250227-OYT1T50322/〔2025/6/18 確認〕
※ 30 The Verge：AI suggested 40,000 new possible chemical weapons in just six hours https://www.theverge.com/2022/3/17/22983197/ai-new-possible-chemical-weapons-generative-models-vx〔2025/6/18 確認〕
※ 31 Bloomberg：AI-Made Bioweapons Are Washington’s Latest Security Obsession https://www.bloomberg.com/news/features/2024-08-02/national-security-threat-from-ai-made-bioweapons-grips-us-government〔2025/6/18 確認〕
※ 32 Havard Sussex Program：Artificial Intelligence Technologies and Chemical and Biological Weapons: A Chronology of Events (2000-Present) http://hsp.sussex.ac.uk/new/_uploads/publications/AI_and_CBW_Chronology_January_2025.pdf〔2025/6/18 確認〕
※ 33 原文では a simulated version of the Uniform Bar Examination と表記。
※ 34 OpenAI 社：GPT-4 Technical Report https://doi.org/10.48550/arXiv.2303.08774〔2025/6/18 確認〕
※ 35 Tiffany H. Kung et al.：Performance of ChatGPT on USMLE: Potential for AI-assisted medical education using large language models https://doi.org/10.1371/journal.pdig.0000198〔2025/6/18 確認〕
※ 36 Eric Martínez：Re-evaluating GPT-4’s bar exam performance https://doi.org/10.1007/s10506-024-09396-9〔2025/6/18 確認〕
※ 37 Jesutofunmi A. Omiye et al：Large language models propagate race-based medicine https://doi.org/10.1038/s41746-023-00939-z〔2025/6/18 確認〕
※ 38 日本経済新聞：ChatGPT で資料作成、実在しない判例引用 米国の弁護士 https://www.nikkei.com/article/DGXZQOGN30E450Q3A530C2000000/〔2025/6/18 確認〕
※ 39 NHK：虐待疑われる子どもの保護判定 AI の導入見送りに こども家庭庁 https://www3.nhk.or.jp/news/html/20250303/k10014738301000.html〔2025/6/18 確認〕
※ 40 Reuters：焦点：アマゾンがAI採用打ち切り、「女性差別」の欠陥露呈で https://jp.reuters.com/article/amazon-jobs-ai-analysis-idJPKCN1ML0DN/〔2025/6/18 確認〕
※ 41 Joy Buolamwini & Timnit Gebru：Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification https://proceedings.mlr.press/v81/buolamwini18a.html〔2025/6/18 確認〕
※ 42 Julia Dressel & Hany Farid：The accuracy, fairness, and limits of predicting recidivism https://doi.org/10.1126/sciadv.aao5580〔2025/6/18 確認〕
※ 43 Marc Cheong et al：Investigating Gender and Racial Biases in DALL-E Mini Images https://doi.org/10.1145/3649883〔2025/6/18 確認〕
※ 44 UNESCO：CI/DIT/2024/GP/01 Challenging systematic prejudices: an investigation into bias against women and girls in large language models https://unesdoc.unesco.org/ark:/48223/pf0000388971〔2025/6/18 確認〕
※ 45 Chris Lu et al.：The AI Scientist: Towards Fully Automated Open-Ended Scientific Discovery https://doi.org/10.48550/arXiv.2408.06292〔2025/6/18 確認〕
※ 46 Sakana AI 株式会社：AI CUDA Engineer：エージェントによる CUDA カーネルの発見、最適化、生成 https://sakana.ai/ai-cuda-engineer-jp/〔2025/6/18 確認〕
※ 47 TechCrunch：Sakana walks back claims that its AI can dramatically speed up model training https://techcrunch.com/2025/02/21/sakana-walks-back-claims-that-its-ai-can-dramatically-speed-up-model-training/〔2025/6/18 確認〕
※ 48 ITmedia：OpenAI と Apollo Research、「o1」は自分の目的のために嘘をつくと報告 https://www.itmedia.co.jp/news/articles/2412/06/news169.html〔2025/6/18 確認〕
※ 49 Kevin Zheyuan Cui et al.：The Productivity Effects of Generative AI: Evidence from a Field Experiment with GitHub Copilot https://doi.org/10.21428/e4baedd9.3ad85f1c〔2025/6/18 確認〕
※ 50 Shakked Noy & Whitney Zhang：Experimental evidence on the productivity effects of generative artificial intelligence https://doi.org/10.1126/science.adh2586〔2025/6/18 確認〕
※ 51 bloomberry：The jobs being replaced by AI – an analysis of 5M freelancing jobs https://bloomberry.com/i-analyzed-5m-freelancing-jobs-to-see-what-jobs-are-being-replaced-by-ai/〔2025/6/18 確認〕
※ 52 The Econometric Society：Tasks, Automation, and the Rise in U.S. Wage Inequality https://www.econometricsociety.org/publications/econometrica/2022/09/01/Tasks-Automation-and-the-Rise-in-US-Wage-Inequality〔2025/6/18 確認〕
※ 53 Epoch AI：How Much Does It Cost to Train Frontier AI Models? https://epoch.ai/blog/how-much-does-it-cost-to-train-frontier-ai-models〔2025/6/18 確認〕
※ 54 Google 社：Nebraska, USA – Google Data Center Location https://datacenters.google/locations/nebraska/〔2025/6/18 確認〕
※ 55 Governor of Missouri：Governor Parson Announces Google's Selection of Kansas City for New Data Center https://web.archive.org/web/20240330150938/https://governor.mo.gov/press-releases/archive/governor-parson-announces-googles-selection-kansas-city-new-data-center〔2025/6/18 確認〕
※ 56 Post by Richland Parish Data Center https://www.facebook.com/RichlandParishDataCenter/posts/pfbid0u2WXqnJgziQZFEvZbz1qg1Pgbw3xLSkaFGJ26xgNqaLp4YojFWsJqHbvTy8fKJDn〔2025/3/11 確認〕
※ 57 Statista：Amazon and Microsoft Stay Ahead in Global Cloud

Market https://www.statista.com/chart/18819/worldwide-market-share-of-leading-cloud-infrastructure-service-providers/〔2025/6/18 確認〕
※ 58 JPNIC：SPOF とは https://www.nic.ad.jp/ja/basics/terms/spof.html〔2025/6/18 確認〕
※ 59 Jens Malmodin et al.：ICT Sector Electricity Consumption and Greenhouse Gas Emissions – 2020 Outcome https://dx.doi.org/10.2139/ssrn.4424264〔2025/6/18 確認〕
※ 60 IEA:Electricity 2024 https://www.iea.org/reports/electricity-2024〔2025/6/18 確認〕
※ 61 IEA：Word Energy Outlook 2024 https://www.iea.org/reports/world-energy-outlook-2024〔2025/6/18 確認〕
※ 62 Google 社：2024 Environmental Report https://sustainability.google/reports/google-2024-environmental-report/〔2025/6/18 確認〕
※ 63 NHK：グーグル 原発から電力調達へ AI 活用で電力需要高まる https://www3.nhk.or.jp/news/html/20241015/k10014610011000.html〔2025/6/18 確認〕
※ 64 Constellation：Constellation to Launch Crane Clean Energy Center, Restoring Jobs and Carbon-Free Power to The Grid https://www.constellationenergy.com/newsroom/2024/Constellation-to-Launch-Crane-Clean-Energy-Center-Restoring-Jobs-and-Carbon-Free-Power-to-The-Grid.html〔2025/6/18 確認〕
※ 65 NHK:エネルギー基本計画決定 “再エネ最大電源に 原子力も活用” https://www3.nhk.or.jp/news/html/20250218/k10014725571000.html〔2025/6/18 確認〕
※ 66 Nicholas Carlini et al.：Extracting Training Data from Large Language Models https://www.usenix.org/conference/usenixsecurity21/presentation/carlini-extracting〔2025/6/18 確認〕
※ 67 Office of the Privacy Commissioner of Canada：Joint statement on data scraping and the protection of privacy https://www.priv.gc.ca/en/opc-news/speeches-and-statements/2023/js-dc_20230824/〔2025/6/18 確認〕
※ 68 McKinsey&Company：Generative AI in healthcare: Adoption trends and what's next https://www.mckinsey.com/industries/healthcare/our-insights/generative-ai-in-healthcare-adoption-trends-and-whats-next〔2025/6/18 確認〕
※ 69 Global Market Insights：AI In Video Surveillance Market Size https://www.gminsights.com/industry-analysis/ai-in-video-surveillance-market〔2025/6/18 確認〕
※ 70 Federal Trade Commission：FTC Staff Report Finds Large Social Media and Video Streaming Companies Have Engaged in Vast Surveillance of Users with Lax Privacy Controls and Inadequate Safeguards for Kids and Teens https://www.ftc.gov/news-events/news/press-releases/2024/09/ftc-staff-report-finds-large-social-media-video-streaming-companies-have-engaged-vast-surveillance〔2025/6/18 確認〕
※ 71 Federal Trade Commission：FTC Says Ring Employees Illegally Surveilled Customers, Failed to Stop Hackers from Taking Control of Users' Cameras https://www.ftc.gov/news-events/news/press-releases/2023/05/ftc-says-ring-employees-illegally-surveilled-customers-failed-stop-hackers-taking-control-users〔2025/6/18 確認〕
※ 72 PC Watch：写真素材サイト大手が Stable Diffusion を提訴。「1,200 万枚以上の写真を無断で複製」 https://pc.watch.impress.co.jp/docs/news/1476475.html〔2025/6/18 確認〕
※ 73 BBC NEWS JAPAN：米紙ニューヨーク・タイムズがオープン AI とマイクロソフトを提訴 著作権侵害で https://www.bbc.com/japanese/67831445〔2025/6/18 確認〕
※ 74 Richard Fletcher：How many news websites block AI crawlers? https://reutersinstitute.politics.ox.ac.uk/how-many-news-websites-block-ai-crawlers〔2025/6/18 確認〕
※ 75 GOV.UK：AI Safety Institute: overview https://www.gov.uk/government/publications/ai-safety-institute-overview〔2025/6/18 確認〕
※ 76 日本 AISI：AI セーフティ年次レポート 2024 https://aisi.go.jp/assets/pdf/j-aisi_report_2024_ja.pdf〔2025/7/15 確認〕
※ 77 足立浩規他：［サーベイ論文］Adversarial Training http://mprg.jp/data/MPRG/F_group/F20220310_adachi.pdf〔2025/6/18 確認〕
※ 78 リスクマネジメントの考え方は EU AI Act の Article 9: Risk Management System で簡潔に整理されている。
EU Artificial Intelligence Act：Article 9: Risk Management System https://artificialintelligenceact.eu/article/9/〔2025/6/18 確認〕
※ 79 JIS Q 31000:2019 等を念頭に説明を翻案している。
※ 80 JIS Q 31010 が同規格の邦訳となっている。
※ 81 一般社団法人日本品質管理学会 AI 品質アジャイルガバナンス研究会著 編「AI リスクアセスメント ガイドブック」
※ 82 NIST：AI Risk Management Framework https://airc.nist.gov/airmf-resources/airmf/〔2025/6/18 確認〕
上記の Web サイトは関連資料へのリンクも集約したポータルサイトとなっている。AI RMF 1.0 版へのリンクは次のとおり。NIST：Artificial Intelligence Risk Management Framework (AI RMF 1.0) https://doi.org/10.6028/NIST.AI.100-1〔2025/6/18 確認〕
※ 83 NIST：NIST AI 600-1 Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile https://doi.org/10.6028/NIST.AI.600-1〔2025/6/18 確認〕
※ 84 経済産業省：AI 事業者ガイドライン https://www.meti.go.jp/shingikai/mono_info_service/ai_shakai_jisso/20240419_report.html〔2025/6/18 確認〕
※ 85 日本 AISI：AI 事業者ガイドラインと米国 NIST AI リスクマネジメントフレームワーク（RMF）のクロスウォーク2 https://aisi.go.jp/output/output_information/240918_1/〔2025/6/18 確認〕
※ 86 AI モデル：ディープニューラルネット（DNN）を前提にした場合、DNN は線形写像と非線形写像（活性化関数）の組を 1 層としてこれを多層化した構造となっており、各層を定義するパラメーター式と、層と層の間のつながりを記述したデータをまとめたものを一般に AI モデルと呼ぶ。なお、AI 事業者ガイドライン1. 1版では、AI システムを「活用の過程を通じて様々なレベルの自律性をもって動作し学習する機能を有するソフトウェアを要素として含むシステム」、AI モデルを「AI システムに含まれ、学習データを用いた機械学習によって得られるモデルで、入力データに応じた予測結果を生成する」ものと定義している。
※ 87 日本 AISI：AI セーフティに関する評価観点ガイド（第 1.01 版） https://aisi.go.jp/effort/effort_framework/guide_to_evaluation_perspective_on_ai_safety/〔2025/6/18 確認〕
※ 88 GAO：GAO-24-106946 Artificial Intelligence: Generative AI Technologies and Their Commercial Applications https://www.gao.gov/products/gao-24-106946〔2025/6/18 確認〕
※ 89 Yupeng Chang et al.：A Survey on Evaluation of Large Language Models https://doi.org/10.48550/arXiv.2307.03109〔2025/6/18 確認〕
※ 90 Richard Harang：Practical LLM Security: Takeaways From a Year in the Trenches https://www.blackhat.com/us-24/briefings/schedule/#practical-llm-security-takeaways-from-a-year-in-the-trenches-39468〔2025/6/18 確認〕
※ 91 「AI セーフティに関するレッドチーミング手法ガイド」ではレッドチーミングを「攻撃者がどのように AI システムを攻撃するかの観点で、AI セーフティへの対応体制及び対策の有効性を確認する評価手法」と定義している。
※ 92 https://aisi.go.jp/effort/effort_framework/guide_to_red_teaming_methodology_on_ai_safety/〔2025/6/18 確認〕
※ 93 NIST：NIST AI 100-2 e2023 Adversarial Machine Learning: A Taxonomy and Terminology of Attacks and Mitigations https://doi.org/10.6028/NIST.AI.100-2e2023〔2025/6/18 確認〕
AI プロダクト品質保証コンソーシアム：AI プロダクト品質保証ガイドライン (QA4AI Guidelines) https://www.qa4ai.jp/download〔2025/6/18 確認〕
国立研究開発法人産業技術総合研究所：機械学習品質マネジメントガイドライン 第 4 版 https://www.digiarc.aist.go.jp/publication/aiqm/guideline-rev4.html〔2025/6/18 確認〕
※ 94 日本 AISI：データ品質マネジメントガイドブック（ドラフト版） https://aisi.go.jp/effort/effort_information/250207_2/〔2025/6/18 確認〕
※ 95 Ada Lovelace Institute：Under the radar? https://www.adalovelaceinstitute.org/report/under-the-radar/〔2025/6/18 確認〕
※ 96 2025 年 2 月に AI Security Institute へと改称した。
※ 97 英国 AISI：Inspect https://inspect.ai-safety-institute.org.uk/〔2025/6/18 確認〕
※ 98 英国 AISI：Advanced AI evaluations at AISI: May update https://www.aisi.gov.uk/work/advanced-ai-evaluations-may-update〔2025/6/18 確認〕
英国 AISI：Early Insights from Developing Question-Answer Evaluations for Frontier AI https://www.aisi.gov.uk/work/early-insights-from-developing-question-answer-evaluations-for-frontier-ai〔2025/6/18 確認〕
英国 AISI：Early lessons from evaluating frontier AI systems https://www.aisi.gov.uk/work/early-lessons-from-evaluating-frontier-ai-systems〔2025/6/18 確認〕
英国 AISI：Pre-Deployment Evaluation of Anthropic's Upgraded Claude 3.5 Sonnet https://www.aisi.gov.uk/work/pre-deployment-evaluation-of-anthropics-upgraded-claude-3-5-sonnet〔2025/6/18 確認〕

英国 AISI：Pre-Deployment Evaluation of OpenAI's o1 Model https://www.aisi.gov.uk/work/pre-deployment-evaluation-of-openais-o1-model〔2025/6/18 確認〕
※ 99 UNESCO：Recommendation on the Ethics of Artificial Intelligence https://www.unesco.org/en/articles/recommendation-ethics-artificial-intelligence〔2025/6/18 確認〕
※ 100 内閣府 科学技術・イノベーション推進事務局：「米国の AI 権利章典（AI Bill of Rights）について https://www8.cao.go.jp/cstp/ai/ningen/r4_2kai/siryo3.pdf〔2025/6/18 確認〕
※ 101 Federal Register：Executive Order 14110 Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence https://www.federalregister.gov/documents/2023/11/01/2023-24283/safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence〔2025/6/18 確認〕
※ 102 国連：Governing AI for Humanity https://doi.org/10.18356/9789211067873〔2025/6/18 確認〕
※ 103 The White House：FACT SHEET: Vice President Harris Announces New U.S. Initiatives to Advance the Safe and Responsible Use of Artificial Intelligence https://bidenwhitehouse.archives.gov/briefing-room/statements-releases/2023/11/01/fact-sheet-vice-president-harris-announces-new-u-s-initiatives-to-advance-the-safe-and-responsible-use-of-artificial-intelligence/〔2025/6/18 確認〕
※ 104 経済産業省：AI セーフティ・インスティテュートを設立しました https://www.meti.go.jp/press/2023/02/20240214002/20240214002.html〔2025/6/18 確認〕
※ 105 Singapore AI Safety Institute:https://sgaisi.sg/〔2025/6/18 確認〕
※ 106 Canadian Artificial Intelligence Safety Institute：https://ised-isde.canada.ca/site/ised/en/canadian-artificial-intelligence-safety-institute〔2025/6/18 確認〕
※ 107 韓国 AISI：AI Safety Institute https://www.aisi.re.kr/eng〔2025/6/18 確認〕
※ 108 Ministère de l'Économie des Finances et de la Souveraineté industrielle et numérique：La France se dote d'un Institut national pour l'évaluation et la sécurité de l'intelligence artificielle (INESIA) https://www.economie.gouv.fr/actualites/la-france-se-dote-dun-institut-national-pour-levaluation-et-la-securite-de-lintelligence〔2025/6/18 確認〕
※ 109 European Commission：European AI Office https://digital-strategy.ec.europa.eu/en/policies/ai-office〔2025/6/18 確認〕
※ 110 JETRO：米商務省、AI 安全研究所コンソーシアムの設置発表 https://www.jetro.go.jp/biznews/2024/02/eb4c5eff3da218bd.html〔2025/6/18 確認〕
※ 111 The White House：FACT SHEET: Biden-Harris Administration Secures Voluntary Commitments from Leading Artificial Intelligence Companies to Manage the Risks Posed by AI https://bidenwhitehouse.archives.gov/briefing-room/statements-releases/2023/07/21/fact-sheet-biden-harris-administration-secures-voluntary-commitments-from-leading-artificial-intelligence-companies-to-manage-the-risks-posed-by-ai/〔2025/6/18 確認〕
※ 112 GOV.UK：AI Safety Summit 2023 https://www.gov.uk/government/topical-events/ai-safety-summit-2023〔2025/6/18 確認〕
※ 113 GOV.UK：AI Safety Summit 2023: The Bletchley Declaration https://www.gov.uk/government/publications/ai-safety-summit-2023-the-bletchley-declaration〔2025/6/18 確認〕
※ 114 GOV.UK：AI Seoul Summit 2024 https://www.gov.uk/government/topical-events/ai-seoul-summit-2024〔2025/6/18 確認〕
※ 115 GOV.UK：Seoul Declaration for safe, innovative and inclusive AI: AI Seoul Summit 2024 https://www.gov.uk/government/publications/seoul-declaration-for-safe-innovative-and-inclusive-ai-ai-seoul-summit-2024〔2025/6/18 確認〕
※ 116 GOV.UK：Frontier AI Safety Commitments, AI Seoul Summit 2024 https://www.gov.uk/government/publications/frontier-ai-safety-commitments-ai-seoul-summit-2024〔2025/6/18 確認〕
※ 117 https://aisi.go.jp/effort/effort_information/250207_1/〔2025/6/18 確認〕
※ 118 The American Presidency Project Republican Party：2024 Republican Party Platform https://www.presidency.ucsb.edu/documents/2024-republican-party-platform〔2025/6/18 確認〕
※ 119 Federal Register：Executive Order 14148 Initial Rescissions of Harmful Executive Orders and Actions https://www.federalregister.gov/documents/2025/01/28/2025-01901/initial-rescissions-of-harmful-executive-orders-and-actions〔2025/6/18 確認〕
※ 120 Federal Register：Executive Order 14179 Removing Barriers to American Leadership in Artificial Intelligence https://www.federalregister.gov/documents/2025/01/31/2025-02172/removing-barriers-to-american-leadership-in-artificial-intelligence〔2025/6/18 確認〕
※ 121 GOV.UK：AI Opportunities Action Plan https://www.gov.uk/government/publications/ai-opportunities-action-plan〔2025/6/18 確認〕
※ 122 GOV.UK：Tackling AI security risks to unleash growth and deliver Plan for Change https://www.gov.uk/government/news/tackling-ai-security-risks-to-unleash-growth-and-deliver-plan-for-change〔2025/6/18 確認〕
※ 123 この動きの直前に、AI が生成した子供の性的虐待画像を取り締まる法案が国会に提出されている。
BBC NEWS JAPAN：イギリス、AI 生成の子ども性虐待画像を取り締まる法案を発表 https://www.bbc.com/japanese/articles/ceve7dk8e38o〔2025/6/18 確認〕
※ 124 Artificial Intelligence Action Summit：https://www.elysee.fr/en/sommet-pour-l-action-sur-l-ia〔2025/6/18 確認〕
※ 125 elysee.fr：Statement on Inclusive and Sustainable Artificial Intelligence for People and the Planet https://www.elysee.fr/en/emmanuel-macron/2025/02/11/statement-on-inclusive-and-sustainable-artificial-intelligence-for-people-and-the-planet〔2025/6/18 確認〕
※ 126 BBC NEWS JAPAN：AI アクションサミット、英米は共同声明に署名せず https://www.bbc.com/japanese/articles/czx8ze7lx9no〔2025/6/18 確認〕
※ 127 GOV.UK：Frontier AI: capabilities and risks – discussion paper https://www.gov.uk/government/publications/frontier-ai-capabilities-and-risks-discussion-paper〔2025/6/18 確認〕
GOV.UK：International Scientific Report on the Safety of Advanced AI: interim report https://www.gov.uk/government/publications/international-scientific-report-on-the-safety-of-advanced-ai〔2025/6/18 確認〕
GOV.UK：International AI Safety Report 2025 https://www.gov.uk/government/publications/international-ai-safety-report-2025〔2025/6/18 確認〕
※ 128 SlashNext, Inc.：SlashNext's 2023 State of Phishing Report Reveals a 1,265% Increase in Phishing Emails Since the Launch of ChatGPT in November 2022, Signaling a New Era of Cybercrime Fueled by Generative AI https://slashnext.com/press-release/slashnexts-2023-state-of-phishing-report-reveals-a-1265-increase-in-phishing-emails-since-the-launch-of-chatgpt-in-november-2022-signaling-a-new-era-of-cybercrime-fueled-by-generative-ai/〔2025/6/18 確認〕
※ 129 日経クロステック：ガードレールなしの生成 AI が相次ぎ出現、サイバー犯罪者「御用達」の使い道とは https://xtech.nikkei.com/atcl/nxt/column/18/00676/080600141/〔2025/6/18 確認〕
※ 130 善意の開発・運用者が作った AI システムの狭義の AI セキュリティを高めるという意味ではこの限りではないが、最初から悪意を持ってそのための AI を開発することを AI ガバナンスで防ぐことはできない。
※ 131 NIST Glossary：Cyber Threat https://csrc.nist.gov/glossary/term/cyber_threat〔2025/6/18 確認〕
※ 132 ENISA は Threat Landscape という脅威サーベイ報告を毎年発刊している。また、中長期の脅威傾向の予測も行っている。詳しくは次の Web サイトを参照。
ENISA：Threat Landscape https://www.enisa.europa.eu/topics/cyber-threats/threat-landscape〔2025/6/18 確認〕
※ 133 Centre for Emerging Technology and Security：Research and Publications https://cetas.turing.ac.uk/research-and-publications〔2025/6/18 確認〕
※ 134 ITmedia：生成 AI の弱点が相次ぎ発覚 ChatGPT や Gemini がサイバー攻撃の標的に 情報流出や不正操作の恐れも https://www.itmedia.co.jp/news/articles/2403/22/news069.html〔2025/6/18 確認〕
※ 135 Security NEXT：機械学習フレームワーク「PyTorch」に不正プログラム混入のおそれ https://www.security-next.com/142623〔2025/6/18 確認〕
※ 136 Wiz：Wiz Research finds architecture risks that may compromise AI-as-a-Service providers and consequently risk customer data; works with Hugging Face on mitigations https://www.wiz.io/blog/wiz-and-hugging-face-address-risks-to-ai-infrastructure〔2025/6/18 確認〕
※ 137 ITmedia：DeepSeek、チャット履歴含む 100 万件超のログが外

部から閲覧できた可能性　米セキュリティ企業が指摘　https://www.itmedia.co.jp/news/articles/2501/30/news173.html〔2025/6/18確認〕
※ 138 SlashNext, Inc.：The State of Phishing 2024 Report　https://slashnext.com/the-state-of-phishing-2024/〔2025/6/18確認〕
※ 139 日本経済新聞：［FT］テレビ会議、AI技術でなりすまし　英企業40億円被害　https://www.nikkei.com/article/DGXZQOCB056460V00C24A6000000/〔2025/6/18確認〕
※ 140 NHK：AIが詐欺師に!? チャットサービス悪用の新手口 被害も　https://www3.nhk.or.jp/news/special/tag-digital_deceive/article/33_01.html〔2025/6/18確認〕
※ 141 BBC NEWS JAPAN：ルーマニアの憲法裁判所、大統領選第1回投票を無効と判断　勝利候補への影響工作が明るみに　https://www.bbc.com/japanese/articles/cg4zzk1d1nxo〔2025/6/18確認〕
※ 142 JETRO：ルーマニア大統領選、ロシア介入やSNS不正操作で憲法裁判所が無効判断　https://www.jetro.go.jp/biznews/2024/12/5539b706af134586.html〔2025/6/18確認〕
Președintele României：Analiza unor riscuri la adresa securității nationale generate de actiunile unor actori cibernetici statali si non-statali asupra unor infrastructuri IT&C, suport pentrú procesul electoral　https://www.presidency.ro/files/userfiles/Documente%20CSAT/Document%20CSAT%20SIE.pdf〔2025/6/18確認〕
※ 143 Reuters：アングル：ルーマニア大統領選、親ロ極右候補躍進でTikTokに疑惑の目　https://jp.reuters.com/economy/2RCUJT3UGRLBFFP62GUM4KCQFE-2024-12-04/〔2025/6/18確認〕
※ 144 読売新聞オンライン：フォロワー5万人のインフルエンサー「報酬もらった」…ルーマニア大統領選巡り「関与を後悔」　https://www.yomiuri.co.jp/world/20241209-OYT1T50188/〔2025/6/18確認〕
※ 145 Centre for Emerging Technology and Security：Evaluating Malicious Generative AI Capabilities　https://cetas.turing.ac.uk/publications/evaluating-malicious-generative-ai-capabilities〔2025/6/18確認〕
※ 146 NIST SP 800-218：Secure Software Development Framework (SSDF) Version 1.1: Recommendations for Mitigating the Risk of Software Vulnerabilities　https://doi.org/10.6028/NIST.SP.800-218〔2025/6/18確認〕
※ 147 NIST SP 800-218A：Secure Software Development Practices for Generative AI and Dual-Use Foundation Models: An SSDF Community Profile　https://doi.org/10.6028/NIST.SP.800-218A〔2025/6/18確認〕
※ 148 NCSC：Machine learning principles　https://www.ncsc.gov.uk/collection/machine-learning-principles〔2025/6/18確認〕
※ 149 Joint Cybersecurity Information（U/OO/143395-24）：Deploying AI Systems Security　https://media.defense.gov/2024/apr/15/2003439257/-1/-1/0/csi-deploying-ai-systems-securely.pdf〔2025/6/18確認〕
※ 150 CSET：Securing Critical Infrastructure in the Age of AI　https://cset.georgetown.edu/publication/securing-critical-infrastructure-in-the-age-of-ai/〔2025/6/18確認〕
※ 151 総務省：情報通信白書令和5年版　第1部第3節　インターネット上での偽・誤情報の拡散等　https://www.soumu.go.jp/johotsusintokei/whitepaper/ja/r05/html/nd123140.html〔2025/7/16確認〕
※ 152 大澤淳：「サイバー領域の安全保障政策の方向性」株式会社ウェッジ、2024年1月、「新領域安全保障」、p.185
※ 153 https://rm.coe.int/information-disorder-toward-an-interdisciplinaryframework-for-researc/168076277c〔2025/7/16確認〕
※ 154 一般社団法人セーファーインターネット協会：Disinformation対策フォーラム報告書　https://www.saferinternet.or.jp/wordpress/wp-content/uploads/Disinformation_report.pdf〔2025/7/16確認〕
※ 155 EEAS：1st EEAS Report on Foreign Information Manipulation and Interference Threats　https://www.eeas.europa.eu/eeas/1st-eeas-report-foreign-information-manipulation-and-interference-threats_en〔2025/7/16確認〕
※ 156 https://www3.weforum.org/docs/WEF_The_Global_Risks_Report_JP_2024.pdf〔2025/7/16確認〕
※ 157 UNESCO：Survey on the impact of online disinformation and hate speech　https://www.unesco.org/sites/default/files/medias/fichiers/2023/11/unesco_ipsos_survey.pdf〔2025/7/16確認〕
※ 158 NHK放送文化研究所：シリーズ「ロイター・デジタルニュースリポート2024」(3)～偽情報・誤情報に対する意識は～【研究員の視点】#552　https://www.nhk.or.jp/bunken-blog/200/671694.html〔2025/7/16確認〕
※ 159 Tobias Bunde, Sophie Eisentraut, and Leonard Schütte, “Munich Security Index 2024,” in: Tobias Bunde/Sophie Eisentraut/Leonard Schütte (eds.), Munich Security Report 2024: Lose-Lose?, Munich: Munich Security Conference, February 2024, 26–45, doi.org/10.47342/BMQK9457, 32–33.
※ 160 Munich Security Conference：AI-pocalypse Now?　https://securityconference.org/en/publications/analyses/ai-pocalypse-disinformation-super-election-year/#:~:text=AI,a%20global%20scale%E2%80%9D%20in%202024〔2025/7/16確認〕
※ 161 U.S. NAVAL INSTITUTE：Future Warfare: The Rise of Hybrid Wars　https://www.usni.org/magazines/proceedings/2005/november/future-warfare-rise-hybrid-wars〔2025/7/16確認〕
※ 162 大澤淳「サイバー情報操作の脅威から日本をどう守るのか」中央公論新社、中央公論、2022年4月号、pp.154-161
※ 163 公益財団法人笹川平和財団安全保障研究グループ：政策提言“外国からのディスインフォメーションに備えを!～サイバー空間の情報操作の脅威～”　https://www.spf.org/global-data/user172/cyber_security_2021_web1.pdf〔2025/7/16確認〕
※ 164 EEAS：Information Integrity and Countering Foreign Information Manipulation & Interference (FIMI)　https://www.eeas.europa.eu/eeas/information-integrity-and-countering-foreign-information-manipulation-interference-fimi_en#:~:text=match%20at%20L791%20evolve%20and,society%20responses%20to%20the%20threat〔2025/7/16確認〕
※ 165 Hybrid CoE：Hybrid threats as a concept　https://www.hybridcoe.fi/hybrid-threats-as-a-phenomenon/〔2025/7/16確認〕
※ 166 EEAS：1st EEAS Report on Foreign Information Manipulation and Interference Threats　https://www.eeas.europa.eu/sites/default/files/documents/2023/EEAS-DataTeam-ThreatReport-2023..pdf
※ 167 Meta社「Q3 2024 ADVERSARIAL THREAT REPORT『Integrity Reports, Third Quarter 2024』（https://transparency.meta.com/ja-jp/integrity-reports-q3-2024/〔2025/7/16確認〕）において「Adversarial Threat Report」の「You can read the full report here.」の「here」をクリックすると表示される。
※ 168 Google LLC：TAG Bulletin: Q2 2024　https://blog.google/threat-analysis-group/tag-bulletin-q2-2024〔2025/7/16確認〕
※ 169 OpenAI社：Influence and cyber operations: an update　https://cdn.openai.com/threat-intelligence-reports/influence-and-cyber-operations-an-update_October-2024.pdf〔2025/7/16確認〕
※ 170 OpenAI社：Disrupting a Covert Iranian Influence Operation.　https://openai.com/index/disrupting-a-covert-iranian-influence-operation/〔2025/7/16確認〕
OpenAI社：Disrupting malicious uses of our models: an update February 2025　https://cdn.openai.com/threat-intelligence-reports/disrupting-malicious-uses-of-our-models-february-2025-update.pdf〔2025/7/16確認〕
※ 171 BBC NEWS JAPAN：【解説】イギリスの騒乱はなぜ起きたのか　https://www.bbc.com/japanese/articles/cp8nl49lpypo〔2025/7/16確認〕
※ 172 NPR：UK's worst riots in years were incited by online disinformation about asylum seekers　https://www.npr.org/2024/08/10/nx-s1-5066896/uks-worst-riots-in-years-were-incited-by-online-disinformation-about-asylum-seekers〔2025/7/16確認〕
※ 173 ODNI：100 Days Until Election 2024　https://www.dni.gov/files/FMIC/documents/ODNI-Election-Security-Update-20240729.pdf〔2025/7/16確認〕
※ 174 ODNI：Joint ODNI, FBI, and CISA Statement on Russian Election Influence Efforts　https://www.odni.gov/index.php/newsroom/press-releases/press-releases-2024/4014-pr-28-24〔2025/7/16確認〕
※ 175 FDD：America Resilient in the Face of Aggressive Foreign Malign Influence Targeting the 2024 U.S. Elections　https://www.fdd.org/analysis/2024/12/18/america-resilient-in-the-face-of-aggressive-foreign-malign-influence-targeting-the-2024-u-s-elections/〔2025/7/16確認〕
※ 176 U.S. Department of the Treasury：Treasury Sanctions Entities in Iran and Russia That Attempted to Interfere in the U.S. 2024 Election　https://home.treasury.gov/news/press-releases/jy2766〔2025/7/16確認〕
※ 177 NPR：China is pushing divisive political messages online using fake U.S. voters　https://www.npr.org/2024/09/03/nx-s1-5096151/china-tiktok-x-fake-voters-influence-campaign〔2025/7/16確認〕

※ 178 U.S. Department of Justice：Three IRGC Cyber Actors Indicted for 'Hack-and-Leak' Operation Designed to Influence the 2024 U.S. Presidential Election https://www.justice.gov/archives/opa/pr/three-irgc-cyber-actors-indicted-hack-and-leak-operation-designed-influence-2024-us〔2025/7/16 確認〕
※ 179 NPR：Foreign influence efforts reached a fever pitch during the 2024 elections https://www.npr.org/2024/11/09/nx-s1-5181965/2024-election-foreign-influence-russia-china-iranp〔2025/7/16 確認〕
※ 180 US Department of Justice：Justice Department Announces Murder-For-Hire and Related Charges Against IRGC Asset and Two Local Operatives https://www.justice.gov/archives/opa/pr/justice-department-announces-murder-hire-and-related-charges-against-irgc-asset-and-two〔2025/7/16 確認〕
※ 181 U.S. Government Publishing Office：118th Congress 2nd Session：S. 5365 To require the President to notify Congress and take certain actions in response to any attempt by a country of concern to affect United States elections. https://www.congress.gov/118/bills/s5365/BILLS-118s5365is.htm#:~:text=,intelligence%20agencies%20stated%20the〔2025/7/16 確認〕
※ 182 GMF：Bribes and Lies: Foreign Interference in Europe in 2024 https://securingdemocracy.gmfus.org/bribes-and-lies-foreign-interference-in-europe-in-2024/〔2025/7/16 確認〕
※ 183 EEAS：3rd EEAS Report on Foreign Information Manipulation and Interference Threats https://www.eeas.europa.eu/eeas/3rd-eeas-report-foreign-information-manipulation-and-interference-threats-0_en〔2025/7/16 確認〕
GMF：Bribes and Lies: Foreign Interference in Europe in 2024 https://securingdemocracy.gmfus.org/bribes-and-lies-foreign-interference-in-europe-in-2024/〔2025/7/16 確認〕
※ 184 Spiegel：Alternative gegen Deutschland https://www.spiegel.de/politik/deutschland/afd-spionageaffaere-russland-und-china-im-fokus-neue-enthuellungen-belasten-die-partei-a-46042b96-2d61-4bb4-ac25-ead57d7d6285〔2025/7/16 確認〕
※ 185 U.S. Department of Justics：Justice Department Leads Efforts Among Federal, International, and Private Sector Partners to Disrupt Covert Russian Government-Operated Social Media Bot Farm https://www.justice.gov/archives/opa/pr/justice-department-leads-efforts-among-federal-international-and-private-sector-partners〔2025/7/16 確認〕
※ 186 EURACTIV：Russia targets social media during French legislative campaign https://www.euractiv.com/section/politics/news/russia-targets-social-media-during-french-legislative-campaign/〔2025/7/16 確認〕
※ 187 France Diplomacy：Foreign digital interference – Result of investigations into the Russian propaganda network Portal Kombat (15 February 2024) https://www.diplomatie.gouv.fr/en/french-foreign-policy/security-disarmament-and-non-proliferation/news/2024/article/foreign-digital-interference-result-of-investigations-into-the-russian〔2025/7/16 確認〕
※ 188 Secrétariat général de la défense et de la sécurité nationale：PORTAL KOMBAT A structured and coordinated pro-Russian propaganda network https://www.sgdsn.gouv.fr/files/files/Publications/20240214_NP_SGDSN_VIGINUM_PORTAL-KOMBAT-NETWORK_PART2_ENG_VF.pdf〔2025/7/16 確認〕
※ 189 Recorded Future：Sombres Influences：Russian and Iranian Influence Networks Target French Elections https://go.recordedfuture.com/hubfs/reports/TA-2024-0628.pdf〔2025/7/16 確認〕
※ 190 France Diplomacy：France strongly condemns Russian actors and their intermediaries implicated in the Storm-1516 manoeuvres (6 May 2025) https://www.diplomatie.gouv.fr/en/country-files/russia/news/2025/article/france-strongly-condemns-russian-actors-and-their-intermediaries-implicated-in〔2025/7/16 確認〕
※ 191 BBC NEWS JAPAN：オーストリア下院、首相不信任案を可決 スキャンダルで https://www.bbc.com/japanese/48428991〔2025/7/16 確認〕
※ 192 The Record：Austria uncovers alleged Russian disinformation campaign spreading lies about Ukraine https://therecord.media/austria-uncovers-russian-disinfo-campaign〔2025/7/16 確認〕
※ 193 The Wall Street Journal：A Den of Spies: Vienna Emerges as Hub for Russian Espionage https://www.wsj.com/world/a-den-of-spies-vienna-emerges-as-hub-for-russian-espionage-9dda8b4d?st=2fd7fnlb9sgbnqs〔2025/7/16 確認〕
※ 194 BBC：Romania hit by major election influence campaign and Russian cyber-attacks https://www.bbc.com/news/articles/cgq18w507dko〔2025/7/16 確認〕
※ 195 Curtea Constituțională a României：COMUNICAT DE PRESĂ, 6 decembrie 2024 https://www.ccr.ro/comunicat-de-presa-6-decembrie-2024/〔2025/7/16 確認〕
※ 196 Curtea Constituțională a României：privind anularea procesului electoral cu privire la alegerea Președintelui României din anul 2024 https://web.archive.org/web/20241206195537/https://www.ccr.ro/wp-content/uploads/2024/12/HCC-32-2024.pdf〔2025/7/16 確認〕
※ 197 Președintele României：Analiza unor riscuri la adresa securității nationale generate de actiunile unor actori cibernetici statali si non-statali asupra unor infrastructuri IT&C, suport pentrú procesul electoral https://www.presidency.ro/files/userfiles/Documente%20CSAT/Document%20CSAT%20SIE.pdf〔2025/7/16 確認〕
※ 198 Democratic Erosion Consortium：Russia fueling Democratic Struggles in Romania through the Media https://democratic-erosion.org/2025/04/24/russia-fueling-democratic-struggles-in-romania-through-the-media/〔2025/7/16 確認〕
※ 199 Reuters：EU opens investigation into TikTok over election interference https://www.reuters.com/business/eu-opens-investigation-into-tiktok-over-election-interference-2024-12-17/〔2025/7/16 確認〕
※ 200-1 Ruters：Romania braces for wave of disinformation ahead of election second round https://www.reuters.com/world/europe/romania-braces-wave-disinformation-ahead-election-second-round-2025-05-05〔2025/7/16 確認〕
※ 200-2 OpenMinds：End of Democracy: How Pro-Russian Telegram Channels Influence Romanian Elections https://www.openminds.ltd/reports/end-of-democracy-how-pro-russian-telegram-channels-influence-romanian-elections〔2025/7/25 確認〕
※ 201 Svidomi：One fake news story fuelled riots across the country. How Russian propaganda affects the UK https://svidomi.in.ua/en/page/one-fake-news-story-fuelled-riots-across-the-country-how-russian-propaganda-affects-the-uk〔2025/5/19 確認〕
※ 202 The Telegraph：The obscure Russian-linked 'news' outlet fuelling violence on Britain's streets https://www.telegraph.co.uk/news/2024/08/03/obscure-russian-linked-news-outlet-fuelling-violence/〔2025/7/16 確認〕
※ 203 The Bureau of Investigative Journalism：Did Russian disinformation fuel the Southport protests? https://www.thebureauinvestigates.com/stories/2024-08-02/did-russian-disinformation-fuel-the-southport-protests/〔2025/7/16 確認〕
※ 204 Independent：Former security minister raises concerns Putin behind Southport far-right disinformation https://www.independent.co.uk/news/uk/politics/southport-far-right-disinformation-russia-b2589041.html〔2025/7/16 確認〕
※ 205 openDemocracy：Great Replacement & boogaloo: The ideology driving the modern far right https://www.opendemocracy.net/en/far-right-riots-great-replacement-boogaloo/〔2025/7/16 確認〕
※ 206 The Soufan Center：QUANTIFYING THE Q CONSPIRACY: A Data-Driven Approach to Understanding the Threat Posed by Qanon https://thesoufancenter.org/wp-content/uploads/2021/04/TSC-White-Paper_QAnon_16April2021-final-1.pdf〔2025/7/16 確認〕
※ 207 GOV.UK：UK children and adults to be safer online as world-leading bill becomes law https://www.gov.uk/government/news/uk-children-and-adults-to-be-safer-online-as-world-leading-bill-becomes-law〔2025/7/16 確認〕
※ 208 UK Parliament：Social media, misinformation and harmful algorithms https://committees.parliament.uk/work/8641/social-media-misinformation-and-harmful-algorithms/news/〔2025/7/16 確認〕
※ 209 FOCUS TAIWAN：China steps up disinformation campaign in 2024: NSB report https://focustaiwan.tw/politics/202501030012〔2025/7/16 確認〕
※ 210 Reuters：Chinese state media stoked allegation Taiwan's president would flee war https://www.reuters.com/world/asia-pacific/chinese-state-media-stoked-allegation-taiwans-president-would-flee-war-2024-04-01/〔2025/7/16 確認〕
※ 211 産経新聞：台湾・頼清徳政権に偽情報攻撃 「日米と組んで『独立』図る」 認知戦拡大を専門家ら警戒 https://www.sankei.com/

article/20240922-J42HRMJH7ZKUZDICHKXPDCHKPM/〔2025/7/16 確認〕
※ 212 Taiwan Factcheck Center：Who are Japanese Taiwanese? The Chinese disinformation that fixated on the ties between Taiwan and Japan https://en.tfc-taiwan.org.tw/who-are-japanese-taiwanese-the-chinese-disinformation-that-fixated-on-the-ties-between-taiwan-and-japan/〔2025/7/16 確認〕
※ 213 Taiwan Factcheck Center：【錯誤】網傳照片「總統賴清德在金門致詞，講台背板出現日本軍旗」？ https://tfc-taiwan.org.tw/fact-check-reports/migration-10958/〔2025/7/16 確認〕
※ 214 FORMOSA NEWS：China reportedly offered payment for propaganda piece on Taiwan-Tuvalu relations https://english.ftvnews.com.tw/news/2024128W06EA〔2025/7/16 確認〕
※ 215 Islands Business：Taiwan envoy says Tuvalu ties 'rock solid' post-election https://islandsbusiness.com/news-break/taiwan-envoy-says-tuvalu-ties-rock-solid-post-election/〔2025/7/16 確認〕
※ 216 FOCUS TAIWAN：Taiwan's embassy in Tuvalu blasts Beijing over 'disinformation' https://focustaiwan.tw/politics/202501150004〔2025/7/16 確認〕
※ 217 オーストラリア戦略政策研究所（Australian Strategic Policy Institute）：Russia and China co-ordinate on disinformation in Solomon Islands elections https://www.aspistrategist.org.au/russia-and-china-co-ordinate-on-disinformation-in-solomon-islands-elections/〔2025/7/16 確認〕
※ 218 Sputnik：Is US Plotting Electoral Coup in Solomon Islands? https://sputnikglobe.com/20240409/is-us-plotting-electoral-coup-in-solomon-islands-1117758198.html〔2025/7/16 確認〕
※ 219 Global Times：Allegations of US interference emerge ahead of pivotal election in Solomon Islands https://www.globaltimes.cn/page/202404/1310521.shtml〔2025/7/16 確認〕
※ 220 The Intercept：Hacked Russian Files Reveal Propaganda Agreement With China https://theintercept.com/2022/12/30/russia-china-news-media-agreement/〔2025/7/16 確認〕
※ 221 EEAS：3rd EEAS Report on Foreign Information Manipulation and Interference Threats https://www.eeas.europa.eu/sites/default/files/documents/2025/EEAS-3nd-ThreatReport-March-2025-05-Digital-HD.pdf〔2025/7/16 確認〕
International Republican Institute：The Authoritarian Nexus – How Russia and China Undermine Democracy Worldwide https://www.iri.org/wp-content/uploads/2024/04/The-Authoritarian-Nexus.pdf〔2025/7/16 確認〕
※ 222 外務省：日・ウクライナ経済復興推進会議 https://www.mofa.go.jp/mofaj/erp/c_see/ua/pageit_000001_00299.html〔2025/7/16 確認〕
※ 223 SOMPO CYBER SECURITY：親ロシア・グループの日本への攻撃（2024 年 2 月） https://www.sompocybersecurity.com/column/column/pro-russia-hacktivists-ddos-japan-2024-feb〔2025/7/16 確認〕
※ 224 NHK：岸田首相の偽画像などが SNS で相次ぎ拡散 注意を呼びかけ https://www3.nhk.or.jp/news/special/article/society20240218-01.html〔2025/7/16 確認〕
※ 225 NHK：軍事侵攻 2 年 ロシア支持サイトなどの記事 拡散割合増える https://www3.nhk.or.jp/news/html/20240222/k10014368131000.html〔2025/7/16 確認〕
※ 226 読売新聞オンライン：選挙イヤー2024で強まるインフルエンスオペレーション（影響力工作）への警戒感 https://www.yomiuri.co.jp/column/matsurigoto/20240313-OYT8T50001/3/〔2025/7/16 確認〕
※ 227 CYBLE：NATO's 75th Anniversary Washington Summit Draws Ire of Hacktivist Groups https://cyble.com/blog/natos-75th-anniversary-washington-summit-draws-ire-of-hacktivist-groups/〔2025/7/16 確認〕
※ 228 Graphika Technologies, Inc.：Summit Old, Summit New https://22006778.fs1.hubspotusercontent-na1.net/hubfs/22006778/Report%20PDFs/graphika_report_summit_old_summit_new.pdf?utm_campaign=Report%20Demand%20Gen&utm_medium=email&_hsenc=p2ANqtz-_FHUKYRyaYcED9iUc17fp8Gh48e6z62SV8y4yiyhXuFDoqvvVWRxRF-9oVdtVRIcwAPSrF4WvklrBc-kYuDhWnjyM7Wqb-H5-uBVBG_e3r9W19RpU&_hsmi=315775437&utm_content=315775437&utm_source=hs_automation〔2025/7/16 確認〕
※ 229 VOA：Russia-tailed AI-generated deepfake videos target US presidential elections, NATO https://www.voanews.com/a/7633946.html〔2025/7/16 確認〕
※ 230 SOMPO CYBER SECURITY：【ブログ】親ロシア・ハクティビストが日本の金融機関・政党・鉄道会社等 Web サイトを攻撃（2024 年 7 月） https://www.sompocybersecurity.com/column/column/pro-russia-hacktivists-ddos-japan-2024-jul〔2025/7/16 確認〕
※ 231 EEAS：Paint It Black – Pro-Kremlin take on the NATO Summit https://euvsdisinfo.eu/paint-it-black-pro-kremlin-take-on-the-nato-summit/〔2025/7/16 確認〕
※ 232 Sputnik:ウクライナが 10 年以内に NATO に加盟するとは誰も言っていない＝NATO 事務総長 https://x.com/sputnik_jp/status/1810143279754977488〔2025/7/16 確認〕
※ 233 Sputnik：NATO 諸国が警戒する、「対ロシア」で越えてはならないデッドライン＝元自衛官、矢野義昭氏 https://sputniknews.jp/20240628/nato-18723847.html〔2025/7/16 確認〕
※ 234 Sputnik：高まる NATO 東京事務所開設の可能性 岸田氏の NATO サミット出席 https://sputniknews.jp/20240706/natonato-18768338.html〔2025/7/16 確認〕
※ 235 日本ファクトチェックセンター：パンデミック条約でワクチン強制接種？ 繰り返し否定されている誤情報【ファクトチェック】 https://www.factcheckcenter.jp/fact-check/health/false-pandemic-treaty-vaccine-mandate/〔2025/7/16 確認〕
※ 236 日本ファクトチェックセンター：新型コロナのレプリコンワクチンは死亡率がファイザー製の 75 倍？元資料の誤読【ファクトチェック】 https://www.factcheckcenter.jp/fact-check/health/false-replicon-vaccine-claim/.〔2025/7/16 確認〕
※ 237 PRESIDENT Online：「コロナワクチンで 50 万人が死亡」「日本で人体実験している」…反ワク派の主張を専門家と徹底検証した結果 https://president.jp/articles/-/90457?page=1〔2025/7/16 確認〕
※ 238 Radio Free Asia：Did Japan classify mRNA vaccines 'deadliest drug' in history? https://www.rfa.org/english/factcheck/2024/12/09/afcl-japan-mrna-vaccine-covid/〔2025/7/16 確認〕
※ 239 U.S. Embassy & Consulates in Italy：Fact vs. Fiction: Russian Disinformation on Ukraine https://it.usembassy.gov/fact-vs-fiction-russian-disinformation-on-ukraine/〔2025/7/25 確認〕
※ 240 The New York Times：Deepfake of U.S. Official Appears After Shift on Ukraine Attacks in Russia https://www.nytimes.com/2024/05/31/us/politics/deepfake-us-official-russia.html〔2025/7/16 確認〕
※ 241 National Security Archive：The Kremlin's Efforts to Covertly Spread Disinformation in Latin America https://nsarchive.gwu.edu/document/32130-25-state-dept-kremlin-covert-disinformation-latin-america〔2025/7/16 確認〕
※ 242 Prosperity Institute：Information at War: From China's Three Warfares to NATO's Narratives https://www.prosperity.com/media-publications/information-at-war-from-chinas-three-warfares-to-natos-narratives/〔2025/7/16 確認〕
※ 243-1 The Associated Press：Misinformation about the Israel-Hamas war is flooding social media. Here are the facts https://apnews.com/article/israel-hamas-gaza-misinformation-fact-check-e58f9ab8696309305c3ea2bfb269258e?utm_source=Email&utm_medium=share〔2025/7/16 確認〕
※ 243-2 BBC NEWS JAPAN：【解説】「ウクライナは生物兵器を開発している」 ロシアの主張をファクトチェック https://www.bbc.com/japanese/features-and-analysis-60733307〔2025/7/25 確認〕
※ 244 RAND Corporation：Lies, Misinformation Play Key Role in Israel-Hamas Fight https://www.rand.org/pubs/commentary/2023/10/lies-misinformation-play-key-role-in-israel-hamas-fight.html〔2025/7/16 確認〕
※ 245 PBS NEWS：How misinformation about Israel and Gaza has evolved in the yearlong war https://www.pbs.org/newshour/world/how-misinformation-about-israel-and-gaza-has-evolved-in-the-yearlong-war〔2025/7/16 確認〕
※ 246 日本ファクトチェックセンター：イスラエル・パレスチナ情勢をめぐり大量の誤情報 / 偽情報 検証方法を解説【ファクトチェックまとめ】 https://www.factcheckcenter.jp/fact-check/international/israel-palestine-conflict-fact-check-summary/〔2025/7/16 確認〕
※ 247 The New York Times：In a Worldwide War of Words, Russia, China and Iran Back Hamas https://www.nytimes.com/2023/11/03/technology/israel-hamas-information-war.html〔2025/7/16 確認〕
※ 248 POLITICO：Russia-linked fake videos spread German election fraud claims, authorities warn https://www.politico.eu/article/russia-linked-fake-videos-spread-german-election-fraud-claims-authorities-warn/〔2025/7/16 確認〕

付録

第20回 IPA「ひろげよう情報セキュリティコンクール」2024 受賞作品

ひろげよう情報セキュリティコンクールは、情報セキュリティをテーマとした作品制作を通じて、全国における児童・生徒等の情報セキュリティに関する意識醸成と興味喚起を図ることを目的として開催しています。ここでは、全30,636点の応募作品の中から、IPAが授与している最優秀賞と優秀賞をご紹介いたします。

最優秀賞

〈標語部門〉

パスワード
意味ない配列
意味がある

板野 早希さん　東京都 東京都立上野高等学校

〈ポスター部門〉

多要素認証があなたを守る

岩永 陽翔さん　東京都 国際基督教大学高等学校

優秀賞

〈標語部門〉

パスワード　よりふくざつに　足すワード

佐藤 海璃さん
宮城県 南三陸町立志津川小学校

謎メール　軽いクリック　重い代償

酒井 翔琉さん
茨城県 北茨城市立中郷中学校

多要素認証　そのひと手間が　漏洩防ぐ

一ノ瀬 玲央さん
北海道 北海道旭川東高等学校

〈ポスター部門〉

タップの前に疑って!!

今岡陽菜歌さん　大阪府 大阪市立大淀小学校

覗き見に注意

井上羽南さん 茨城県 茨城県立並木中等教育学校

同じ鍵は危険です

杉本瑞季さん 愛知県 愛知県立安城南高等学校

IPAの便利なツールとコンテンツ

情報セキュリティ対策ベンチマーク
https://www.ipa.go.jp/security/sec-tools/benchmark.html

診断

用途・目的	自組織のセキュリティレベルを診断
利用対象者	情報セキュリティ担当者
特 長	• 他組織と比較した自組織のセキュリティレベルが判る • 自組織に不足しているセキュリティ対策が判る

概要

「セキュリティ対策の取り組み状況に関する評価項目」27 問と「企業プロフィールに関する評価項目」19 問、計 46 問に回答すると以下の診断結果を表示します。

■提供される診断結果
- セキュリティレベルを示したスコア（最高点 135 点、最低点 27 点）
- 企業規模、業種が自組織と近い他組織と診断項目別にスコアを比較
- 結果に応じた推奨される取り組み

脆弱性体験学習ツール「AppGoat」
https://www.ipa.go.jp/security/vuln/appgoat/

学習

用途・目的	脆弱性に関する基礎的な知識の学習
利用対象者	• アプリケーション開発者 • Web サイト管理者
特 長	脆弱性の概要や対策方法等、脆弱性に関する基礎的な知識を実習形式で体系的に学べるツール

概要

SQL インジェクション、クロスサイト・スクリプティング等 の 12 種類の Web アプリケーションに関連する脆弱性について学習できるツールです。
利用者は学習テーマ毎の演習問題に対して、埋め込まれた脆弱性の発見、プログラミング上の問題点の把握、対策手法を学べます。

■活用方法例
- Web アプリケーション用学習ツール（個人学習モード）を利用した、自宅等での個人学習
- Web アプリケーション用学習ツール（集合学習モード）を利用した、学校の講義や組織内のセミナー等、複数人での学習

脆弱性対策情報データベース「JVN iPedia」
https://jvndb.jvn.jp/

対策

用途・目的	自組織で使用しているソフトウェア製品の脆弱性の確認と対策
利用対象者	• システム管理者 • 製品・サービスの保守を担う担当者
特 長	国内外で公開されたソフトウェア製品の脆弱性対策情報が掲載された、キーワード検索可能なデータベース

概要

■掲載情報例
- 脆弱性の概要
- 脆弱性がある製品名とそのベンダー名
- 共通脆弱性識別子 CVE
- 脆弱性の深刻度 CVSS 基本値
- 本脆弱性に関わる製品ベンダー等のリンク

■活用方法例
- ネット記事等に記載された CVE 番号を JVN iPediaで検索し、脆弱性の詳細を確認
- 自組織で使用している製品名で検索し、脆弱性の詳細を確認

MyJVN バージョンチェッカ for .NET
https://jvndb.jvn.jp/apis/myjvn/vccheckdotnet.html

対策

用途・目的	パソコンにインストールされたソフトウェア製品のバージョンが最新かどうかの確認
利用対象者	パソコン利用者全般
特 長	インストールされている対象製品が最新バージョンかどうかをまとめて確認できる

概要

■判定対象ソフトウェア製品

- Adobe Reader
- JRE
- Lhaplus
- Mozilla Firefox
- Mozilla Thunderbird
- iTunes
- Lunascape
- Becky! Internet Mail
- OpenOffice.org
- VMware Player
- Google Chrome
- LibreOffice

■活用方法例

毎朝、MyJVN バージョンチェッカを実行して、使用しているソフトウェアが最新かどうかをチェックし、最新でなければそのソフトウェアを更新する

注意警戒情報サービス
https://jvndb.jvn.jp/alert/

対策

用途・目的	脆弱性対策に必要な最新情報の収集
利用対象者	• システム管理者 • 製品・サービスの保守を担う担当者
特 長	国内で広く利用され、脆弱性が悪用されると影響の大きいサーバー用オープンソースソフトウェアのリリース情報と IPA が発信する「重要なセキュリティ情報」を提供

概要

■掲載情報例

- Apache HTTP Server
- Apache Struts
- Apache Tomcat
- BIND
- Joomla!
- OpenSSL
- WordPress
- 重要なセキュリティ情報

■活用方法例

定期的に自組織で使用しているオープンソースソフトウェアのリリース情報や IPA が発信する「重要なセキュリティ情報」が公表されているかどうかを確認し、公表されていれば内容の確認、必要に応じ対応を行う

サイバーセキュリティ注意喚起サービス「icat for JSON」
https://www.ipa.go.jp/security/vuln/icat.html

対策

用途・目的	IPA が発信する「重要なセキュリティ情報」のリアルタイム取得
利用対象者	• システム管理者 • サービスの保守を担う担当者 • 個人利用者
特 長	Web ページに HTML タグを埋め込むと、Web ページから IPA が発信する「重要なセキュリティ情報」を配信

概要

■「重要なセキュリティ情報」発信例

- 利用者への影響が大きい製品の脆弱性情報
- 広く使われる製品のサポート終了情報
- サイバー攻撃への注意喚起

■活用方法例

icat を自組織の従業員がよくアクセスする Web ページ（イントラページ等）に表示させ、ソフトウェア更新等の対策を促す

MyJVN 脆弱性対策情報フィルタリング収集ツール（mjcheck4）
https://jvndb.jvn.jp/apis/myjvn/mjcheck4.html

対策

用途・目的	自組織で使用しているソフトウェア製品の脆弱性の確認と対策
利用対象者	• システム管理者 • 製品・サービスの保守を担う担当者
特 長	JVN iPedia に登録されている脆弱性対策情報をフィルタリングして自社システムに関連する脆弱性情報を効率よく収集

概要

■フィルタリング例
• 製品名　　• CVSSv3　　• 公開日　等

■活用方法例
• 自組織が利用しているオープンソースソフトウェア製品の脆弱性対策情報収集
• 情報システム部門が運用しているシステムの脆弱性対策情報の収集

Web サイトの攻撃兆候検出ツール「iLogScanner」
https://www.ipa.go.jp/security/vuln/ilogscanner/

診断

用途・目的	Web サイトに対する攻撃の痕跡、攻撃の可能性を検出
利用対象者	Web サイト運営者
特 長	Web サイトのアクセスログ、エラーログ、認証ログを解析し、攻撃の痕跡や攻撃に成功した可能性があるログを解析結果レポートに表示

概要

■アクセスログ、エラーログから検出可能な項目例
• SQL インジェクション　　• OS コマンド・インジェクション
• ディレクトリ・トラバーサル　　• クロスサイト・スクリプティング

■認証ログ（Secure Shell、FTP）から検出可能な項目例
• 大量のログイン失敗　　• 短時間の集中ログイン
• 同一ファイルへの大量アクセス　　• 認証試行回数

■活用方法例
定期的に iLogScanner を実行し、自組織の Web サイトを狙った攻撃が行われているか確認する

5 分でできる！情報セキュリティ自社診断
https://www.ipa.go.jp/security/guide/sme/5minutes.html

診断

用途・目的	自社の情報セキュリティ対策状況を診断
利用対象者	中小企業・小規模事業者の経営者、管理者、従業員
特 長	• 設問に答えるだけで自社のセキュリティ対策状況を把握することができる • 診断後は、診断結果に即した対策が確認できる

概要

「5 分でできる！情報セキュリティ自社診断」は、情報セキュリティ対策のレベルを数値化し、問題点を見つけるためのツールです。
25 の質問に答えるだけで診断することができ、解説編を参照することで、自社で対応していない場合に生じる情報セキュリティ上のリスクと、今後どのような対策を設けるべきかを把握することができます。

情報セキュリティ・ポータルサイト「ここからセキュリティ！」

学習 診断 対策

https://www.ipa.go.jp/security/kokokara/

用途・目的	• 情報セキュリティや情報リテラシーに関する情報収集 • 国内の主なレポート、ガイドライン、学習・診断等のツール等の利用
利用対象者	• インターネットの一般利用者（小学生～大人） • 企業の管理者／一般利用者
特 長	情報セキュリティ関連の民間及び公的な団体が公開する無償の資料、情報、ツールを網羅的に掲載。目的別、用途別、役割別に情報を選択し利用が可能

概要

• セキュリティベンダー、公的機関、政府等から発信される注意喚起や、資料・動画・ツール等のコンテンツを網羅的に掲載したポータルサイト
• コンテンツを「被害に遭ったら」「対策する」「教育・学習」「セキュリティチェック」「データ＆レポート」に分類。必要な情報が見つけやすい
• 教育学習は対象者を細分化し、それぞれに適した教育学習コンテンツを紹介

サイバーセキュリティ経営可視化ツール

診断

https://www.ipa.go.jp/security/economics/checktool.html

用途・目的	セキュリティ対策の実施状況のセルフチェック
利用対象者	原則として、従業員 300 名以上の企業の CISO 等、サイバーセキュリティ対策の実施責任者
特 長	サイバーセキュリティ経営ガイドライン Ver3.0 に準拠したセキュリティ対策の実施状況を成熟度モデルで自己診断し、レーダーチャートで可視化

概要

経営者がサイバーセキュリティ対策を実施する上で責任者となる担当幹部（CISO 等）に指示すべき“重要 10 項目”が、適切に実施されているかどうかを 5 段階の成熟度モデルで自己診断し、その結果をレーダーチャートで可視化するツールです。
診断結果は、経営者への自社のセキュリティ対策の実施状況の説明資料として利用できます。経営者が対策状況を定量的に把握することで、サイバーセキュリティに関する方針の策定や適切なセキュリティ投資の検討、投資家等ステークホルダとのコミュニケーション等に役立てることができます。

■提供される主な機能
・重要 10 項目の実施状況の可視化
・診断結果と業種平均との比較
・対策を実施する際の参考事例
・グループ企業同士の診断結果の比較

5 分でできる！情報セキュリティポイント学習

学習

https://www.ipa.go.jp/security/sec-tools/5mins_point.html

用途・目的	自社の情報セキュリティ教育の実施
利用対象者	中小企業の経営者、管理者、従業員等
特 長	• 自社診断の質問を 1 テーマ 5 分で学べる • インストール不要、無料の学習ツール

概要

情報セキュリティについて学習できるツールです。
身近にある職場の日常の 1 コマを取り入れた親しみやすい学習テーマで、情報セキュリティに関する様々な事例を疑似体験しながら適切な対処法を学ぶことができます。

安心相談窓口だより
https://www.ipa.go.jp/security/anshin/attention/index.html

学習 対策

用途・目的	最新の「ネット詐欺」等の手口を知り被害防止につなげる
利用対象者	スマートフォン、パソコンの一般利用者
特 長	実際に相談窓口に寄せられる、よくある相談内容に関して「手口」と「被害にあった場合の対処」「被害にあわないための対策」を学べる

概要

IPA 情報セキュリティ安心相談窓口では、寄せられる相談に関して手口を実際に検証し、そこで得られた知見をその後の相談対応にフィードバックするとともに、注意喚起等、情報発信にも活かしています。
「安心相談窓口だより」では中でも多く相談が寄せられる相談内容の「手口」「対処」「対策」について、パソコンやスマートフォンの操作等にあまり詳しくない人でも理解できるように分かりやすく説明を行っています。
記事は不定期に公開されますので、「安心相談窓口だより」を定期的に確認することで、最新のネット詐欺等の手口や対策を知り、被害の未然防止に役立てることができます。
手口に関する内容以外にも、被害にあわないための日ごろから気を付けるポイントについての記事も公開しています。

映像で知る情報セキュリティ
https://www.ipa.go.jp/security/videos/list.html

学習

用途・目的	動画の視聴により、情報セキュリティの脅威、手口、対策等を学ぶ
利用対象者	スマートフォンやパソコンを使用する一般利用者 組織の経営者、対策実践者、啓発者、従業員等
特 長	組織内の研修等で利用できる10 分前後の動画を公開。情報セキュリティ上の様々な脅威・手口、対策をドラマ等の動画を通じで学べる

概要

「サイバー攻撃」「内部不正」「ワンクリック請求」「偽警告」等の脅威をテーマにした動画のほか、「中小企業向け情報セキュリティ対策」「新入社員向け」「保護者／小学生／中高生向け」といった訴求対象者別の動画を公開しています。動画の視聴により、様々な情報セキュリティ上の脅威・手口、対策を学ぶことができます。
情報セキュリティの自己研さんを目的とした個人の視聴のほか、組織内の研修用としての利用が可能です。

■動画のタイトル例
- 今そこにある脅威～組織を狙うランサムウェア攻撃～
- 今そこにある脅威～内部不正による情報流出のリスク～
- What' s BEC ？～ビジネスメール詐欺 手口と対策～
- あなたのパスワードは大丈夫？～インターネットサービスの不正ログイン対策～

索引

F

G

I

J

L

M

N

O

P

さ

た

な

は

ま

や

ら

著作・製作　独立行政法人情報処理推進機構（IPA）

編集責任　高柳 大輔　沖田 孝裕　小山 明美　涌田 明夫　白石 歩
井上 佳春　渋谷 環

執筆者　IPA
伊藤 彰朗　伊藤 さやか　伊藤 忠彦　伊藤 吉史　井上 佳春
入来 星衣　大久保 直人　奥村 明俊　大海 健太　小川 賢一
小川 隆一　沖田 孝裕　金木 陽一　金子 成徳　加納 諒也
神谷 健司　亀山 友彦　菅野 和哉　菊池 秀一　小杉 聡志
小山 明美　小山 祐平　佐藤 栄城　渋谷 環　白石 歩
新保 淳　鷲見 拓哉　銭谷 謙吾　田島 凛　辻 宏郷
豊田 亮子　長迫 智子　西尾 秀一　野村 春佳　平本 健二
冨士 愛恵里　藤井 明宏　古居 敬大　松島 伸彰　宮本 冬美
森貞 夏樹　守屋 真人　籔口 春南　山下 恵一　吉原 正人
吉本 賢樹

三菱電機株式会社　神余 浩夫
デジタル庁 戦略・組織グループ セキュリティ危機管理チーム　中村 元洋
順天堂大学 健康データサイエンス学部　満塩 尚史
一般社団法人 JPCERT コーディネーションセンター　米澤 詩歩乃

協力者　IPA
浅見 侑太　井上 真弓　板橋 博之　伊藤 真一　江島 将和
大澤 淳　小野塚 直人　甲斐 成樹　釜谷 誠　唐亀 侑久
神田 雅透　岸野 照明　北村 弘　桐淵 直人　黒岩 俊二
桑名 利幸　佐川 陽一　貞広 憲一　篠塚 耕一　白井 綾
瀬光 孝之　高見 穣　高柳 大輔　田口 聡　田中舘 隼
田村 智和　土屋 正　遠山 真　中島 尚樹　西原 栄太郎
西村 奏一　日向 英俊　福原 聡　松岡 光　松田 修平
宮崎 卓行　安田 進　渡邉 祥樹
サイバーレスキュー隊 J-CRAT(ジェイ・クラート)
AISI 事務局 戦略・企画チーム

一般財団法人日本情報経済社会推進協会　大熊 三恵子
NRI セキュアテクノロジーズ株式会社　北原 幸彦
一般財団法人日本情報経済社会推進協会　崎村 夏彦
一般社団法人 JPCERT コーディネーションセンター　染川 夕貴
NTT 株式会社　永井 彰
国立研究開発法人情報通信研究機構　中尾 康二
総務省 サイバーセキュリティ統括官室
国立研究開発法人情報通信研究機構 サイバーセキュリティ研究所
経済産業省 商務情報政策局 サイバーセキュリティ課

おわりに

2024年度は、仕事や日々の生活での生成AIの活用が本格化し、「日常が一変」したという方も多いのではないでしょうか。 その一方で、総合エンターテインメント企業がランサムウェア攻撃で多大な被害を受けた事例のように、1回のサイバー攻撃で、いままでの「日常が一変」することも起こっています。 良くも悪くも「一変する日常」に私達は対応していかないといけない、そしてその日常を支えるのは個々人や個々の組織だけでは難しいことから、サブタイトルを「一変する日常:支える仕組みを共に築こう」としました。
IPAでは2025年3月にIoT製品のセキュリティレベルを可視化する新たな制度「セキュリティ要件適合評価及びラベリング制度（JC-STAR）」を開始し、5月には適合ラベルの交付が開始されました。サブタイトル後半の日常を「支える仕組み」の一つとして、本制度が浸透し、安全なIoT機器が積極的に選ばれることで、DDoS攻撃等のサイバー攻撃の被害を減らす一助になればと思います。

編集子

情報セキュリティ白書 2025

一変する日常：支える仕組みを共に築こう

2025 年 9 月 10 日　先行公開版発行

企画・著作・制作・発行　**独立行政法人情報処理推進機構（IPA）**
〒 113-6591
東京都文京区本駒込2丁目 28 番8号
文京グリーンコートセンターオフィス 16 階
URL　https://www.ipa.go.jp/
電話　03-5978-7503
E-Mail spd-book@ipa.go.jp

表紙デザイン／本文 DTP・編集　**伊藤 千絵、久磨 公治、涌田 明夫、北林 俊平**