

2026

サイバー防御新時代：諸刃の進化に備えよ



独立行政法人
情報処理推進機構
Innovation Platform Agency, Japan

Beyond Digital

「情報セキュリティ白書2026」の刊行にあたって

デジタルトランスフォーメーション(DX)や AI の活用が急速に広がり、サイバー空間とフィジカル空間の融合は新たな段階を迎えています。デジタル技術は、人々の暮らしや企業活動、社会基盤の在り方を大きく変え、新たな価値や成長の機会を生み出しています。更に、量子コンピューティングを始めとする先端技術の発展は、イノベーションの可能性を一層広げつつあります。

一方で、デジタル社会の発展に伴い、サイバー空間を取り巻くリスクも複雑化・深刻化しています。ランサムウェアによる被害は依然として高水準にあり、サプライチェーンや委託先を狙った攻撃、システムの脆弱性を悪用した攻撃はますます高度化・巧妙化しています。また、AI の利活用の拡大に伴う新たなサイバーリスクも顕在化しています。加えて、社会や産業のデジタル化が進む中で攻撃対象領域は拡大し、ひとたび重大なインシデントが発生した場合には、その影響が広範囲に波及する可能性も高まっています。

更に、先端技術の実用化は大きな可能性をもたらす反面、既存の暗号や認証による安全性の前提を揺るがし、セキュリティの在り方そのものに変革を迫っています。技術の進化による恩恵を最大限に生かしながら、安全・安心をいかに確保していくかが、これまで以上に重要な課題となっています。

私達が目指すのは、脅威の存在を前提としながらも、人々や組織が安心して挑戦を続けられる社会です。攻撃を受けても止まらず、回復し、更にレジリエンスを社会全体で高めていくことが、挑戦と成長を支えるセキュリティの本質です。サイバー空間は持続的な経済社会の発展を支える不可欠な基盤であり、その安全・安心はもはや個々の組織だけで守れるものではありません。社会全体で支えるべき公共的な基盤として、セキュリティをとらえることが求められています。

こうした社会的要請の高まりを受け、政府はサイバーセキュリティを危機管理と成長戦略の両面を支える重要な投資分野として位置付け、産学官が連携して人材・技術・制度の強化を進めています。IPA もまた、技術・人材・制度・信頼を結び付けるハブとして、デジタル化の進展とセキュリティ確保の両立を図り、攻撃や障害に直面しても止まらず、回復する、更なるレジリエントな社会の実現に向けて、その役割を果たしてまいります。

「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を巡る政策や脅威動向、被害の実態等を取りまとめ、多くの皆様に活用いただいております。本白書が、2025 年度の情勢への理解を深め、それぞれの立場で取り組まれる対策や実践の一助となり、より安全で信頼されるデジタル社会の実現につながることを心より願っております。

2026年9月

独立行政法人情報処理推進機構(IPA)

理事長 齊藤 裕

序章 2025年度の情報セキュリティの概況	6
第1章 国内外のサイバー脅威の動向	8
1.1 2025年度に観測されたインシデント状況	8
1.1.1 世界におけるサイバーセキュリティインシデント状況	8
1.1.2 国内におけるサイバーセキュリティインシデント状況	14
1.2 インシデント事例や脆弱性・攻撃の動向と対策	19
1.2.1 ランサム攻撃	19
1.2.2 標的型攻撃	27
1.2.3 DDoS攻撃	34
1.2.4 情報システムの脆弱性を悪用した攻撃	37
1.2.5 不正アクセスや内部不正による情報漏えい	42
1.2.6 個人を狙う騙しの手口	46
1.2.7 企業・組織を狙う騙しの手口	51
第2章 サイバー空間を巡る動向	62
2.1 サイバー安全保障関連施策の状況	62
2.1.1 能動的サイバー防御の実現に向けた取り組み	62
2.1.2 経済安全保障関連施策の状況	66
2.2 AIの安全性確保の取り組み	69
2.2.1 AIの発展と社会への普及	69
2.2.2 AIのリスク	70
2.2.3 AIの安全性とガバナンス	76
2.2.4 AIのセキュリティ	79
2.3 偽・誤情報対策の動向	88
2.3.1 偽・誤情報の定義	88
2.3.2 偽・誤情報の最新情勢	89
2.3.3 2025年度の注目事例	91
2.3.4 偽・誤情報への対策の動向	93
2.3.5 状況のまとめと今後の見通し	95

第3章 国内の政策及び取り組みの動向	106
3.1 国内のサイバーセキュリティ政策の状況	106
3.1.1 政府の政策動向	106
3.1.2 経済産業省の政策	114
3.2 サイバーセキュリティ人材の現状と育成	122
3.2.1 サイバーセキュリティ人材を取り巻く現状と課題	122
3.2.2 サイバーセキュリティ人材育成のための国家試験、国家資格制度	125
3.2.3 セキュリティ人材育成のための活動	126
3.3 製品・サービスの評価・認証制度、及び暗号技術の動向	133
3.3.1 セキュリティ要件適合評価及びラベリング制度(JC-STAR)	133
3.3.2 ITセキュリティ評価及び認証制度(JISEC)	138
3.3.3 政府情報システムのためのセキュリティ評価制度(ISMAP)	139
3.3.4 暗号技術の動向	141
3.4 中小企業のサイバーセキュリティ対策強化の取り組み	145
3.4.1 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(SCS評価制度)	145
3.4.2 中小企業のサイバーセキュリティ対策の課題とそれらに対する施策	147
3.4.3 中小企業向け支援策	151
第4章 国際的な政策及び取り組みの動向	160
4.1 国際的なサイバーセキュリティ政策の状況	160
4.1.1 国際社会と連携した日本の取り組み	160
4.1.2 米国の政策	165
4.1.3 欧州の政策	172
4.1.4 中国の政策	180
4.1.5 アジア太平洋地域でのCSIRTの動向	183
4.2 国際標準化活動	188
4.2.1 様々な標準化団体の活動	188
4.2.2 情報セキュリティ、サイバーセキュリティ、プライバシー保護関係の規格の標準化 (ISO/IEC JTC 1/SC 27)	189

付録・索引	199
第21回IPA「ひろげよう情報セキュリティコンクール」2025受賞作品	200
索引	201

コラム

ランサムウェア被害から学ぶ、経営者に求められる実践的な対処・対応とは？	18
IoT機器は輸入禁止になるくらい危険なのでしょうか？	54
AI時代のデジタル・シティズンシップと責任	87
「AIの利用をめぐるサイバーリスク」が「情報セキュリティ10大脅威」に初登場	97
データを守ることは、事業を守り強化すること	131
サプライチェーンのセキュリティ管理のための五つのポイント	154
セキュリティを「脅威」ではなく「我々」を主語に語り直す	193



情報セキュリティ白書

●序章 2025年度の情報セキュリティの概況

●第1章 国内外のサイバー脅威の動向

- 1.1 2025年度に観測されたインシデント状況
- 1.2 インシデント事例や脆弱性・攻撃の動向と対策

●第2章 サイバー空間を巡る動向

- 2.1 サイバー安全保障関連施策の状況
- 2.2 AIの安全性確保の取り組み
- 2.3 偽・誤情報対策の動向

●第3章 国内の政策及び取り組みの動向

- 3.1 国内のサイバーセキュリティ政策の状況
- 3.2 サイバーセキュリティ人材の現状と育成
- 3.3 製品・サービスの評価・認証制度、及び暗号技術の動向
- 3.4 中小企業のサイバーセキュリティ対策強化の取り組み

●第4章 国際的な政策及び取り組みの動向

- 4.1 国際的なサイバーセキュリティ政策の状況
- 4.2 国際標準化活動

序章

2025年度の情報セキュリティの概況

2025年度の情報セキュリティを巡る状況は、生成AIや量子コンピューター関連技術の進展、社会全体のデジタル化の加速を背景として、大きな変化の中にあった。これらの技術は、業務の高度化や効率化、新たな価値の創出をもたらす一方で、攻撃の巧妙化・効率化、偽情報の拡散、将来的な暗号基盤への懸念等、新たなリスクも生み出している。技術の進化がもたらす恩恵と脅威の双方を見据え、情報セキュリティを経営、事業継続、社会基盤、安全保障にまたがる課題としてとらえる必要性は、これまで以上に高まった。

国内外ではランサム攻撃、不正アクセス、DDoS攻撃等が継続して発生し、機密性、完全性、可用性のいずれに対しても深刻な影響を及ぼした。

2025年5月には、外食事業者のテイクアウトサイトに対する不正アクセスにより、クレジットカード情報等2,270件が漏えいした可能性があることが公表された。同年6月には研究開発支援事業者、同年7月にはホスティング事業者に対するDDoS攻撃が確認され、サービス提供の継続そのものを脅かす事案が続いた。2026年1月には農林水産省でのメール誤送信、同年2月には総務省が実施したアンケート調査における委託先の設定ミスにより個人情報漏えいが発生しており、従来型の漏えい事案も引き続き発生している。

2025年度は、企業活動の中核を担う基幹システムが攻撃を受けた場合の影響の大きさも改めて浮き彫りとなった。2025年8月には海外自動車メーカーが認証情報の悪用をきっかけにランサム攻撃を受け、生産、販売、物流に影響が及び、長期間の操業停止を招いた。国内では、同年9月に大手飲料メーカーがランサム攻撃を受け、約2ヵ月間のシステム停止が発生した。同年10月には大手eコマース企業に対するランサム攻撃により、約74万件の情報漏えいが発生した。こうした事案は、業務のデジタル化が進展する中で、一部のシステムの停止による影響が、受発注、物流、顧客対応等サプライチェーンの広範囲に及び得ることを示している。

国際的には、国家を背景とするサイバー攻撃が引き続き警戒された。2025年8月にはAPT攻撃グループ「Salt Typhoon」によるサイバー攻撃に関する国際アドバイザリーが共同署名の形で公表された。サイバー攻撃は、

機密情報の窃取にとどまらず、通信基盤や重要サービスへの影響を通じて、国家安全保障や経済安全保障にも関わる問題となっている。こうした情勢のもと、国際的な連携と情報共有の重要性は一層高まっている。

新技術を巡っては、AIの利活用拡大に伴う課題も顕在化した。2025年5月には、米国の新聞でAIのハルシネーションによる存在しない書籍の書評の掲載が発生した。AIは生産性向上や高度な分析を可能にする一方で、誤情報の生成・拡散による社会的信頼の毀損といったリスクも抱えている。サイバーセキュリティ上の課題は、システムやデータの保護にとどまらず、人の認知活動や社会的信頼の維持にも及ぶものとなっている。

こうした状況を受け、国内では防御側の体制整備も進展した。2025年5月には、サイバー対処能力強化法及び同整備法が公布された。同年7月には国家サイバー統括室(NCO)が発足し、サイバーセキュリティ戦略本部が改組された。更に12月には「サイバーセキュリティ戦略」が閣議決定され、平時からの探知、分析、対処を重視する方向へ、我が国の体制整備は一段と進んだ。2026年3月には、経済産業省が「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に関する制度構築方針」(SCS評価制度の構築方針)を公表した。SCS評価制度が活用されることにより、サプライチェーン全体でのサイバーセキュリティ対策が強化されることが期待される。

制度面では、新技術への対応も着実に進められた。2025年7月には欧州で「General-Purpose AI Code of Practice (汎用AIの行動規範)」が、米国では「America's AI Action Plan」が公表された。日本でも同年6月に「人工知能関連技術の研究開発及び活用の推進に関する法律」(AI法)が公布・一部施行され、同年9月には全面施行された。同年12月には同法に基づき、「人工知能基本計画」が閣議決定され、政府の具体的な方針が示された。また、同年10月にはIoTラベリング制度の推進を図る国際的な枠組みとして「グローバル・サイバーセキュリティ・ラベリング・イニシアティブ(GCLI)」が発足した。同年11月には内閣官房が「政府機関等における耐量子計算機暗号(PQC)への移行について(中間とりまとめ)」を公表し、2035年を目途にPQCへの移行を行うことが示された。

2025年度の主なインシデント・事件・政策・イベント

序章

2025年度の情報セキュリティの概況

	● 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
2025年 4月	● 証券会社をかたるフィッシングメールにより甚大な被害が発生(1.1.2)	● 「情報流通プラットフォーム対処法」の施行(2.3.4) ● IPA「企業組織向けサイバーセキュリティ相談窓口」開設(1.2.7、3.1.2)
5月	● 外食事業者のテイクアウトサイトへの不正アクセスにより、クレジットカード情報等 2,270 件の漏えいが発生した可能性(1.2.5) ● 米国の新聞において、AI のハルシネーションによる存在しない書籍の書評の掲載が発生(2.2.2)	● 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（サイバー対処能力強化法）、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（サイバー対処能力強化法整備法）の公布(2.1.1)
6月	● 研究開発支援事業者への DDoS 攻撃を確認(1.2.3)	● 第 10 回日米サイバー対話の開催(4.1.1)
7月	● ホスティング事業者が DDoS 攻撃を確認(1.2.3)	● 国家サイバー統括室（NCO）の発足と、サイバーセキュリティ戦略本部の改組(2.1.1、3.1.1) ● 欧州「General-Purpose AI Code of Practice（汎用 AI の行動規範）」を公表(2.2.3) ● 米国「America's AI Action Plan」を公表(2.2.3)
8月	● 英国自動車メーカーへの認証情報悪用による不正アクセスとランサム攻撃により、生産活動、販売、物流業務への影響及び長期間の操業停止が発生(1.2.2)	● 「Salt Typhoon」によるサイバー攻撃に関する国際アドバイザリーの共同署名を行い、パブリック・アトリビューションとして、本件アドバイザリーを公表(1.2.2、4.1.1)
9月	● 大手総合飲料メーカーへのランサム攻撃により、約 2 ヶ月間のシステム停止が発生(1.2.1)	● 「人工知能関連技術の研究開発及び活用の推進に関する法律」（AI 法）の全面施行(2.2.3)
10月	● 大手 e コマース企業へのランサム攻撃により、約 74 万件の情報漏えいが発生(1.2.1)	● 「グローバル・サイバーセキュリティ・ラベリング・イニシアティブ(GCLI)」の発足(3.3.1)
11月	● 国立国会図書館のシステム更新で委託先が管理する開発環境への不正アクセスを確認(1.2.5)	● 内閣官房「政府機関等における耐量子計算機暗号（PQC）への移行について（中間とりまとめ）」を公表(3.3.4)
12月		● 「人工知能基本計画」の閣議決定(2.2.3) ● 「サイバーセキュリティ戦略」の閣議決定(3.1.1)
2026年 1月	● 農林水産省でのメール誤送信により、マイナンバー等を含む個人情報漏えいが発生(1.2.5)	● 経済産業省「経済安全保障経営ガイドライン」を公表(2.1.2)
2月	● 総務省が実施したアンケート調査での委託先の設定ミスにより、回答者の個人情報漏えいが発生(1.2.5)	● 「International AI Safety Report 2026」の公開(4.1.1)
3月		● 「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に関する制度構築方針」の公表(3.1.2、3.4.1)

※表には、2025 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

サイバー空間を巡る動向

本章では最初に、政府が注力している能動的サイバー防御と、経済安全保障に関する取り組みを取り上げる。2025年度はAI技術の進展と社会への普及が進んだ一方、AIのリスクが顕在化し、攻撃者によるAIの悪用が進んだ年であった。また、偽・誤情報の拡散が

引き続き国際的な公共への脅威であったが、官民の偽・誤情報への対策が進化した年でもあった。そこで本章では、2024年度に引き続き、AIと偽・誤情報の動向を取り上げる。

2.1 サイバー安全保障関連施策の状況

本節では、政府が特に注力しているサイバー安全保障関連施策について取り上げる。具体的には、能動的サイバー防御に関する取り組みと、経済安全保障に関する取り組みを紹介する。

2.1.1 能動的サイバー防御の実現に向けた取り組み

政府は、2022年に閣議決定された「国家安全保障戦略^{*1}」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を掲げた。具体的には、能動的サイバー防御の導入のため、①官民連携の強化、②通信情報の利用、③攻撃者のサーバー等への侵入・無害化、④内閣サイバーセキュリティセンター（NISC：National center of Incident readiness and Strategy for Cybersecurity）の発展的改組及びサイバー安全保障分野の政策を一元的に総合調整する新たな組織の設置等の実現に向け検討を進めるとされた。本項では、これらの検討を実施した「サイバー安全保障分野での対応能力の向上に向けた有識者会議^{*2}」により取りまとめられた「サイバー安全保障分野での対応能力の向上に向けた提言^{*3}」、及び同提言を踏まえて制定された「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^{*4}」（以下、サイバー対処能力強化法）、「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^{*5}」（以下、整備法）の概要、及び「重要電子計算機に対する特定不正行為による被害の防止のための基本的な方針^{*6}」について紹介する。

(1) サイバー安全保障分野での対応能力の向上に向けた提言

「国家安全保障戦略」に基づき新たな取り組みの実現に必要な法制度の整備等の検討のため、「サイバー安全保障分野での対応能力の向上に向けた有識者会議」が立ち上げられ、2024年6月から11月にかけて4回の全体会議、及び①官民連携の強化、②通信情報の利用、③アクセス・無害化をテーマとした計9回のテーマ別会合が開催された。

その結果、実現すべき具体的な方向性について、上記①～③に、及びそれぞれのテーマについて議論を行う中で明らかになった④横断的課題を加えて、「サイバー安全保障分野での対応能力の向上に向けた提言」が2024年11月に取りまとめられた。

同提言に基づき、後述のサイバー対処能力強化法及び整備法が制定された。また、サイバーセキュリティ戦略本部は、2025年5月、同提言等を踏まえ、「サイバー空間を巡る脅威に対応するため喫緊に取り組むべき事項^{*7}」を決定し、現行制度下において喫緊に取り組むべき施策の方向性を取りまとめるとともに、改組後のサイバーセキュリティ戦略本部において2025年内を目途に新たな「サイバーセキュリティ戦略」を策定することとした。同決定で示された施策の方向性の概要は以下のとおりである。

- サイバーセキュリティに係る新たな司令塔機能の確立
- 巧妙化・高度化するサイバー攻撃に対する官民の対策・連携強化
- サイバーセキュリティを支える人的・技術的基盤の強化
- 緊密な国際連携を通じた我が国のプレゼンス強化

新たな司令塔機能の確立のため、2025年7月1日には、NISCが発展的に改組され、サイバー安全保障も含め、官民を通じたサイバーセキュリティの確保に関する司令塔として、内閣サイバー官を長とする国家サイバー統括室(NCO: National Cybersecurity Office)が発足した^{*8}。

各テーマの概要及び同提言の概要については、「情報セキュリティ白書 2025^{*9}」の「3.1.1 (1) (a) サイバー安全保障分野での対応能力の向上に向けた提言」を参照されたい。

2025年12月に閣議決定された新たな「サイバーセキュリティ戦略^{*10}」については、本白書の「3.1.1 (1) 『サイバーセキュリティ戦略』に基づく取り組み」を参照されたい。

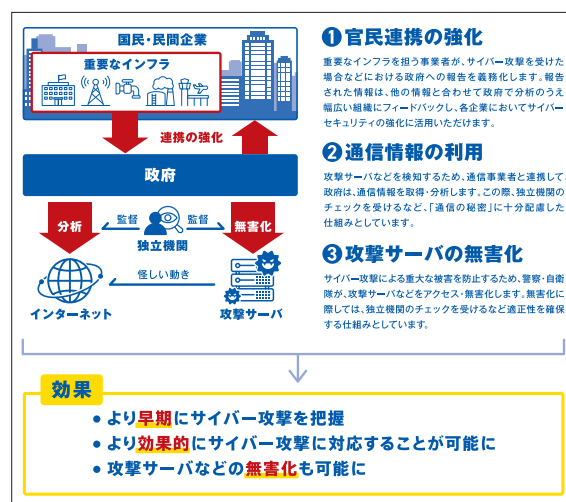
(2) サイバー対処能力強化法及び整備法

「国家安全保障戦略」及び前述の提言を踏まえて2025年2月、サイバー対処能力強化法案及び整備法案が閣議決定され、同年5月に成立し^{*11}、同月、サイバー対処能力強化法及び整備法として公布された^{*12}。

両法律の制定の背景としては、上述のとおり「国家安全保障戦略」において能動的サイバー防御の導入に向けた検討が閣議決定されたことに加え、IT系システムの侵害、有事に備えた重要インフラ等への侵入、機微情報の窃取等のサイバー安全保障に関わる攻撃が巧妙化・深刻化するとともに、サイバー攻撃関連通信や被害数が年々増加傾向にあり、質・量両面でサイバー攻撃の脅威が増大していること、欧米主要国が官民連携や通信情報の利用、アクセス・無害化について先行した取り組みを行っていることが挙げられる^{*12}。そこで、国民生活や経済活動の基盤及び国家及び国民の安全をサイバー攻撃から守るべく、欧米主要国における取り組みも参考に、能動的サイバー防御を実施する体制を整備するため、両法律が制定された。

サイバー対処能力強化法は①官民連携の強化、②通信情報の利用、③分析情報・脆弱性情報の提供等を規定し、整備法は④アクセス・無害化措置、⑤組織・体制整備等を規定（関係法律の改正）している。図2-1-1は、両法律により実現する能動的サイバー防御のポイントを示す図である。

サイバー対処能力強化法及び整備法における①～⑤に係る規定の概要は以下のとおりである。内容が広範にわたるため、詳細については内閣官房の説明資料等^{*14}及び「情報セキュリティ白書 2025」の「3.1.1 (1) (b) サイバー対処能力強化法及び整備法」を参照されたい。



■ 図 2-1-1 能動的サイバー防御のポイント
(出典)内閣官房「みんなで備えよう。新・サイバー防御、はじまる。」^{*13}

① 官民連携の強化

- ・ 特別社会基盤事業者^{*15}によるインシデント報告等
- ・ 情報共有・対策のための協議会の設置
- ・ 電子計算機の使用者に対する情報共有
- ・ 脆弱性対応の強化等
- ・ 罰則の整備

② 通信情報の利用

- ・ 特別社会基盤事業者との協定(同意)に基づく通信情報の取得
- ・ (同意によらない)通信情報の取得
- ・ 自動的な方法による機械的情報の選別の実施
- ・ 独立機関による事前審査・継続的検査
- ・ 罰則の整備

③ 分析情報・脆弱性情報の提供等

④ アクセス・無害化措置

- ・ 警察によるアクセス・無害化措置(警察官職務執行法の改正)
- ・ 防衛省・自衛隊によるアクセス・無害化措置(自衛隊法の改正)

⑤ 組織・体制整備等(関係法律の改正)

- ・ 政府を挙げた取り組みを推進するための体制整備

サイバー対処能力強化法及び整備法は、一部を除き、公布の日(2025年5月23日)から起算して1年6ヵ月以内(サイバー対処能力強化法のうち、サイバー通信情報監理委員会は2026年4月に設置済、特定重要電子計算機の届出義務開始は公布日から1年6ヵ月以内、通信情報の利用については2年6ヵ月以内、整備法のうち、サイバーセキュリティ戦略本部の改組及び内閣サイバー

官の設置は2025年7月施行済)で、政令で定める日から施行される。

(3) 重要電子計算機に対する特定不正行為による被害の防止のための基本的な方針

政府は、サイバー対処能力強化法に基づき、重要電子計算機に対する特定不正行為による被害防止のための基本的な方針案や法施行その他必要な事項について意見を聴くため、「サイバー対処能力強化法の施行等に関する有識者会議」を設置し、2025年9月から12月にかけて4回の会合を開催した^{*16}。その検討結果を踏まえ、サイバー対処能力強化法第3条に基づき2025年12月に「重要電子計算機に対する特定不正行為による被害の防止のための基本的な方針」が閣議決定された。同方針は、法に基づく各般の施策を実施することにより、情報の収集、情報の整理・分析、及び情報の提供を抜本的に強化するものであり、以下の章で構成されている。

- ①重要電子計算機に対する特定不正行為による被害の防止に関する基本的な事項
- ②当事者協定の締結に関する基本的な事項
- ③通信情報保有機関における通信情報の取り扱いに関する基本的な事項
- ④情報の整理及び分析に関する基本的な事項
- ⑤総合整理分析情報の提供に関する基本的な事項
- ⑥協議会の組織に関する基本的な事項
- ⑦その他重要電子計算機に対する特定不正行為による被害の防止に関し必要な事項

同方針の概要を以下に示す。

- ①重要電子計算機に対する特定不正行為による被害の防止に関する基本的な事項
 - ・法に基づく各般の施策実施により、サイバーセキュリティの対策に資する情報の収集、収集した情報の整理・分析、及び整理・分析を踏まえて作成した情報の提供の各機能を抜本的に強化する。
 - ・NCOによる総合調整のもと、政府内の緊密な連携協力を進めるとともに、特別社会基盤事業者はもとより、中小企業も含めて広く様々な事業者との官民連携、同盟国・同志国等の関係機関・団体との連携を強化する。
 - ・法の適用にあたっては、法目的を達成するために必要な最小限度において厳格にその権限を行使する。関連業務に携わる通信情報保有機関のすべ

ての関係職員は、通信の秘密を尊重しつつ厳格にその業務に取り組むことを徹底する。

- ・重要電子計算機や特定重要電子計算機の定義を明確化する。特に特定重要電子計算機は、事業者等との協議を経て、特別社会基盤事業者の業態別に明確化する。
- ・IPアドレス等通信履歴に係る情報等の機械的情報は、分析に必要な情報を適切に含めるとともに、意思疎通の本質的な内容が理解できないものに限定する。

②当事者協定の締結に関する基本的な事項

- ・当事者協定の制度は、特別社会基盤事業者等におけるサイバーセキュリティの確保のための官民相互の協力の推進に資する枠組みであることから、内閣府は、本制度が有効に活用されるよう、特別社会基盤事業者等との間で丁寧な協議を行いつつ、当事者協定の締結を進める。
- ・重要性の高い特別社会基盤事業者等から締結に向けた協議を求める。
- ・締結が事実上の強制とならないよう配慮し、締結しなかった者に対して不当に不利益な取り扱いをしない。
- ・分析の対象となる選別後通信情報を例外的に「他目的利用」する際は、内閣府は、協定当事者から個別に具体的かつ明確な同意を得て、その範囲内で選別後通信情報を他目的利用する。なお、本法に基づく他目的利用には犯罪捜査のための選別後通信情報の利用は含まれ得ない。

③通信情報保有機関^{*17}における通信情報の取り扱いに関する基本的な事項

- ・通信情報保有機関は、通信の秘密等に十分に配慮して制度の適正な運用を図る。
- ・内閣府及び関係行政機関は、必要なシステム・設備の整備を進め、運用に係る職員の確保・育成を図る。
- ・電気通信事業者の協力による負担が過度にならないよう配慮し回避策を検討する。
- ・通信情報保有機関では、透明性の向上、安全管理措置の遵守、提供用選別後情報の活用、個人情報保護法等の法令遵守に取り組む。また、関係職員は通信情報の利用に際し、法に規定する規律について厳格に遵守する。
- ・サイバー通信情報監理委員会を設置し、必要な規模の体制を構築し、法律に基づき所掌事務を実施

する。また、適切に国会への報告を行い、内容の充実に努める。

④情報の整理及び分析に関する基本的な事項

- 内閣府は、法の規定に基づき収集した情報等について、重要電子計算機に対する特定不正行為による被害の防止に有効に活用されるよう、総合的かつ業種横断的に整理及び分析を実施する。
- 特定重要電子計算機の届出は脆弱性情報の提供に活用するため、個別事業者向けの専用設計品等は不要。
- 特定侵害事象等の報告は、対象範囲を明確化し、タイミングに即して過度な負担とならないよう設定する。また、報告窓口の一元化について調整を進める。
- 総合整理分析情報は、用途・場面、閲読者等に応じて内容等を適切にすること、また政府だからこそ取得や整理・分析が可能な情報に基づく情報、受け手の具体的行動につながる情報の作成に努める。
- 総合整理分析情報を加工して選別後通信情報を含まない提供用の、インフラ事業者等向けに秘密を含まない周知等用の総合整理分析情報を作成する。

⑤総合整理分析情報の提供に関する基本的な事項

- 総合整理分析情報等の提供先と内容は表 2-1-1 のとおり。
- ワンボイスで機関ごとに差異が生じないよう、関係行政機関等の間で緊密に連携を図る。
- 政府は、適切なタイミングで積極的に情報提供するよう努め、情報提供の在り方も不断に改善を図る。

情報提供した事業者等へ積極的にフィードバック等を行い、情報共有がより活発になるよう取り組むとともに、その権利利益保護に配慮する。

- 政府は、通信の当事者その他の者の権利利益の保護に配慮するとともに、選別後通信情報の取り扱いについては、組織的、物理的、技術的な安全管理措置を講ずる。

⑥協議会の組織に関する基本的な事項

- 重要電子計算機に対する特定不正行為による被害防止のため、協議会を組織する。協議会では、被害防止情報の提供、被害防止に資する情報の共有・協議、演習や初動対応支援等の機会提供を行う。
- 柔軟な運営に取り組み、自発的な相互の情報提供・意見交換等を活性化させる。
- 専門家向けの技術情報に限らず、経営層の判断に必要な情報を適切なタイミングで積極的に提供する。
- 構成員との継続的なコミュニケーションを通じてニーズを把握し、それを踏まえた情報提供に取り組む。
- 被害防止の対策や情報を適正に管理するために必要な措置等について、構成員で協議を行う。
- 構成員以外の者にも、秘密を含まない情報の提供を行い、国内のサイバーセキュリティ強化を促す。情報提供した構成員等の権利利益保護に配慮する。

⑦その他重要電子計算機に対する特定不正行為による被害の防止に関し必要な事項

- 施行後3年を目途として、特定侵害事象等の報告、通信情報の取得、当該通信情報の取り扱い等の

(1)行政機関等	内閣府は、行政機関が使用する重要電子計算機の被害発生の可能性を把握した、特別社会基盤事業者の役務提供に支障を及ぼすおそれがある、アクセス・無害化措置に資する等の場合には、行政機関等に対し対策や措置に必要な総合整理分析情報を提供。
(2)外国の政府等	内閣府又は通信情報機関は、①法の規定による提供目的の制限に適合するかを個別かつ適切に判断するとともに、②外国の政府等が法に規定する情報の取扱いに係る適切な措置の実施を明示的に確認した上で、その必要な範囲で総合整理分析情報等を提供。
(3)協議会の構成員	内閣府は、例えば、サイバーの専門家が求める技術情報に限らず、経営層の判断に必要な攻撃の目的や背景等に関する情報などの提供用総合整理分析情報等を適切なタイミングで積極的に提供。構成員のニーズも踏まえた形式・内容での情報提供に取り組んでいく。
(4)特別社会基盤事業者	内閣府から情報提供を受けた特別社会基盤事業者の所管省庁は、特別社会基盤事業者に対して、攻撃技術情報などの周知等用総合整理分析情報を積極的に提供。
(5)電子計算機を使用する者	内閣府は、重要電子計算機の利用者に限らず、特定不正行為に用いられるおそれのある電子計算機の利用者等に対し、周知等用総合整理分析情報を提供。例えば、協議会の構成員に対して提供するような脅威情報に必要な加工を行った上で、情報提供を行う。
(6)電子計算機等供給者	内閣府又は電子計算機等供給者の所管省庁は、必要に応じて、公表前の脆弱性情報を重要電子計算機の供給者に対して迅速に提供。脆弱性情報は、その秘匿性や緊急性も踏まえ、適切な情報提供・情報管理に努めていく。脆弱性関連情報の取扱いについては、政府が本法に基づく官民連携に係る事務を着実かつ効果的に実施できるよう、関係する告示・ガイドラインの必要な見直しを行う。

■表 2-1-1 総合整理分析情報の提供先と提供内容

(出典)内閣府政策統括官(サイバー安全保障担当)「サイバー対処能力強化法に基づく基本方針の概要^{*18)}」

状況を検討し、必要があれば所要の措置を講ずる。

- 政府は、不断に取り組み状況の検証・評価を行い、それに伴う制度の見直しを適時に行う。
- 関係省庁や関係機関は、緊密な連絡・協力が不可欠であり、経済安全保障推進法や個人情報保護法、他関連業法を所管する省庁との連携、合理的な制度設計・運用に努める。
- 内閣府、国の行政機関、IPA、国立研究開発法人情報通信研究機構(NICT: National Institute of Information and Communications Technology) その他関係者は、法その他の法令、基本方針に基づき、相互に連絡・協力する。また、内閣府の事務はNCOの総合調整のもとで実施する。
- 能動的サイバー防御を実現するために必要な措置とされた「アクセス・無害化措置」については、サイバー安全保障担当大臣のもと、NCOが総合調整機能を発揮し、警察及び防衛省・自衛隊が個別のアクセス・無害化措置を執行する。
- 関係行政機関は、平素から必要な連携を図り、アクセス・無害化措置の実施に資すると認められる情報を、情報保全にも配慮しつつ、効果的かつ適正にアクセス・無害化措置の実施に関わる行政機関に提供する。

2.1.2 経済安全保障関連施策の状況

2024年5月に「経済施策を一体的に講ずることによる安全保障の確保の推進に関する法律^{*19}」(以下、経済安全保障推進法)が全面的に施行されるとともに、「重要経済安保情報の保護及び活用に関する法律^{*20}」(以下、重要経済安保情報保護活用法)が公布され、同法の施行に向けた準備が進められた。本項では、両法の概要と2025年度の主な動きを紹介する。

(1) 経済安全保障推進法

経済安全保障推進法に基づき、①重要物資の安定的な供給の確保に関する制度、②基幹インフラ役務の安定的な提供の確保に関する制度、③先端的重要技術の開発支援に関する制度、④特許出願の非公開に関する制度の四つの制度が創設された。

このうち、①に関しては、2026年4月時点で16の特定重要物資について145件の供給確保計画が認定されている^{*21}。

また、②に関して、制度の対象事業(電気、ガス、

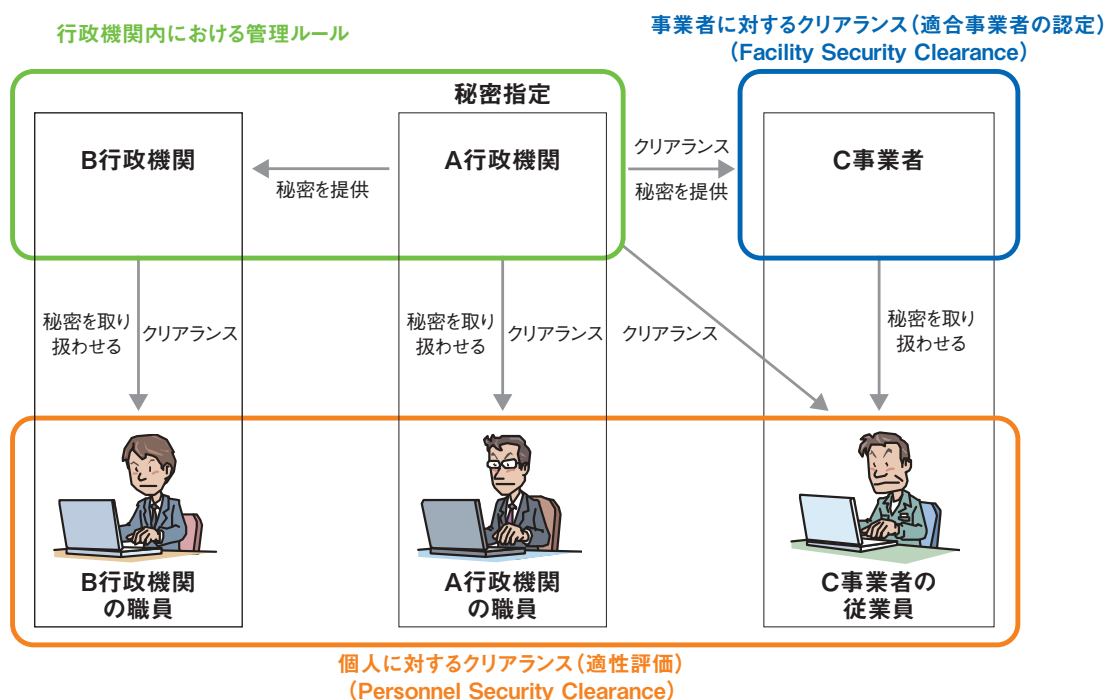
石油、水道、鉄道、貨物自動車運送、外航貨物、航空、空港、電気通信、放送、郵便、金融、クレジットカード)は、港湾運送事業が加わり15分野になった(2025年4月)。更に2026年6月には医療分野を追加する改正法が公布され、公布日から起算して1年6ヵ月を超えない範囲内において政令で定める日から施行される^{*22}。

2025年11月より、経済安全保障推進法の改正に向けて「経済安全保障法制に関する有識者会議^{*23}」が検討を行い、2026年1月に「経済安全保障の更なる推進に向けた提言^{*24-1}」を公表した。この提言を踏まえた「経済施策を一体的に講ずることによる安全保障の確保の推進に関する法律及び株式会社国際協力銀行法の一部を改正する法律案」が2026年3月に閣議決定され、第221回国会(令和8年特別会)に提出され、同年6月に成立・公布された^{*24-2}。同改正法では、①重要な物資の安定的な供給の確保、②基幹インフラ役務の安定的な提供の確保、③先端的重要技術の開発支援、④重要な海外事業の促進、⑤総合的な経済安全保障シンクタンク及び官民協議会の創設という五つの対応を講じるとされた。

(2) 重要経済安保情報保護活用法

安全保障の概念が、防衛や外交という伝統的な領域から経済・技術の分野に大きく拡大し、軍事技術・非軍事技術の境目も曖昧となっている中、国家安全保障のための情報に関する能力の強化は、一層重要になっており、経済安全保障分野においても、厳しい安全保障環境を踏まえた情報漏えいのリスクに万全を期すべく、我が国の情報保全の更なる強化を図る必要がある。そこで、2024年1月に公表された「経済安全保障分野におけるセキュリティ・クリアランス制度等に関する有識者会議」の「最終とりまとめ^{*25}」により、経済安全保障分野におけるセキュリティ・クリアランス制度^{*26}の創設が提言された。セキュリティ・クリアランス制度とは、国家における情報保全措置の一環として、政府が保有する安全保障上重要な情報として指定された情報にアクセスする必要がある者(政府職員及び必要に応じ民間事業者等の従業者)に対して政府による調査を実施し、当該者の信頼性を確認した上でアクセスを認める制度である。これを受けて、2024年5月に重要経済安保情報保護活用法が成立し^{*27}、同月公布され、2025年5月に施行された。同法に基づく重要経済安保情報の提供の流れの概略図を図2-1-2(次ページ)に示す。

同法は、①重要経済安保情報の指定、②重要経済



■図 2-1-2 重要経済安保情報の提供の流れとそれに対する管理ルールのイメージ
(出典)内閣府「適性評価と重要経済安保情報の提供の流れ^{*28}」を基に IPA が編集

安保情報の提供、③重要経済安保情報の取扱者の制限、④適性評価、⑤罰則等を定めている^{*29}。

2025 年 1 月、重要経済安保情報保護活用法第 18 条第 1 項に基づき、「重要経済安保情報の指定及びその解除、適性評価の実施並びに適合事業者の認定に関し、統一的な運用を図るための基準^{*30}」が閣議決定された。同基準は、重要経済安保情報の指定及び解除等、適性評価の実施、適合事業者の認定等について、具体的な判断基準や手順を規定している。また、同基準では、同法の運用にあたって留意すべき事項として、拡張解釈の禁止並びに基本的人権及び報道・取材の自由の尊重、公文書管理法及び情報公開法の適正な運用が規定された。また、適性評価の実施にあたっての基本的考え方として、基本的人権の尊重、プライバシーの保護、「特定秘密の保護に関する法律」で定められた七つの調査事項以外の調査の禁止、適性評価の結果の目的外利用の禁止が規定された。

更に、2025 年 5 月には、各行政機関が定める重要経済安保情報保護規程、重要経済安保情報に係る通報窓口一覧、適性評価に係る苦情受理窓口・相談窓口一覧、行政機関や適合事業者向けのガイドライン、適性評価に関する Q&A が公開された^{*31}。

また、同法の適正な運用を確保するため、同法の施行日より、独立公文書管理監が独立した公正な立場から検証及び監察を行っている^{*32}。検証・監察の結果、

2026 年 3 月、是正の求め等(是正の求め 3 件、うち、重要経済安保情報保護活用法関係は 1 件)を行った^{*33}。

2026 年 6 月には、同法の規定に基づき、重要経済安保情報の指定及びその解除、適性評価の実施並びに適合事業者の認定の状況について国会に報告された^{*34}。

(3) 関連施策

2025 年は、経済安全保障推進法を着実に執行するため、関連する複数の施策が実施された。

(a)「経済安全保障に関する産業・技術基盤強化アクションプラン」再改訂

経済産業省では、経済安全保障に関する産業・技術基盤を強化するための取り組みの方向性と内容を、2023 年 10 月に「経済安全保障に関する産業・技術基盤強化アクションプラン」として取りまとめ、官民での戦略的対話を行ってきた。2024 年 5 月には同アクションプランを改訂したが、その後、我が国を取り巻く経済安全保障環境が大きく変化したことを踏まえ、2025 年 5 月、同アクションプランを再改訂した^{*35}。同アクションプランの再改訂の概要^{*36}を以下に示す。

同アクションプランは、大国が、物資、技術、人材、資金、データ等の「囲い込み」と「自給自足」を通じて、先端技術だけでなく、伝統的な製造業まで対象にした国家戦略を加速している新たな国際環境下での産業・

技術基盤強化に向け、以下の方向性をもって再改訂された。

- 投資、人材、技術を呼び込むための環境整備を拡大し、国内産業・技術基盤を強化
- 「新たな製造業＝（製造業X）」を創出すべく、物資・技術だけでなく、金融、物流、データ等のサービスを含め、バリューチェーン全体を射程にとらえた取り組みを強化
- 同志国間、官民・民間同士の力を有機的につなぐための「プラットフォーム」を作成

また、我が国の自律性・不可欠性確保のため、「3つのP」（産業支援策（Promotion）、産業防衛策（Protection）、国際連携・官民対話（Partnership））を有機的に連携させるべく、以下のように取り組みのフレームワークをアップデートするものとされた。

- 官民連携のもと、新たな国際環境がもたらす経済安全保障上の「脅威・リスク」を分析するとともに、我が国にとつての「機会」を抽出すべく、①我が国が優位性を有する技術の把握、他国の関心も踏まえた技術分析の継続、②物資を超え、サプライチェーン分析から産業バリューチェーン分析に拡大、③シナリオ分析に加え、各国との経済安全保障協力を進めるため、地域情勢分析を追加する等、分析の射程を拡大する。
- 自律性・不可欠性確保の観点から、重要な物資・技術・サービス等を特定した上で、官民連携と同志国や影響力を増すグローバルサウス諸国等との国際連携のもと、バリューチェーンを強化する。

－ 物資・技術アプローチ

「破壊的技術革新が進む領域」「我が国が優位性を持つ領域」「対外依存の領域」の考えのもと、重要な物資、技術、サービス等を特定していく。

－ 産業バリューチェーンアプローチ

研究開発、調達、生産、販売のバリューチェーン全体、各層ごとに対応策を補強する。物資の「サプライチェーン」に加えて、知財、ソフトウェア、更に物流、金融、データ等のサービスにも注目する。

－ 官民対話

官民での脅威リスク認識、対応策に関する対話を深化させる。

－ 地域戦略

我が国の自律性・不可欠性確保に向けた各国・地域との相互依存関係・補完関係を構築するため、例えば、日本のシーレーン確保のための要衝国・

地域との連携強化や経済安全保障の観点からの戦略的地域枠組み等を検討する。

(b)「経済安全保障経営ガイドライン」取りまとめ

企業が経済安全保障対応を進めていく中で、経済安全保障上のリスクに起因する損失を中長期的に抑え、企業価値の維持・向上も見据えた経営戦略を考える上の推奨事項として、2026年1月、経済産業省が「経済安全保障経営ガイドライン（第1版）^{*37}」を取りまとめた。

本ガイドラインでは、経営層が以下の三つの原則を認識し、主体的に対策を進めることが重要としている。

- 自社ビジネスを正確に把握し、リスクシナリオを策定する。
- 経済安全保障への対応を単なるコストではなく、投資ととらえる。
- マルチステークホルダーとの対話を欠かさない。

また、企業の経済安全保障への対応は多岐にわたるが、本ガイドラインでは、「自律性確保の取組」「不可欠性確保の取組」、及び「経済安全保障対応におけるガバナンス強化」の3点について、個社の取り組みとして推奨される事項を記載している。

(c)「経済安全保障グローバルフォーラム・ウィークス」開催

経済安全保障分野における国内外の連携を強化するため、経済産業省では、2025年9月末から12月にかけて、政府・独立行政法人・民間シンクタンク等が協調して経済安全保障に関する国際イベントを開催する「経済安全保障グローバルフォーラム・ウィークス^{*38}」を実施した。具体的には、会議・展示会、国際フォーラム、セミナー、シンポジウム等が多数開催された。

(d)「技術流出対策ガイダンス」策定

産業界が技術流出対策を講じる際の一助となることを目的に、2025年5月、経済産業省が「技術流出対策ガイダンス第1版」を公表した。その後、2026年1月に公表された「経済安全保障経営ガイドライン」と足並みを揃え、「各章で共通する技術流出対策」「人を通じた技術流出への対策」等について内容を大幅に充実させるとともに、企業における国内外との「共同研究」や調達時の「すり合わせ」に伴う技術流出対策に関するニーズを踏まえ、これらの内容の記述を新設した「技術流出対策ガイダンス第2版^{*39}」を2026年4月に策定した。

2.2 AIの安全性確保の取り組み

本白書が執筆された約10年前を振り返ると、囲碁AI（Artificial Intelligence：人工知能）のAlphaGoがプロ棋士に相次いで勝利し、AIの可能性を世界に印象付けた^{*40}。2022年には、OpenAI, Inc.（以下、OpenAI社）がChatGPTを公表し、専門知識を持たない一般利用者でも、自然言語による対話を通じて生成AIを容易に利用できるようになった^{*41}。現在では、多くの人がAIによる社会の変化を実感しているところだろう。

AIの利活用が広がる一方、IPAによる「情報セキュリティ10大脅威2026^{*42}」に「AIの利用をめぐるサイバーリスク」が初めて選ばれた。急速に発展し社会に普及するAIのリスクが、現実の脅威として認知され始めていると言えよう。本節ではAIの発展とリスク、安全性確保の取り組み等について、2025年度以降の動向や事例を中心に、2026年5月上旬時点までに公表された情報に基づいて概観する。

2.2.1 AIの発展と社会への普及

AIの技術は、古くは1950年代の研究から発展してきたと言える。その歴史的な概観については「情報セキュリティ白書2025」の「2.1.1 (1) 20世紀中のAI技術」や「2.1.1 (2) 機械学習とディープラーニングで発展したAI」を参照いただきたい。ここでは、直近の生成AIの発展やAIエージェントの実用化に着目する。

(1) 生成AIの発展

最初にAIの基礎に関して簡潔に触れると、AIの動作は、学習（訓練）と推論に分けられる。学習とは、大量のデータから傾向や関係性を見だし、未知の入力に対しても適切に応答できるようにする過程である。推論とは、実際の入力に対して、学習した情報に基づき具体的な出力を導く過程である。こうした学習と推論を担う仕組みはAIモデルと呼ばれ、学習によって得られた情報をパラメーターとして内部に保持している。

生成AI（Generative AI）とは、「文章、画像、プログラム等を生成できるAIモデルにもとづくAIの総称^{*43}」であり、自然な文章を生成するLLM（Large Language Model：大規模言語モデル）はその代表例である。LLMの学習では、AIモデルのパラメーター数、学習に用いるデータ量、そして学習に使用される計算量を増や

す程、応答の精度が向上するスケールリング則^{*44}が知られている。この経験則に沿ってLLMは規模を拡大しながら発展してきた。ただし、その基本的な仕組みは、学習した膨大なテキストに基づいて、与えられた文章に続く語句を統計的に予測するものであり、「確率的オウム」とすら呼ばれた^{*45}。特に、計算問題等、論理的思考を要する質問は苦手とされてきた。

こうした課題に対応するため、複雑な問題を分解し、段階的に推論を行う推論モデル（Reasoning Model）の開発が進んでいる^{*46}。特に、最終的な回答までの中間的な推論のステップを明示的に扱うCoT（Chain of Thought：思考の連鎖）は、論理的な推論を要する質問への回答精度を高める手法として知られている^{*47}。推論モデルの開発には、2024年9月に発表されたOpenAI ol-previewを始め、各社が参入している^{*48}。2026年4月には、OpenAI社、Google LLC（以下、Google社）、Anthropic PBC（以下、Anthropic社）の各AIモデルが東京大学の入試問題における大半の科類で人間の合格最高点を上回り、一部のAIモデルは数学で満点を記録したと報告された^{*49}。このように推論モデルの登場以降、LLMが苦手とされていた推論能力にも大幅な向上が認められる。

生成AIの発展は、LLMによる対話だけでなく、画像や動画の生成等でも進んでいる。例えば、2025年8月にGoogle社が発表したNano Banana^{*50}は、単に画像を生成するだけでなく、対話形式で物体の消去等細かい修正を可能にした。動画に関しても、2025年後半に登場したOpenAI社のSora 2^{*51}やGoogle社のVeo 3.1^{*52}は、利用者が入力した写真を参考に、人物の動画を生成できるようになった。

更に、テキストや画像、動画等異なる種類のデータを統合的に扱うマルチモーダルAIも普及してきている。例えば、Google社は音声によるAIとの対話に加え、スマートフォン等の画面共有やカメラを使った動画による対話機能の提供を2025年4月に開始した^{*53}。また、Microsoft Corporation（以下、Microsoft社）はWindowsのデスクトップ画面をAIと共有して対話する機能を2025年7月から提供しており^{*54}、マルチモーダルAIの発展が人間にとってより自然でコミュニケーションに使いやすいAIを実現しつつあると言えよう。

(2) AI エージェントの実用化

自然な回答やコンテンツ生成を行う AI は、単に質問に答えるだけでなく、人間の作業を代行する能力も高めている。AI エージェントと呼ばれるこうした AI は、LLM の高い推論能力を基盤として、自律的に計画を立案し、外部のシステムを操作しながら、与えられた目標に向かって一連の処理を実行する点を特徴としている^{*55}。

2025 年以降、こうした AI エージェントの特徴を持つサービスが相次いで実用化されている。例えば、OpenAI 社の ChatGPT エージェントは、Web サイトの操作機能を備え、情報収集や分析からフォーム入力、予約等を一貫して実行できる AI エージェントとして 2025 年 7 月に登場した^{*56}。また、Microsoft 社が提供する Microsoft 365 Copilot は、組織内のデータや業務システムと連携し、問い合わせ対応やデータ分析等の業務を支援する^{*57}。Anthropic 社も、ユーザーの指示に沿って多様なデスクトップ業務を遂行する Claude Cowork を提供している^{*58}。更にソフトウェア開発の分野では、GitHub, Inc. が提供する GitHub Copilot^{*59} や Anthropic 社の Claude Code^{*60} 等、コード生成だけでなく、テストや修正を含む開発作業を自律的に実行する AI エージェントが登場している。このように AI エージェン

トは、日常生活から専門的な業務に至るまで幅広い領域で活用され、作業の効率化や高度化を推し進めている。

2.2.2 AI のリスク

AI の発展は幅広い恩恵をもたらすようになった一方で、特有のリスクも指摘されている。総務省と経済産業省が取りまとめた「AI 事業者ガイドライン^{*61}」では、事業者が AI によるリスクの対策を検討できるよう、その体系的な分類案を示している(表 2-2-1)。

本項では、主に 2025 年度以降の事例や報告を、表 2-2-1 の中分類のリスク例に対応付けて紹介する。なお、以降で紹介するリスクは、表 2-2-1 のリスク例のすべてではないため、記載のないリスク例の説明については、「AI 事業者ガイドライン」を参照いただきたい。

(1) 学習及び入力段階のリスク

汚染されたデータで AI を学習させて誤動作を引き起こすデータポイズニング攻撃や、悪用可能な応答を AI から引き出すプロンプトインジェクション攻撃等、AI 特有の攻撃が知られている。これらの事例は、「2.2.4(1)(a) AI の技術的特徴を狙った攻撃」で後述する。

大分類	中分類	リスク例
技術的リスク (=主に AI システム特有のもの)	学習及び入力段階のリスク	データ汚染攻撃等の AI システムへの攻撃
	出力段階のリスク	バイアスのある出力、一貫性のない出力等 ハルシネーション等による誤った出力
	事後対応段階のリスク	ブラックボックス化、判断に関する説明の不足
社会的リスク (=既存のリスクが AI においても発生又は AI によって増幅するもの)	倫理・法に関するリスク	個人情報の不適切な取扱い等
		生命等に関わる事故の発生
		差別的出力
		過度な依存
		悪用
	経済活動に関するリスク	知的財産権等の侵害
		金銭的損失
		機密情報の流出
		労働者の失業
		データや利益の集中
	情報空間に関するリスク	資格等の侵害
		偽・誤情報等の流通・拡散
		民主主義への悪影響
		フィルターバブル及びエコーチェンバー現象
	環境に関するリスク	多様性・包摂性の喪失
		バイアス等の再生成
	環境に関するリスク	エネルギー使用量及び環境の負荷

■表 2-2-1 AI によるリスク例の体系的な分類案

(出典)総務省、経済産業省「AI 事業者ガイドライン(第 1.2 版)別添(付属資料)^{*62}」

(2) 出力段階のリスク

AI の出力には誤りが含まれる可能性があり、事実と異なることをもって回答する「ハルシネーション」のリスクがある。2025 年 12 月 8 日に青森県沖で発生した地震では、気象庁から発表された津波警報や津波注意報が解除される前にもかかわらず AI による Web サイトの要約が、「大津波警報・津波警報・津波注意報はすべて解除されており、発表されていません」と誤った内容を表示した^{*63}。また、航空会社から送信されたメールの内容について、ユーザーの利用する AI が予約と異なる空港名に誤訳したり、オンラインチェックインの内容を誤って欠航と要約したりする事象が確認されている^{*64}。

対話形式で回答を出力する AI とは異なり、AI エージェントでは、与えられた指示に応えるために周囲のシステムを自律的に操作する場合があるため、システムに対して意図しない操作をするリスクがある。ソフトウェアのコーディングを行う AI エージェントが、一連の作業中に特定のフォルダーの削除ではなく誤って最上位のフォルダーから削除してしまい多くのファイルが滅失した事例^{*65}や、指示を無視してデータベースを削除した上に事実と反してロールバックができないと主張した事例^{*66}が報じられている。

(3) 事後対応段階のリスク

AI の出力根拠が不明確な場合、出力が人材の採用、保険給付等の判断に用いられると、判断結果によって不利益を受けた本人が原因を把握しにくくなる。判断根拠の不透明さや説明不足が問題となり、訴訟に至るケースも見受けられる。

米国では、Workday, Inc. が提供する、AI を使った採用支援サービスが、連邦の差別禁止法に違反し、人種、年齢、障害に基づく不利益な影響を求職者に与えたとして、同社に対する訴訟が進行している。2025 年 5 月、連邦地方裁判所は、年齢差別に関する請求について、原告と同様の立場の求職者が訴訟に参加できる暫定的な手続きを認めた^{*67}。医療保険においても、UnitedHealth Group, Inc. に対する訴訟で、原告は AI を用いたツールの判断によって、加入者の治療に必要な保険給付が不当に拒否されたと主張している。2026 年 3 月、連邦地方裁判所は、同社に対して AI を用いた給付判断プロセスに関する広範な文書の提出を命じた^{*68}。

これらの事案は執筆時点でいずれも係争中であり、AI による不当な判断や違法性が認定されたものではないが、AI による判断に対する根拠や説明の不足が訴

訟を含めたリスクにつながり得ることを示している。

(4) 倫理・法に関するリスク

AI の活用が、人の安全やプライバシー、主体的な判断等に深く関わる場面へ広がる中で、適切な管理や配慮を欠いた利用によって、倫理的・法的に保護されるべき利益が損なわれるおそれがある。本項では、AI の悪用、AI における個人情報の不適切な取り扱い、AI を用いたことによる事故、AI への過度な依存について、実際の事例とリスクを述べる。

(a) 悪用

AI の悪用事例は、後述するサイバー攻撃(「2.2.4(2) AI を悪用したサイバー攻撃への対処」参照)や偽・誤情報の生成(「2.3 偽・誤情報対策の動向」参照)に加え、詐欺やわいせつ画像の作成と多岐にわたる。

詐欺に関しては、生成 AI で実在する人物になります事例が相次いでいる。ビデオ会議に参加した偽の CFO (Chief Financial Officer: 最高財務責任者) の送金指示で、ある海外の企業が約 38 億円を騙し取られた事件^{*69}が 2024 年に発生し話題となった^{*70}。日本企業に対しても 2025 年に詐欺グループが AI で合成された社長の音声で部下に電話をかけて送金の指示をした事例^{*71}が報告されている。2026 年 3 月には、他人になりました人物が国内 IT 企業のオンライン面接を受けていたことが報じられた^{*72}。国外においても生成 AI を悪用してオンライン面接で他人になります事例が確認されており、米国連邦捜査局 (FBI: Federal Bureau of Investigation) は北朝鮮の IT 技術者が外貨獲得を目的として身分を偽装していることを指摘している^{*73}。こうした事例の中には、公開された写真や動画だけでなく、SNS 等を通じて対象の人物を偽の会議に誘導し、そこで音声や映像を取得することで、それらを元にした AI によるリアルタイムの音声や動画生成でその対象者になります手口の存在も知られている^{*74}。また、なりすまし以外にも、インド国内から日本人が狙われたサポート詐欺事件において、ポップアップウィンドウの作成や日本語への翻訳等に生成 AI が使われたことが報告されている^{*75}。

詐欺とは別の AI の悪用として、わいせつ画像の作成も後を絶たない。2025 年 4 月、男女 4 人が画像生成 AI で実在しない女性のわいせつな画像を作成し、インターネットオークションで販売した疑いにより、生成 AI によるわいせつ物販売事件として全国で初めて逮捕された^{*76}。また、実在する人物を対象として、特に児童の

学校行事や卒業アルバムの写真を AI で性的画像に加工する事例が相次いでいる^{*77}。警察庁の調査によると、全国の都道府県警察が取り扱った事案数は 2025 年に 114 件となり、被害児童の同級生または同じ学校の生徒が行為者となっている事案が 65 件と過半数を占めた^{*75}。この数値は若年層が容易に AI を悪用でてしまう状況を如実に表していると言えよう。AI-CSAM (Child Sexual Abuse Material: シーサム) と呼ばれる生成 AI を悪用した児童性的虐待コンテンツは、海外においても問題視されており、SNS に投稿した子供の写真が知らないうちに AI で性的画像に加工されてしまう危険性も指摘されている^{*78}。

(b) 個人情報の不適切な取り扱い等

AI の利用に伴い、多様な個人情報が取得・分析される場面が増えている。特に、AI が個人識別やスコアリング、AI モデルの改善等のために、利用者や第三者のデータを収集する場合、本人への通知や同意、利用目的の明確化等が問題となり得る。

米国では、2025 年 12 月、AI 会議アシスタントを提供する Fireflies.ai Corp. に対して、オンライン会議の参加者の生体音声の録音・分析・文字起こし・保存を行い、話者識別のための生体情報を本人の同意なく収集していたとして、訴訟が提起された^{*79}。特に、Fireflies サービスのアカウントを持たず、利用規約にも同意していない会議参加者の音声も処理されたと主張されている。また、2026 年 1 月には採用支援サービスを提供する Eightfold AI Inc. に対し、AI を用いて本人の同意なしに個人情報を収集して応募者のプロフィールやランキングを作成し、雇用主に提供していたとして、応募者らが集団訴訟を提起したと報じられている。訴状によれば、応募書類に記載された以上の情報を SNS 等から収集し、応募者の好みや適性についての推論やランク付けを行って雇用主に提供し、それらが候補者の絞り込みに使われたという^{*80}。なお、いずれの事案も係争中であり、執筆時点で原告の主張が認定されたものではないことに留意が必要である。

(c) 生命等に関わる事故の発生

AI の動作や応答が生命に深刻な損害を与える可能性がある。米国では、Tesla, Inc. のオートパイロットシステムを搭載した自動車が、時速約 100 キロメートルで交差点に突進し、死亡事故を発生させた訴訟に関して、2025 年 8 月、同社に事故の 33% の責任があったとした陪審評決が下された。同社側はオートパイロットではなく

無謀な運転手の責任だと主張したが、2026 年 2 月には、連邦地裁が同社による評決取り消しの申し立てを退け、2 億 4,300 万ドル (388 億 8,000 万円^{*81}) の評決を維持したと報じられている^{*82}。また、対話型 AI による医療の助言が死亡事故につながったとして訴訟に発展した事例もある。米国で 19 歳の大学生が ChatGPT の助言に従った薬物の過剰摂取により死亡したとして、2026 年 5 月に遺族が OpenAI 社等を提訴したと報じられている。訴状によると、ChatGPT が薬物や抗不安薬等を混ぜて飲むようアドバイスし、致死量となる摂取量を提案したという^{*83}。ただし、この事件は執筆時点で係争中であり、原告の主張が認定されたものではない。

(d) 過度な依存

AI は、利用者の相談、調査、文章作成、意思決定等を支援する有用な手段となり得る一方で、利用者が AI との対話や AI の回答に過度に依存することにより、精神的な不安定さが増幅されたり、自ら判断する力や自信が低下したりするリスクも指摘されている。

米国では、2025 年 8 月、16 歳の少年の自殺について、少年が ChatGPT との対話に深く依存し、同サービスが自殺念慮を助長した等として、少年の両親が OpenAI 社等を提訴したと報じられた^{*84}。また、同月には、ChatGPT との対話に没頭していた男性が母親を殺害した後に自殺した事件について、AI との対話が男性の被害妄想を悪化させた可能性があると報じられた^{*85}。更に、2026 年 5 月には、フロリダ州の銃乱射事件の遺族が ChatGPT による犯行助長を主張して OpenAI 社を提訴した事例も報じられている^{*86}。ただし、これらはいずれも執筆時点で訴訟または報道段階のものであり、AI との法的因果関係が確定したものではない点には留意が必要である。

また、過度な依存による悪影響は、利用者の精神的な不安定さの増幅だけではない。2026 年 4 月に米国心理学会が公表した研究では、AI に依存して業務を行うことは、認知能力そのものを低下させるとは限らない一方で、自らの独立した推論能力に対する自信や、成果物に対する主体感を損なう可能性があるとされている^{*87}。このように、AI は利用者の作業や意思決定を支援する一方で、精神的健康や作業への自信・主体感に悪影響を及ぼすおそれがある。

(5) 経済活動に関するリスク

AI の活用に伴い、事業活動を支える情報・資産・

人材等を取り巻く環境が変化し、経済活動の安定性が揺らいだり、効率性がかえって損なわれたりするリスクがある。以下では、実際に発生した事例とリスクについて述べる。

(a) 知的財産権等の侵害

AI 利用に伴う知的財産権上のリスクのうち、特に著作権を巡る問題が顕在化している。以下で取り上げる各事例には、著作権以外の法的論点や複雑な事実関係を含むものもあるが、ここでは著作権上の論点に焦点を当てて紹介する。主な論点としては、以下の「(ア) AI の出力・表示が他者の著作権を侵害するリスク」「(イ) 他者の著作物を AI の学習に利用することを巡るリスク」「(ウ) AI 生成物の著作権保護・権利帰属を巡るリスク」が挙げられる。なお、著作権法の解釈は、個別具体的な事案に応じて司法により行われるべきものであることに留意されたい。

(ア) AI の出力・表示が他者の著作権を侵害するリスク

生成 AI は、利用者の指示に応じて画像、動画、文章、要約等を生成・表示するが、その内容が既存の著作物との関係で著作権侵害に該当するかが問題となる場合がある。

画像生成 AI に関しては、Disney Enterprises, Inc. 等複数のエンターテインメント企業が、AI による画像生成サービスを提供する Midjourney, Inc. を 2025 年 6 月に提訴した^{*88}。この訴訟では、同社が著作権で保護されたキャラクターを無断で学習させ、利用者がこれらに類似する画像を生成できるようにしていると主張されている。

動画生成 AI についても、同様の懸念が示されている。2025 年 10 月には、OpenAI 社が提供を開始した動画生成 AI である Sora 2 を巡り、一般社団法人日本動画協会、公益社団法人日本漫画家協会、出版社 17 社が「生成 AI 時代の創作と権利のあり方に関する共同声明^{*89}」を発表した。同声明では、Sora 2 により生成・公開された動画の中に、著名な創作物や表現への依拠性・類似性が強く疑われるものが散見されることに加え、権利者が明示的にオプトアウトしない限り著作物が生成・公開され得る仕組みへの懸念が示されている。また、「第二、第三の Sora 2」とも言うべき新たな生成 AI の登場が容易に予見されるとしている。その後、OpenAI 社は、Sora 2 の Web 版・アプリ版の提供を 2026 年 4 月 26 日に終了し、API についても同年 9 月 24 日に終了予定であると公表した^{*90}。また、2026 年 3 月には、ByteDance

(字节跳动有限公司) の動画生成 AI である Seedance 2.0 についても、著作権上の問題によりグローバル展開が一時停止したと報道された^{*91}。

新聞記事に関する事例としては、株式会社日本経済新聞社と株式会社朝日新聞社による Perplexity AI, Inc. に対する訴訟がある。両社は、AI を利用した検索サービスにおいて新聞記事が無断で取得・保存・利用され、回答として要約が表示されたことが、著作権侵害等にあたると主張している。この事例は、画像や動画のように視覚的に既存作品と類似する出力だけでなく、記事本文の要約の表示についても、AI の出力・表示が著作権侵害の疑義を招き得ることを示している。なお、Perplexity AI, Inc. は 2026 年 5 月にこの訴訟に関して争う姿勢を示している^{*92}。

一方、AI の出力が既存作品の「作風」の類似にとどまる場合には、直ちに著作権侵害にあたるとは限らない。2025 年 3 月には、OpenAI 社の GPT-4o による画像生成機能の提供をきっかけに、株式会社スタジオジブリの作風を模した画像が SNS 上に多数投稿され、著作権上の問題が議論された^{*93}。一般論として、著作権法は創作的表現を保護するものであり、単に作風が似ているだけでは著作権侵害にあたるとは言えない。他方で、AI の出力について、特定の著作物との類似性及び依拠性が認められる場合には、著作権侵害となり得るとされる^{*94}。

(イ) 他者の著作物を AI の学習に利用することを巡るリスク

AI を巡る著作権問題は、生成されたコンテンツだけでなく、AI の開発・学習段階にも及ぶ。書籍、画像、音楽、記事、映像等他者の著作物を、権利者の許諾なく AI の学習に利用することが著作権侵害にあたるとか、あるいは許容されるのかが問題となる。

米国では、作家らによる Anthropic 社を相手取った書籍の無断学習に関する集団訴訟が注目された。2025 年 6 月、カリフォルニア州連邦地裁は、同訴訟において、適法に取得された書籍の AI の学習への利用についてはフェアユースを認める一方、海賊版サイト等から書籍を取得・保管したことについては別途問題があると判断した。その後、この訴訟では、海賊版由来の書籍利用に関する請求を巡り、同社が和解のために 15 億ドル (2,400 億円) 規模の支払いに合意したと報じられている^{*95}。

この事例は、学習段階における著作物利用について、

単に「AI 学習であれば適法」または「無断学習は常に違法」と単純化できないことを示している。特に、利用目的、利用された著作物の性質、取得経路、保存・複製の態様、権利者の市場への影響等が、個別具体的に問題となる。なお、この訴訟は米国法上のフェアユースに関する判断を含むものであり、法体系や著作権法の規定が異なる日本法における結論と直ちに同一に扱うことはできない。日本においては、AI の学習等のニーズを背景に、2018 年の著作権法改正によって著作物の利用の円滑化を図るための柔軟な権利制限規定が整備された。文化庁は、こうした AI と著作権に関する考え方や、リスク低減や権利保全のためのチェックリストとガイダンスを示している^{*96}。

(ウ) AI 生成物の著作権保護・権利帰属を巡るリスク

AI を用いて生成されたコンテンツ自体に著作権が認められるのか、また、認められる場合に誰が権利を持つのかという問題がある。これは、他者の権利を侵害するリスクとは異なり、自社や利用者が AI で作成した成果物を十分に保護できるのかという観点から重要である。

AI 生成物を巡っては、「AI 自体が著作者となり得るか」という問題がしばしば取り上げられる。米国では、AI システムを単独の著作者として登録できるかが争われた訴訟において、2026 年 3 月、米国連邦最高裁判所が裁量上訴を認めなかった。これにより、人間の著作者性を必要とする米国著作権局及び下級審の判断が維持された^{*97}。

実務上は、人間が道具として AI を利用した場合、どの程度の創作的関与があれば、その成果物が著作物として保護されるのかが重要である。文化庁が示した考え方^{*96}によれば、人間の「創作意図」と「創作的寄与」が認められる場合に、著作物として保護される余地がある。

日本国内では、2025 年 11 月、生成 AI で作成された画像を無断で複製したとして、神奈川県男性が著作権法違反の疑いで摘発された事例が報道された^{*98}。この事例は、AI 生成物の著作物性が問題となり得る事例として注目される一方、執筆時点では捜査段階であり、裁判所が AI 生成物の著作物性について判断を示したものではない点に留意が必要である。

(b) 労働者の失業

AI の発展により、人間の仕事内容や役割が変化し、労働者の負担軽減や生産性向上が期待できる一方で、

雇用の減少につながるリスクも懸念されている。

大手テクノロジー企業を中心に、従業員の解雇や採用停止が相次いで報じられている。一例として、Salesforce, Inc. については、2025 年に 1,000 人以上の人員削減が報じられたほか、同社 CEO は、AI による生産性向上を背景に、同年中はエンジニアを新規採用しない趣旨の発言をしている^{*99}。一方、2025 年 10 月に約 14,000 人の人員削減計画を発表した Amazon, Inc. は、その理由について、AI による代替ではなく、組織階層を減らして機動性を高めるためと説明している^{*100}。このように、企業によって背景は異なり、また、その背景を公にしているかも異なるため、一概に AI が人員削減の原因とは言えない点には留意が必要である。

AI による失業の懸念の対象はテクノロジー企業に限らない。例えば、AI により生成された架空の女優が登場したことに対して、俳優らの間に仕事を奪われる懸念が広がっている。2025 年 9 月には全米映画俳優組合も、AI で生成された女優について、人間の俳優を失業させるとして批判する声明を発表している^{*101}。

(c) 機密情報の流出

AI の学習や推論で扱われた機密情報が、セキュリティやガバナンスの不足によって、意図しない形で流出してしまうリスクがある。特に、攻撃者が機密情報の窃取を目的として、AI を狙った攻撃の事例については「2.2.4 (1) AI に係る安全性確保」の中で扱う。

(d) 低品質な生成物の氾濫

2025 年を象徴する言葉の一つとして、米国の辞書出版社である Merriam-Webster, Inc. は、「slop (スロップ)」を選定した^{*102}。本来 slop は、泥水や残飯等を意味する英単語であるが、同社はこれを「一般に AI によって大量に生成される低品質なデジタルコンテンツ」と定義し、2025 年には、AI によって作成された低品質な動画や書籍、報告書等が広く流通したと指摘した。日本においても、生成 AI で作成した営業メールの大量送信が迷惑行為として報じられている^{*103}。また、セキュリティの分野では、AI が生成した大量の低品質な脆弱性レポートが開発者や保守担当者の負担を増大させていることが指摘されている。これについては、「2.2.4 (3) AI を活用したサイバーセキュリティ確保」で後述する。このように、本来は利用者の作業効率向上が期待される AI であっても、その出力が低品質なまま大量に流通すると、受け手にとっては、かえって情報の選別や確認の負担を

増大させ、作業の妨げになる場合がある。

(6) 情報空間に関するリスク

AIによって情報の生成・流通・受容を巡る環境が変化し、人々の認識や社会の意思形成における偏りや歪みが助長されるリスクがある。以下では、実際に発生した事例とリスクについて述べる。

(a) 偽・誤情報等の流通・拡散

主に、生成 AI による精緻な画像や動画等が、事実と異なるにも関わらず、事実として人々に広まるリスクがあり、世論操作との関連も指摘されている。AI が生成したものに限定しない偽・誤情報の動向については、「2.3 偽・誤情報対策の動向」を参照いただきたい。ここでは、悪意ある生成・拡散ではなく、生成 AI の利用に伴う確認不足により、誤った情報が報道機関等の信頼性のある媒体を通じて流通した事例を紹介する。

2025 年 5 月、米国シカゴで 2 番目の規模を持つ新聞、シカゴ・サンタイムズ紙に存在しない書籍の書評が掲載された。記事は同紙ではなく別の配信会社が配信したもので、記事作成に関与した担当者が書籍リスト作成の段階から生成 AI を使ったことを認めたという。同紙の関係者は、記事を作成した配信会社との契約を見直すとしている。一方で、SNS 上では、なぜ編集者はこれを見落としていたのか、といった同紙への批判が相次いだ^{*104}。この事例は、「2.2.2(2) 出力段階のリスク」で述べたハルシネーションが、そのまま記事として多くの読者に届けられたために、誤情報の拡散という事態の悪化につながってしまった事例と言える。

日本国内では、一般社団法人共同通信社が、2025 年 10 月に配信した記事の写真について、生成 AI により加工されたものと確認したとして取り消した。写真は、屋久島でのタヌキによるウミガメの食害に関する記事に掲載されたもので、もともとは監視カメラで撮影されていた。掲載された写真は、子ガメをくわえるタヌキの画像が鮮明になるよう生成 AI で加工され、子ガメの向き等が変わっていた。同社の社内指針では、報道用の写真に生成 AI を原則使用しないとしているが、取材で提供を受けた写真について確認をしていなかった^{*105}。この事例は、生成 AI による加工が、画像の鮮明化を目的とする場合であっても、被写体の形状や位置関係等、事実認定に関わる要素を変化させれば、報道写真の真正性を損ない得ることを示している。

(b) 民主主義への悪影響

AI を使って個人の政治的意見を誘導することで、選挙干渉等の民主主義への悪影響が生じる可能性が指摘されている。これには、「2.2.2(6) (a) 偽・誤情報等の流通・拡散」や「2.3 偽・誤情報対策の動向」で触れた偽・誤情報による世論操作だけでなく、対話型 AI による政治的な意見の説得も含まれている。2025 年 12 月に学術誌 Science に掲載されたオックスフォード大学の研究者らの論文^{*106}では、4 万人を超える英国人が参加する大規模な実験を行い、最大 10 ターンの LLM との会話の前後で、英国の政治課題に対する意見への同意度をパーセンテージで測定した。LLM は、参加者を決められた意見に説得するよう、事前に指示されていた。LLM との会話の平均時間は約 9 分であったが、条件によっては約 12 ポイントも同意度の変化があり、1 ヶ月後にも説得効果の影響が確認された。この論文は、偽・誤情報に限らず対話型 AI による説得が政治的な意見に影響し得ることを実験的に示しており、AI が多方面から健全な民主主義に影響を及ぼす可能性が懸念される。

(c) 多様性・包摂性の喪失

社会全体で類似の AI が広く利用されるようになると、出力される表現や発想が一定の方向に収束し、多様な視点や、地域ごとの文化や考え方が反映されにくくなり、多様性や包摂性が失われる可能性がある。2025 年 4 月にコーネル大学の研究チームが主要国際会議で発表した論文^{*107}では、西洋中心の AI モデルが文章執筆を支援する提案を行った場合、文章の書き方や内容が西洋の規範に近づき、文化表現の差異を弱めることを実験的に示している。実験では米国とインドからの 118 人の参加者を対象として、決められたテーマに関する文章の執筆を求めた。その際、米国とインドの参加者を、それぞれ執筆中に AI からの文章の提案を受けるグループと受けないグループに分けている。論文では、執筆された文章の分析により、AI を用いることで米国とインドの参加者が書いた文章の間の類似度が高まること等が示されている。

(7) 環境に関するリスク

AI には、学習と推論を支える大規模な計算資源が必要であり、その多くはデータセンターに集約されている。そのため、AI の普及は、データセンターの電力需要、冷却水の使用、関連する CO2 排出の増加等を通じて、環境負荷を高める可能性がある。

IEA（International Energy Agency：国際エネルギー機関）が2026年4月に公表した報告書^{*108}によれば、2025年のデータセンターの電力需要は17%増加し、AI向けデータセンターの電力需要はそれを上回るペースで増加した。AIのタスクあたりの消費電力は低下しているものの、利用の増加等により、データセンター全体の電力消費は2030年までに倍増する見込みとされている。これは、個々の処理の効率化が進んでも、利用量の拡大によって総量としての電力需要が増加し得ることを示している。

CO2排出への影響も課題である。IEAは、データセンターの電力消費に伴うCO2排出量について、道路交通や航空と並んで2030年までに増加する数少ない分野の一つとし、2024年時点で約1億8,000万トンと推計している^{*109}。この規模は、オランダの2023年の燃料燃焼由来CO2排出量である約1億1,200万トン^{*110}の約1.6倍にあたる。ただし、前述の数値はデータセンターの排出量であり、AIが占める排出量はその一部である点には留意が必要である。

また、環境負荷の問題は、電力消費やCO2排出に限らない。AI向け計算基盤における冷却方式によっては水使用量の増加も課題となり得る。Microsoft社は、2025年の環境サステナビリティ報告書において、AIやクラウドサービス需要の拡大を背景に、データセンターの設計・建設・運用の見直しを進めていると説明している。同社は、冷却に水を使用しない新しいデータセンター設計により、施設あたり年間推定1億2,500万リットルの水使用を回避できるとしている^{*111}。

2.2.3 AIの安全性とガバナンス

前項で例示したように、AIにはセキュリティに限らない幅広いリスクが指摘されており、AIの活用にあたっては、こうしたリスクの管理が欠かせない。本項ではAIの安全な活用に向けたAIセーフティやAIガバナンスを概説するとともに、関連する国際的な政策動向にも触れる。

(1) AIセーフティの概要

AIのリスクは以前からも指摘されており、2023年11月には英国でAIセーフティサミットが開催された^{*112}。こうしたAIの安全性に対する国際的な意識の高まりを受けて、2024年2月、日本は英国、米国に次いで世界で3番目にAIセーフティ・インスティテュート（AISI：AI Safety Institute）を設立している。日本AISI（エイシー）

は、総務省と経済産業省が取りまとめた「AI事業者ガイドライン」におけるAIに関連する各主体が取り組む事項^{*113}のうち、「人間中心」「安全性」「公平性」「プライバシー保護」「セキュリティ確保」「透明性」をAIセーフティの重要要素として、これらが保たれた状態をAIセーフティと位置付けている^{*114}。

日本AISIは、この定義に基づいてAIセーフティに関する評価手法の検討等を行っており、これらの重要要素に対する評価観点ガイド^{*115}の作成等を行ってきた。2025年以降はAIセーフティの原則論から具体的な評価に向けた取り組みに重点を移しており、AIセーフティの評価ツールの開発と公開^{*116}や、事業実証ワーキンググループの設立による産業界を交えたAIセーフティ評価の枠組みの整備^{*117}等の活動を行っている^{*118}。更に、日本AISIを含む世界各国のAISI相当機関から構成されるAISI国際ネットワーク（The International Network of AI Safety Institutes）は、名称を「The International Network for Advanced AI Measurement, Evaluation and Science」と改め、2026年2月にはAI評価のベストプラクティスに関する国際的な合意を公表し、各国と連携しながらAI評価に関する活動を進めている^{*119}。

(2) AIガバナンスの概要

AIガバナンスとは、「AIの利活用によって生じるリスクをステークホルダーにとって受容可能な水準で管理しつつ、そこからもたらされる正のインパクト（便益）を最大化することを目的とする、ステークホルダーによる技術的、組織的、及び社会的システムの設計並びに運用^{*43}」である。重要な点は、リスクを完全に排除することでも、リスクを避けるためにAIを使わないことでもなく、AIの便益とのバランスを踏まえてリスクを許容可能な水準で管理することである。この考え方は、一般的なリスクマネジメント^{*120}とも共通している。AIセーフティの評価や対策は、AIガバナンスにおいてリスクを把握・管理する重要な要素と位置付けられる。

こうした内容を含めてAIの開発・提供・利用における必要な取り組みの基本的な考え方を示したのが「AI事業者ガイドライン」であり、我が国におけるAIガバナンスの包括的な手引きとなっている。「AI事業者ガイドライン」は継続的に更新されており、本白書の執筆時点で最新の第1.2版では、AIエージェントを含む技術動向の反映や、AIによるリスクと便益の記載の充実等が行われている^{*121}。また、企業活動におけるAIガバナンスの実務マニュアルを含むガイドとして、日本AISIは、AI

による価値創出と責任ある活用の両立を支援するため、主に民間事業者が Chief AI Officer（CAIO：AI 統括責任者）を設置・運用する際の実務指針となる「Chief AI Officer ガイド^{*122}」を提供している。

既に AI は国境を超えて利用されており、その適切な管理には一国だけでなく国際的な連携が欠かせない。日本は 2023 年に開催された G7 広島サミットの結果を踏まえ、AI に関する国際的なガバナンスの形成に向けた広島 AI プロセスを主導してきた。2025 年 2 月からは、広島 AI プロセスの国際行動規範に関する遵守状況を AI 開発企業等が自ら確認し報告する「報告枠組み」の運用が開始され、報告内容は Web サイトで公開されている^{*123}。広島 AI プロセスは、G7 に限らず取り組みに賛同する国々や組織を増やししながら、国際的な枠組みの形成を進めている^{*124}。

こうした国際的なガバナンスの取り組みと並行して、AI に関わる組織の管理体制や評価プロセスの標準化も進められている。2023 年 12 月に発行された ISO/IEC 42001 は、組織が AI に関するリスクと便益を継続的に管理するための AI マネジメントシステムの国際規格であり、2025 年 8 月にはこれに対応する日本産業規格 JIS Q 42001 が制定された^{*125}。2025 年 7 月には、AI マネジメントシステムの認証機関に対する国際的な要求事項を定めた ISO/IEC 42006 が発行され^{*126}、認証制度の国際的な整備が進められている。また、2025 年 5 月には AI システムの影響評価を定めた ISO/IEC 42005 も発行されている^{*127}。更に、ISO/IEC の合同技術委員会と ISO の上層委員会である CASCO（Conformity Assessment Committee：適合性評価委員会）が共同して、AI システムに対する認証を含む適合性評価のハイレベル・フレームワークに関する ISO/IEC 42007 の開発を進めており^{*128}、AI ガバナンスに関する標準化が活発に進められている。

(3) AI に関する国際的な政策動向

AI の急速な発展と社会実装の拡大を受け、各国・地域では、AI に関する制度整備や戦略策定が進められている。AI は、産業競争力を左右する一方、社会的リスクや安全保障上の課題も伴うため、各国・地域では、それぞれの産業基盤、制度設計、国際競争上の立場を踏まえた政策対応が進められている。以下では、米国、欧州、中国、日本における AI 政策の動向を概観する。なお、サイバーセキュリティの観点からの日本の政策に関しては「3.1 国内のサイバーセキュリティ政策の

状況」を、米国、欧州、中国の政策に関しては「4.1.2 米国の政策」「4.1.3 欧州の政策」「4.1.4 中国の政策」を参照いただきたい。

(a) 米国の政策動向

米国には OpenAI 社や Anthropic 社等の新興 AI 企業や Google 社等の大手テクノロジー企業等、世界を代表する AI 関連企業が集積している。そうした中、2023 年の Joe Biden 政権下での大統領令^{*129}では、AI の安全性評価、政府による監督等、AI の安全性や信頼性が重視されていた。一方、2025 年 1 月に発足した第二次 Donald Trump 政権は、こうした規制色の強いアプローチを撤回し^{*130}、AI 開発を加速させる方向へと政策を転換している。

2025 年 7 月に Trump 政権が公表した「America's AI Action Plan^{*131}」では、「AI 競争に勝利すること」が国家戦略上の重要目標として掲げられており、AI を経済競争力及び国家安全保障上の中核技術と位置付けている。政策の柱としては、AI イノベーションの加速、AI インフラの構築、国際的な主導権の確保が示されている。これらの政策には、中国との技術競争も背景の一つにあり、AI に必要な先端半導体の輸出管理の強化、同盟国や友好国との連携強化も含まれている。

また、AI のセキュリティに関する取り組みも強化されている。2025 年 6 月には、米国 AISI が CAISI（Center for AI Standards and Innovation）に改組されること発表され、AI に関するイノベーションの強化とともに国家安全保障上のセキュリティリスクに対処していく方向性が示された^{*132}。「America's AI Action Plan」では、CAISI との連携を含め、AI 関連の脅威情報を共有する AI-ISAC（Information Sharing and Analysis Center）の構築や AI インシデントレスポンスの強化、更に、AI を利用する軍事及び情報機関のための高度なセキュリティを備えたデータセンターの構築が盛り込まれている。

このように、Trump 政権での米国の AI 政策は、AI の包括的な安全性規制よりも、AI 分野での競争力や国家安全保障を優先する方向へと軸足を移しており、その中でセキュリティも重要な政策課題となっている。

(b) 欧州の政策動向

欧州連合（EU：European Union）では、2024 年 5 月に成立し、同年 8 月に発効した「EU AI Act^{*133}」の段階的な適用が進んでいる。「EU AI Act」は、基本的人権や安全性を重視した包括的な AI 規制であり、

EU 域内での AI の用途を四つのリスクレベルに分類したリスクベースの規制に加え、幅広い用途に利用可能な汎用 AI に対する義務が定められている。

2025 年 2 月には、最も高いリスク区分である「許容できないリスク」に該当する AI の禁止規定の適用が開始された。ここには、サブリミナル技術を使用する AI や個人の犯罪リスクを予測する AI 等が含まれる^{*134}。同年 7 月には、汎用 AI 向けの「General-Purpose AI Code of Practice (汎用 AI の行動規範)」が公表された。この規範は、「透明性」「著作権」「安全性・セキュリティ」から構成され、汎用 AI の提供者を対象としている。規範自体に強制力はないが、準拠することで「EU AI Act」の関連する義務遵守の証明となり、他の方法による証明と比べて対応負担の軽減と法的確実性の向上を享受できるとされている^{*135}。

一方、EU 域内では AI を始めとする各種規制による企業負担の増加やイノベーション停滞の懸念も高まっている。そのため、欧州委員会(European Commission)は「Digital Omnibus」において、AI、サイバーセキュリティ、データに関するルールの簡素化を提案している(「4.1.3 (3) デジタルオムニバス規則案」参照)。また、高リスク AI に関する規定の適用については、標準化された技術仕様やガイダンスの整備のため、最長 16 ヶ月延期する方針が発表されている^{*136}。

また、2020 年に EU を離脱した英国については、「EU AI Act」のような包括的な AI 規制法は導入されていない。2025 年 1 月に Keir Starmer 政権は「AI Opportunities Action Plan」を発表し、イノベーション重視の姿勢を明確にすると同時に、適切な規制と効果的な保証が信頼性のある AI の実現につながり、規制当局もイノベーションを支援する役割を担うとした^{*137}。同年 2 月には、英国 AI Safety Institute を AI Security Institute と改称し、国家安全保障や犯罪利用等セキュリティリスクへの対応を重視する方向性を示している^{*138}。

このように、欧州では、基本的人権や安全性を重視した AI 規制とイノベーション促進の両立を図る方向で政策の調整が進められている。

(c) 中国の政策動向

中国政府は、AI を国家戦略上の重要技術と位置付け、研究開発、産業育成、社会実装を強力に推進している。2017 年に公表された「次世代人工知能発展計画^{*139}」では、2030 年までに世界主要 AI イノベーションセンターとなる目標が掲げられた。その後も、米国との

技術競争や輸出規制を背景に、AI 及び半導体分野での「自立・自強(self-reliance)」を重要政策として位置付け、先端 AI 技術への投資や人材育成、計算資源・データセンター等のインフラ整備が進められている^{*140}。2025 年 8 月には、主要産業に AI を融合させる「AI プラス」の行動指針が発表され、経済成長の柱となる AI のインフラ強化等が改めて示された^{*141}。こうした環境の中、Alibaba Group Holding Limited (阿里巴巴集团控股有限公司)、Huawei Technologies Co. Ltd. (華為技術有限公司) 等の大手テクノロジー企業に加え、DeepSeek 社(杭州深度求索人工智能基础技术研究有限公司)等の新興 AI 企業も含めた AI 技術開発が加速している。

一方で、中国政府は、社会安定、国家安全保障の観点から規制や管理も強化している。2023 年に施行された「生成 AI サービス管理暫定弁法」では、生成 AI サービス提供者に対して、国家権力や社会主義体制の転覆を扇動する内容を含めた違法・有害情報の生成防止、学習データや AI モデルの管理等を求めている^{*142}。また、2025 年 9 月から「AI 生成合成コンテンツラベル弁法」が施行され、AI 生成コンテンツへの識別表示(ラベル付け)が義務化された^{*143}。これにより、生成コンテンツの識別可能性や偽情報対策が一層強化されている。

このように、中国の AI 政策は、国家主導による AI 産業育成と技術自立を推進する一方で、社会秩序の維持や国家安全保障を重視した管理・規制を並行して強化する点に特徴がある。また、米中間の技術競争の激化を背景として、AI 分野における安全保障と産業競争力の双方を重視する方向性が一層鮮明となっている。

(d) 日本の政策動向

日本では、AI を経済成長や安全保障の基盤となる重要技術と位置付け、イノベーション促進と安全性確保の両立を図る政策が進められている。2025 年 9 月には「人工知能関連技術の研究開発及び活用の推進に関する法律」(以下、AI 法)が全面施行された。内閣府は、AI 技術の発展が国民生活の向上及び国民経済の発展に寄与する一方、国内の AI 開発・活用の遅れや AI リスクに対する国民の不安があるとしており、AI 法は、AI のイノベーションを促進しつつリスクに対応するための制度的基盤と位置付けられている。同法では、「人工知能関連技術の研究開発及び活用の推進に関する基本理念」「国・地方公共団体・研究開発機関・事業者等の責務」「人工知能基本計画の策定」「人工知能戦

略本部の設置」等が定められている^{*144}。

AI法に基づき、2025年12月には「人工知能基本計画^{*145}」が閣議決定された。同計画では、信頼できるAIの追求と世界で最もAIを開発・活用しやすい国の実現を基本構想とし、AI利活用の加速的推進、AI開発力の戦略的強化、AIガバナンスの主導、AI社会に向けた継続的変革を柱として、AIの利活用と技術革新の好循環を実現することが掲げられている。

こうした方針のもと、政府全体として、生成AIを含むAIの研究開発と利活用を幅広い分野で推進している。「人工知能基本計画」では、行政、医療・介護、防災、教育等の社会課題解決分野におけるAI利活用の加速、国内のAI開発力の戦略的強化、データの集積・利活用・共有、AI人材の育成・確保等が盛り込まれている。また、政府自らが先導的にAIを利活用する方針も示されている。具体的な取り組みの例として、デジタル庁では政府職員が安全・安心にAIを活用できる基盤としてガバメントAI「源内」を構築しており、2027年度の本格利用に向けて、2026年度には全府省庁の約18万人の政府職員を対象とした大規模実証が進められている^{*146}。

「人工知能基本計画」では、AIガバナンスに関して、「2.2.3(2) AIガバナンスの概要」で述べた「AI事業者ガイドライン」や日本AISIによるAIセーフティに関する取り組みを踏まえ、これらを更に強化する方向性が示されている。特に、日本AISIを抜本的に強化し、AIモデルの技術的評価、AIがもたらすリスクの実態把握、必要な措置の実施を進める方針が掲げられている。また、AIの安全性確保に加え、AIを利用した攻撃への対応が新たなサイバーセキュリティ上の課題であることも踏まえ、体制整備を含めた措置を講じる方針とされている^{*147}（サイバーセキュリティ対策については「2.2.4(4) AI性能の高度化を踏まえた我が国におけるセキュリティ対策の強化」参照）。

このように日本のAI政策は、AIの研究開発・利活用を促進しつつ、リスク対応や安全性確保を進める構成となっている。また、前項で述べた広島AIプロセスやAISI国際ネットワーク等の国際的な枠組みとも連動しながら、AIガバナンスやAI評価に関する国際協調を重視している。

(e) 各国の政策動向のまとめ

以上のように、各国・地域では、それぞれの産業基盤や制度設計、安全保障環境を踏まえながら、AIの

開発・利活用の促進とリスク対応を進めている。政策の方向性には違いがあるものの、AIを経済競争力や国家安全保障に関わる重要技術として位置付け、AIの安全性や信頼性、セキュリティを確保する必要性が高まっている点は共通している。次項では、こうした背景を踏まえ、AIのセキュリティに関する主要な観点に沿って、具体的な事例を見ていく。

2.2.4 AIのセキュリティ

2025年12月に閣議決定したサイバーセキュリティ戦略では、AI技術の進展と普及がサイバーセキュリティ分野においても大きなインパクトを与えているとして、その内容を「AIに係る安全性確保」「AIを悪用したサイバー攻撃への対処」「AIを活用したサイバーセキュリティ確保」の三つの観点で整理している^{*10}。本項では、これらの観点から直近の攻撃事例を示すとともに、具体的な対策に関する情報についても紹介する。

(1) AIに係る安全性確保

Security for AIと呼ばれる領域で、AIを含むシステムを標的とした攻撃と対処を扱う。トレンドマイクロ株式会社の調査によれば、AIに関連する脆弱性は、CVE (Common Vulnerabilities and Exposures: 共通脆弱性識別子)の件数において2025年には2,130件となった。これは前年比34.6%の増加で、CVE全体の増加率17.9%のほぼ倍となっている^{*148}。本項では、急増するAIの脆弱性に対する攻撃を「(a) AIの技術的特徴を狙った攻撃」、AIを対象としているが「(b) AIの技術的特徴によらない攻撃」に大別して紹介し、攻撃への対策については「2.2.4(5) AIに関するセキュリティ対策の考え方」で触れる。

(a) AIの技術的特徴を狙った攻撃

AIの技術的な仕組みに対する攻撃や防御は、敵対的機械学習 (Adversarial Machine Learning) と呼ばれる学術分野として研究が進められてきた。ここでは、実際の攻撃に使われている主な手法を事例とともに紹介する。

(ア) プロンプトインジェクション

AIを狙った特徴的な攻撃として、プロンプトインジェクションは数多くの事例が知られている。プロンプトインジェクションとは、細工した指示をLLMに与えることにより、

不正な動作をさせる攻撃である。なお、安全機構を回避する攻撃の目的に着目したジェイルブレイクという表現もよく用いられる^{*149}。細工した指示を与える経路としては、LLMとの対話において直接入力する方法と、メールの本文や組織内の共有ファイル等 LLM が回答の際に参照する情報源に指示を埋め込んでおく方法とがある。後者は間接プロンプトインジェクションと呼ばれ、特に事例が目立っている。

間接プロンプトインジェクションの事例では2025年6月、情報の窃取が可能となる Microsoft 365 Copilot の脆弱性「EchoLeak」が報告された (CVE-2025-32711、同年5月に修正済み)^{*150}。EchoLeak の脆弱性を悪用するために攻撃者が実施するのは、標的のユーザーにメールを送信するだけである。メールには不正な指示が含まれ、ユーザーが Copilot に質問した際に参照される可能性が高まる細工もなされている。Copilot がメールを参照すると、不正な指示に従って機密情報が特定の URL のパラメーターに埋め込まれ、その URL が Web ブラウザー上で画像として表示されることで攻撃者のサーバーに送信される。間接プロンプトインジェクションにおいて細工した指示を埋め込む対象はメールに限らず、マルウェアの実行ファイルに生成 AI による検知を回避するための指示が含まれていた事例^{*151}や、学术论文の査読で AI による評価が高くなるように、本文に見た目には分からない文字で指示が記載されていた事例^{*152}もある。

間接プロンプトインジェクションは、AI エージェントに対しても通用する。例えば、ユーザーに代わって Web サイト上の操作を行う機能を持つ AI エージェント型のブラウザ Comet に対して Web ページ上の見えない文字で細工した指示を与えることで、ユーザーが利用する Web サイトの認証情報が外部に送信されてしまう脆弱性が報告されている^{*153}。また、業務向けの AI エージェントである Claude Cowork が、Word 等の文書ファイルに埋め込まれた指示によって、ユーザーの持つローカルのファイルを攻撃者に送信してしまう脆弱性も報告されている^{*154}。

(イ)コードインジェクション

実行可能なコードを何らかの形で埋め込み、アプリケーションに実行させる攻撃は、従来からコードインジェクションと呼ばれている^{*155}。AI においても、入力に基づいて推論を行う AI モデルのファイルに不正なコードが埋め込まれ、推論時に実行される可能性がある。こうした攻撃は、AI モデルの導入でよく見られる形態である、

Hugging Face 等のサイトで公開されている学習済みの AI モデルをダウンロードして利用するという形態を主な標的としている。攻撃者は、不正なコードを埋め込んだ AI モデルを公開し、被害者がそれをダウンロードして実行することで攻撃を成立させる。このように、被害者の操作に依存した受動的な攻撃ではあるが、任意のコードが実行される危険性を伴うため、攻撃が成立した際の影響は極めて大きいと言える。

AI におけるコードインジェクションの主な事例として、Python の標準ライブラリにおいてデータ保存に使われる pickle 形式に関するものが挙げられる。AI モデルを格納するデータ形式としても広く用いられる pickle は、実行可能なコードも格納できる仕様となっており、Python の公式ドキュメントにおいても、その危険性が警告されている^{*156}。2025年10月には、AI モデルの構築に用いられるフレームワーク Keras において、セーフモードを有効にしているにもかかわらず、細工された pickle 形式の AI モデルを読み込むことで任意のコードが実行される脆弱性が報告された^{*157}。こうした攻撃への対策として、pickle に不正なコードがないかを検出する Picklescan と呼ばれるオープンソースソフトウェア (OSS: Open Source Software) が公開されているが、この検出を回避して任意のコードが実行される脆弱性も報告されている^{*158}。また、不正なコードが埋め込まれる問題に対応するため、AI モデルに必要な情報のみに絞って格納する新たなデータ形式として、safetensors 等が開発されている。ただし、これらの新たなデータ形式についても、同様の攻撃を可能とする脆弱性が複数報告されており^{*159}、AI モデルのファイルは単なるデータではなく実行可能なコードとなり得ることに留意が必要である。

(ウ)モデル抽出

自然な対話や高度な推論を実現する近年の高性能な AI モデルは、大量の学習データと計算資源を用いて開発されており、開発企業のノウハウとコストが注ぎ込まれた、資産価値の高いものと言える。モデル抽出とは、こうした高価値な AI モデルを標的とし、入力と出力 (更には思考過程等の中間出力) を収集して分析することで、元の AI モデルの能力を盗み取り、比較的少ない労力で類似する能力のモデルを開発する攻撃である。

モデル抽出は敵対的蒸留 (adversarial distillation)、あるいは単に蒸留とも呼ばれる。類似の用語として知識蒸留があり、こちらも蒸留と略されるが、知識蒸留は教師となる AI モデルの入出力を利用して、生徒となる比

較的小規模な AI モデルを効率的に開発する正当な学習技術である^{*160}。一方で敵対的蒸留は、攻撃者が正当な権限なく行う知識蒸留を指し、各 AI 提供企業は利用規約でこれを禁じている^{*161}。

先進的な AI を開発する Google 社や Anthropic 社は、自社に対して大規模な敵対的蒸留が行われていること、またそれらの攻撃の検知や対策に取り組んでいることを公表している^{*162}。これらの攻撃については、直接的な被害を受ける AI 開発企業だけでなく、国家間の AI 開発競争の観点から米国政府も懸念を示しており、米国科学技術政策局（Office of Science and Technology Policy）は、中国が米国の AI モデルを敵対的蒸留によって盗用していると指摘し、民間企業と連携した対策を計画している^{*163}。一方、ワシントンの中国大使館は、この指摘に対し、米国による中国企業の不当な抑圧であるとする立場を示しており^{*164}、モデル抽出は国家安全保障の文脈において今後も注目を集めていくと考えられる。

（エ）スロップスクワッティング

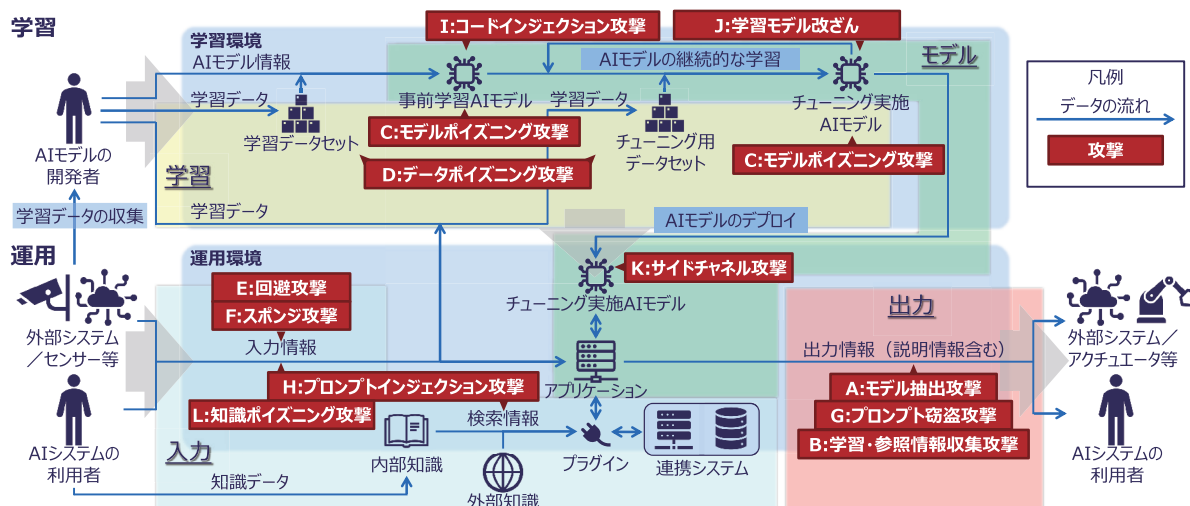
厳密には AI が含まれるシステムそのものへの攻撃ではないが、AI が生成したコードに起因する特徴的な攻撃として、スロップスクワッティング（slopsquatting）がある。AI が生成するコードには、ハルシネーションによって実在しないパッケージ名が含まれる場合がある。この攻撃では、攻撃者がその実在しないパッケージ名で不正なコードを公開しておく。AI の生成したコードが実行されると、もともとは実在しなかったパッケージであるにもかかわらず、攻撃者が用意した不正なコードが読み込まれて実行され、攻撃が成立する。スロップスクワッティングという

用語は、誤入力されやすい URL に悪性の Web サイトを仕込むタイポスクワッティング（typosquatting）になぞらえたものであり、AI による低品質な生成物（slop）に居座る（squatting）という意味を組み合わせた表現である^{*165}。

一見すると遠回りにも思われる攻撃手法だが、AI のハルシネーションは完全にはランダムでなく一定の傾向が見受けられ、特定の実在しないパッケージ名が複数回繰り返し生成されることが実験で確認されている^{*166}。更に、2025 年 10 月に報告された PhantomRaven と呼ばれる攻撃キャンペーンでは、スロップスクワッティングの実例として、AI が生成しやすいパッケージ名の不正なコードが実際に複数公開されていたことが確認されている^{*167}。

（オ）AI 特有のその他の攻撃

AI を含むシステムを標的とした特徴的な攻撃には、これまで紹介した以外にも、学習データや AI モデルの改ざんにより主に推論の誤動作を引き起こすデータポイズニングやモデルポイズニング、推論時に細工した入力を与えることで計算資源を浪費させるスポンジ攻撃等、多種多様な攻撃が知られている。日本 AISI では、これらの攻撃を、学習、モデル、（推論時の）入力、出力の 4 種類の攻撃対象領域（アタックサーフェス）を持つ、12 種類の攻撃手法に大別して俯瞰する資料として「AI システムに対する既知の攻撃と影響^{*168}」を提供している（図 2-2-1）。更に、同資料ではそれぞれの攻撃手法がシステムに与える影響を機密性、完全性、可用性の観点で整理しており、セキュリティ上のリスクアセスメントの参考にも利用できる。具体的な内容については、資料を参照いただきたい。



■ 図 2-2-1 AI システムに対する攻撃の俯瞰
 （出典）日本 AISI「AI システムに対する既知の攻撃と影響^{*169}」

(b) AI の技術的特徴によらない攻撃

AI を標的としたサイバー攻撃の中には、AI そのものの技術的特徴を悪用するのではなく、ソフトウェアに対する従来型の攻撃手法を用いた事例も多く見受けられる。特に LLM や AI エージェント向けのツールが急速に普及しており、これらが新たな攻撃対象となっている。

例えば、LLM アプリケーションの構築で使われる OSS の LangChain では、データ変換の不備によって機密情報が漏えいする脆弱性 (CVE-2025-68664) が報告されている^{*170}。また、AI と外部ツールを接続する MCP (Model Context Protocol) サーバーにおいても、サンドボックス機構が回避される脆弱性 (CVE-2025-53109 等) が確認されている^{*171}。更に、複数の AI エージェントを連携させる CrewAI の脆弱性 (CVE-2026-2275 等)^{*172} や、同じく AI エージェント向けツールである OpenClaw における脆弱性 (CVE-2026-25253 等)^{*173} も報告されている。

また、AI 関連のソフトウェアを狙ったサプライチェーン攻撃も発生している。異なる LLM を統一の API で利用可能にする OSS である LiteLLM では、2026 年 3 月に攻撃者によって開発基盤が侵害され、認証情報の窃取機能等を含む不正なバージョンが公開された^{*174}。

このように、AI の発展によって関連するソフトウェアの開発と普及も急速に進んでおり、攻撃者は AI そのものだけでなく、こうした周辺のソフトウェアについても影響力の大きい攻撃対象として狙っていると考えられる。

(2) AI を悪用したサイバー攻撃への対処

ここまで紹介した AI を狙ったサイバー攻撃だけでなく、AI を悪用した攻撃の事例も相次いで報告されている。こうした攻撃では、標的に AI を含むか否かによらず、攻撃側が効率化や高度化のために AI を利用する。英国 AISI の評価によれば、2026 年 4 月以降に登場した Anthropic 社や OpenAI 社の高度な AI モデルは、従来の AI モデルと比較してサイバー攻撃能力が大幅に向上しているとされる^{*175}。ここでは、AI を悪用したサイバー攻撃の事例を「(a) 参入障壁の低下」「(b) 攻撃の自動化」「(c) 攻撃能力の進化」の三つの軸で紹介し、攻撃への対処については、「2.2.4 (5) AI に関するセキュリティ対策の考え方」で触れる。詐欺等サイバー攻撃によらない AI の悪用に関しては、「2.2.2 (4) (a) 悪用」に記載した。

(a) 参入障壁の低下

英国国家サイバーセキュリティセンター (NCSC: National Cyber Security Centre) は、AI の悪用によ

る影響の一つとして、プログラミング等サイバー攻撃で使われるスキルを AI が補うことで、技術力の高くない者による攻撃の参入障壁を下げ、攻撃の担い手が広がっていると指摘している^{*176}。

2025 年 8 月の Proofpoint, Inc. による報告では、自然言語での指示で Web サイトを簡単に作成できる AI を活用したサービスがフィッシングサイトの作成に悪用され、サイバー犯罪への参入障壁がかつてなく低くなっているとの指摘がなされた。このサービスは、複数の攻撃キャンペーンで認証情報等を狙った Web サイトの作成に悪用され、同サービスは不正利用を検知する仕組みを導入したという^{*177}。

通常、AI はセーフガードやガードレールと呼ばれる安全機構によって、悪用等の不適切な指示を拒否するように設計されており、例えば、AI モデルに善悪の価値観を学習させるアラインメントや、AI モデルの外側のシステムで不適切な入出力を検知して対処する仕組みが施されている。2025 年 11 月の Palo Alto Networks, Inc. による報告では、こうした安全機構が意図的に外され、攻撃のために設計された LLM の事例を紹介し、簡単にマルウェアの作成等ができてしまうと指摘している^{*178}。

AI の悪用事例は国内でも発生しており、2025 年 12 月には、ガードレールを回避するように対話形式の生成 AI に質問することで自作のプログラムを改良し、サイバー攻撃を行ったとして、高校生が逮捕されている^{*179}。

(b) 攻撃の自動化

サイバー攻撃における AI の悪用は、専門知識のない者に対する参入障壁の低下と同時に、熟練した攻撃者に対しても、偵察、侵入、情報抽出といった攻撃の各段階の自動化を実現し、攻撃の大規模化や効率化をもたらした。

2025 年 11 月の Anthropic 社の報告によれば、中国を拠点とするグループと評価される脅威アクターが同社の AI を悪用し、金融機関や政府機関等約 30 組織を標的にした攻撃を実行したという。攻撃者は AI エージェントを活用して一連の攻撃段階を自律的に実行させて攻撃の 80 ～ 90% を自動化し、人間の介入は偵察から積極的な攻撃への移行の承認等、意思決定をする程度であったとされる^{*180}。同社の報告は、AI による攻撃の自動化が驚く程早く進んでいるとされる内容に加えて、それを検証できる情報の不足や誇張を疑う声もあり話題となった^{*181}。それ以降にも、複数の組織から AI の悪用による攻撃の自動化に関する報告が続いている。2026

年2月にはAIで作成されたツールによって600台以上のFortiGateデバイスが侵害された事例^{*182}や、2026年3月には北朝鮮の脅威アクターによるAIを使った自動的な偵察や金銭要求等の悪用が報告された^{*183}。

(c) 攻撃能力の進化

AIの悪用は、自動化によるサイバー攻撃の量的な側面だけでなく、質的な側面への影響も指摘されている。すなわち、AIの悪用によって従来と同様の方法では対処しきれない攻撃が登場する脅威である。

代表例として挙げられるのが、未知の脆弱性の発見能力である。2026年4月のAnthropic社の報告によると、同社のClaude Mythos Preview（以下、Mythos Preview）は、数千件以上の重大な脆弱性を発見し、その中にはセキュリティを重視したOSであるOpenBSDをクラッシュさせる、27年間未発見だったバグも含まれていた^{*184}。従来、脆弱性の発見には高度な知識と長期間の解析が必要だったが、AIによってコードを分析することで、人間が見落としていた重大な問題を短時間で発見できる可能性が示されている。こうした能力が、攻撃者によるシステムの脆弱性の発見と、その脆弱性を狙った攻撃につながる懸念から、本白書の執筆時点で同社はMythos Previewを一般公開していない。代わりに、大手IT企業等一部の組織にのみアクセスを認め、攻撃者に悪用される前に脆弱性を発見して修正することで、システムのセキュリティ向上を目指すProject Glasswingを立ち上げた^{*185}。同じく4月に公表されたOpenAI社のGPT-5.5も高度なサイバーセキュリティ能力を持つとされ、その機能の利用は一定の基準を満たしたユーザーに限定されている^{*186}。

また、未知の脆弱性の発見以外にも、AIの悪用による新たな攻撃の脅威が報告されている。2025年には、感染した環境内においてAIを用いて自律的にコードを生成するマルウェアが複数報告されており、これらはあらかじめ固定された悪性コードを持たないことで、従来型のシグネチャーによるマルウェアの検知を回避する可能性が指摘されている^{*187}。2025年7月には、AIモデル自体の悪用ではないが、よく知られたAIのWebサービスのAPIを攻撃者のサーバーとの通信に悪用し、正規のAIサービス利用に紛れて外部通信を行うマルウェアも発見されている^{*188}。

このように、AIはサイバー攻撃を単に効率化するだけでなく、攻撃能力そのものの進化や攻撃手法の多様化をもたらしつつある。今後は、AIの悪用を前提としたセ

キュリティ対策の重要性が更に高まっていくと考えられる。

(3) AIを活用したサイバーセキュリティ確保

AI for Securityと呼ばれる領域で、これまで述べた攻撃側ではなく、防御側にAIを取り入れる話題を扱う。攻撃と防御は表裏一体であり、例えば、「2.2.4 (2) (c) 攻撃能力の進化」でも触れたようにAnthropic社は高度なサイバーセキュリティ能力を持つMythos Previewを悪用の懸念から執筆時点では一般公開せず、アクセスを限定した上でソフトウェアのセキュリティ向上に活用する取り組みを始めている。2026年4月には、WebブラウザFirefoxの更新において、Mythos Previewが発見した271件の脆弱性を修正しており、開発元のMozillaは「The zero-days are numbered」（ゼロデイ攻撃の終わりは近い）と表現して防御側が優位になり得るとの認識を示している^{*189}。

更に、AIの活用は脆弱性の発見や修正だけでなく、攻撃経路の予測や悪用されやすい脆弱性の分析等、リスクマネジメントの支援に加え、アラートのトリージや詳細調査、攻撃手法の解説等セキュリティ運用を支援する用途にも広がっている^{*190}。従来、セキュリティ運用では大量のアラート分析やログ調査を人手に依存する場面が多かったが、AIが分析や判断を支援することで、運用負荷の軽減や対応速度の向上が期待されている。

しかし、防御側へのAIの導入がかえって問題を引き起こすことも指摘されている。例えば、AIが大量の脆弱性を短期間で発見することで、修正や検証の負荷が急増する可能性がある。2026年1月には、OSSで広く利用されるcurlのバグ報奨金プログラムが終了した。開発者によると、AIを用いた低品質な脆弱性レポートが大量に届き、セキュリティチームに大きな負担をかけているためとされる^{*191}。また、同様の理由で他の報奨金プログラムの新規受付も終了することが報じられている^{*192}。2026年4月には、米国国立標準技術研究所（NIST：National Institute of Standards and Technology）が、脆弱性報告の増加を受けて、NVD（National Vulnerability Database：国家脆弱性データベース）の運用を変更した。従来はすべてのCVEを分析し、重大度のスコアや製品リスト等の詳細を追加していたが、4月15日以降は一定の基準を満たす重要なCVEを優先して扱っている^{*193}。

このように、AIを悪用したサイバー攻撃が急速に拡大する中で、防御側においてもAIの活用が前提となりつつある。一方で、単にAIを導入するだけで直ちに

安全性が向上するわけではなく、防御側の新たな負担や運用上の課題につながる側面もあることから、AIの特性を見極めながら適切に活用していくことが重要になると言える。

(4) AI性能の高度化を踏まえた我が国におけるセキュリティ対策の強化

ここまで、AIを標的としたサイバー攻撃やAIを悪用したサイバー攻撃の事例を紹介するとともに、セキュリティ分野におけるAIの活用にも触れた。本項では、こうした背景を踏まえた我が国におけるセキュリティ対策について、政府が進める取り組みを紹介する。

政府は2026年5月18日、「AI性能の高度化を踏まえたサイバーセキュリティ対策に関する関係省庁会議」において、「Project YATA-Shield^{*194}」と題して、AI性能の高度化を踏まえたサイバーセキュリティ対策の強化について関係省庁・関係機関による施策を取りまとめた^{*195}。

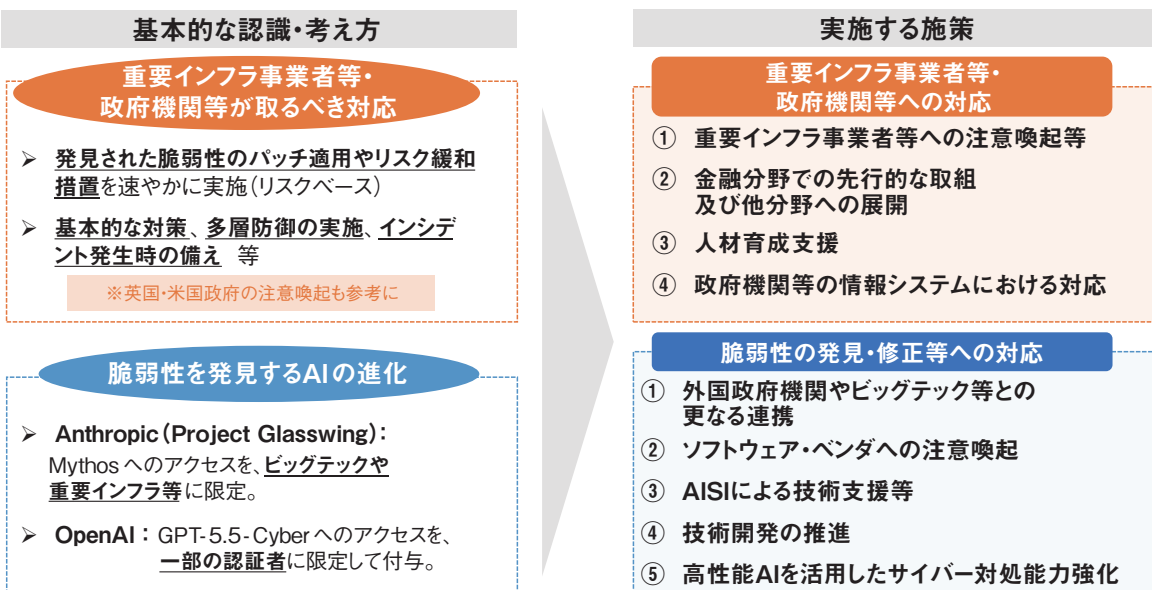
背景には、AI技術は急速に進展・普及しており、サイバー攻撃にAIが悪用されることで、攻撃のスピード・規模が劇的に増加する等、サイバーセキュリティにおける脅威に直面している状況がある。特に「フロンティアAI」と呼ばれる高性能AIによる脆弱性の発見・修正等のサイバーセキュリティ性能の急速な向上を踏まえ、重要インフラ事業者等への対応と脆弱性の発見・修正等

に関する対応の双方に迅速に取り組む必要があるとされた。高性能AIは、ベンダー等における脆弱性の発見・修正等や重要インフラ事業者等における検知・対応等のサイバーセキュリティ対策に活用することにより、我が国のサイバー対処能力の更なる強化が期待できる一方、攻撃者に悪用されることにより、サイバー攻撃がより高速かつ大規模に行われるおそれがある。このため、悪用リスクを前提として、防御側でも高性能AIを積極的に活用する方針が示された。

図2-2-2にProject YATA-Shieldの概要を示す。

Project YATA-Shieldでは、重要インフラ事業者等に対し、経営層のリーダーシップのもと、基本的なサイバーセキュリティ対策の確実な実施や、より高速かつ大量に脆弱性が発見・修正されることを前提とした対策強化を求めている。注意喚起文書^{*197}では、サイバーセキュリティ対策は、企業活動におけるコストや損失を減らすために必要な投資（将来の事業活動・成長に必須な費用）と位置付けることが重要とされ、資産管理、リスクアセスメント、脆弱性管理、認証・アクセス制御、バックアップ、監視・分析、事業継続計画、インシデント対応及び復旧、サプライチェーン・リスクへの対応等の基本的対策の徹底を求めている。加えて、「ゼロトラスト」の考え方に基づくシステム設計・運用への移行、脅威ハンティング等の強化、高性能AIを活用したサイバーセキュリティ対策の強化も挙げている。

- フロンティアAIモデルによるサイバーセキュリティ性能が向上する中においても、我が国のサイバーセキュリティが確保されるよう、**総理指示を踏まえて、政府全体としての対策パッケージを取りまとめ**。（5月18日）



■ 図2-2-2 Project YATA-Shieldの概要
(出典) NCO「AI性能の高度化を踏まえたサイバーセキュリティ対策の強化^{*196}」

具体的な活動として、金融分野について、2026年5月14日に、金融業界とIT事業者、政府・日本銀行等がAI技術の進展による脅威について共通の理解を持ち、対応を検討していくため、実務者レベルでの議論を深めることを目的とした作業部会を開催した^{*198}。また経済産業省でも、同年5月1日に、官民での認識共有を図るための意見交換を行った^{*199}。

政府機関等に対しても、政府統一基準に基づく基本的な対策を徹底するとともに、より高速かつ大量に脆弱性が発見・修正されることを前提とした対策の強化を求めている。特に、セキュリティパッチ（修正プログラム）の速やかな適用は非常に重要であり、必要に応じて、セキュリティパッチ等の対策用ファイルの適時の適用を前提とした運用設計（パッチマネジメント（管理））を見直すこととしている。あわせて、インシデントやその予兆を確認した場合には、NCOへの速やかな報告を求めている^{*200}。

脆弱性の発見・修正等への対応では、ソフトウェアベンダーに対し、セキュア・バイ・デザイン（設計）に基づき、高性能AIも活用しながら、ソフトウェア開発ライフサイクル全体において脆弱性の早期発見・対応に率先して取り組むよう注意喚起している^{*201}。具体的には、リリース前のソフトウェアについては、高性能AIを積極的に活用し、リリース後の脆弱性発見の可能性を低減させた上で、リリースをする。リリース後のソフトウェアについては、脆弱性が発見された場合に（必要に応じ高性能AIも活用して）パッチを早急に作成し、顧客に速やかに提供することを求めている。

更に、日本AISIによる情報収集・評価の実施及びこれを踏まえた情報提供、ガイドラインの策定、外国政府機関やAI開発者等との連携、AIを活用したサイバー対処能力の向上も進めるとしている。政府は、AIを巡る動向が急速に変化する中においても、より実効的な対応を継続的に行うべく、施策の実施状況を機動的に確認し、追加的対応を不断に検証・実施するとしている。

(5) AIに関するセキュリティ対策の考え方

本項では、AIに関するセキュリティ対策の考え方、参考となる情報を紹介する。なお、AIの活用にあたっては、「2.2.2 AIのリスク」に例示したようにセキュリティに限らない幅広いリスクが存在するため、「2.2.3 AIの安全性とガバナンス」で紹介したAIセーフティの考え方重要となる。その上で、ここでは特にセキュリティに焦点を当てる。

前述のとおり、セキュリティの分野においてもAIは様々

な影響を与えている。しかし、従来のセキュリティ対策の考え方が不要になるわけではなく、むしろ引き続き重要であることを強調したい。「2.2.4 (1) (b) AIの技術的特徴によらない攻撃」では、AIの急速な普及に伴い、その周辺のソフトウェアが従来型のサイバー攻撃の標的となっている事例を紹介した。また、「2.2.4 (2) AIを悪用したサイバー攻撃への対処」では、AIによる参入障壁の低下や攻撃の自動化によって、従来型のサイバー攻撃が増加している側面を指摘した。このように従来型のサイバー攻撃は、AIによって淘汰されるところか激化しており、それらへの対策は引き続き重要である。その上で、AIに特徴的な要素を踏まえた対策を新たに取り込んでいく必要がある。

考慮すべき点の一つは、「2.2.4 (1) (a) AIの技術的特徴を狙った攻撃」で挙げた、AIを標的とした攻撃への対策である。2026年3月に総務省が公表した「AIのセキュリティ確保のための技術的対策に係るガイドライン」では、プロンプトインジェクションを中心としたAI特有の攻撃に関する具体的な技術的対策例を示している^{*202}。また、日本AISIでは、AIを含むシステムに施した対策について、セキュリティを含むAIセーフティの観点で攻撃者の視点から評価する「AIセーフティに関するレッドチーミング手法ガイド^{*203}」や、レッドチーミングの支援を行うOSSのツール^{*204}を公開している。更に日本AISIは、AIのブラックボックス性や自律性によって事前対策だけでは被害の最小化が困難であるとして、観測性と制御性を中核としたAIシステム向けインシデントレスポンスの枠組みも提唱している^{*205}。

AIがセキュリティ対策にもたらす別の側面として、ソフトウェア更新への影響もある。「2.2.4 (3) AIを活用したサイバーセキュリティ確保」で示したようにAIが未知の脆弱性を大量に発見する能力を持ち始めている。NCSCは、AIによって発見される大量の脆弱性の修正を「脆弱性パッチの波」と表現し^{*206}、これに備えて組織は大規模で迅速かつ頻繁に更新する計画を立てるべきだとして、脆弱性管理のガイダンスを公表している^{*207}。

更に、ガバナンスの観点からは、「シャドー AI」と呼ばれる、組織の管理部門が認識しないAI利用への対応も考慮する必要がある。民間企業団体である一般社団法人AIガバナンス協会は、シャドー AIに関して従業員による個人的利用だけでなく、既存ソフトウェアの更新によってAI機能が追加されるケースや、外部パートナーによるAI利用についても課題が存在するとする調査結果を公表し、ヒアリング結果を踏まえて整理したAIガバ

ナンスのアプローチを紹介している^{*208}。

最後に、AIとセキュリティを巡る状況は絶えず変化しており、継続的な情報収集も重要と言える。内閣府主導で創設された「経済安全保障重要技術育成プログラム」(通称、K Program)の中では、「安心安全なAI利活用の為の知識・技術の体系化と知識集約環境構築」が進められており、その成果を活用した解説記事や体系化のマップ等が掲載されたポータルサイトが公開されている^{*209}。また、IPAでは「AIセキュリティ短信^{*210}」として注意すべきインシデント事例や把握しておくべき動

向を発信する取り組みを始めている。

本節「2.2 AIの安全性確保の取り組み」の冒頭では、執筆時点から約10年前を振り返り、AlphaGoによる囲碁のプロ棋士への勝利に触れた。当時から2026年の現状を予測するのは困難であったが、AIの発展は更に加速し、現時点では1年後の社会の変化を予測することすら難しくなっている。そのため、日頃から情報収集を行い、将来現れる新たな脅威に対しても進化し続けるAIをうまく活用しながら対応していくことが求められると言えよう。



AI時代のデジタル・シティズンシップと責任

インターネットで多くの人が頻繁に利用する検索機能。いつからか AI による要約が示されるようになり、それを多くの人が自然と受け入れています。生成 AI が驚く程の速さで発達・利用拡大し、気付けば寄り添って手助けしてくれている今、大人達がこれまでの経験から得た常識・考え方・危機管理術のアップデートは急務です。

そこで注目したいのが、「デジタル・シティズンシップ」＝「デジタル技術の利用を通じて、社会に積極的に関与し、参加する（責任ある行動を取る）能力」です。これは、年齢や立場を問わず現代社会を生きる誰もが身に付けたい能力であり、「情報セキュリティ」「情報リテラシー」「情報モラル」も欠かせない要素となります。学校では、教育活動全体をとおしてデジタル・シティズンシップ教育が実施されるようになったものの、大人が学ぶ機会はなかなかありません。ランサムウェア、特殊詐欺、闇バイトへの勧誘等といった様々な危機が、ネットの向こうから待ったなしで襲ってくる時代、青少年を導く側でもある大人だからこそ、デジタル・シティズンシップを身に付けたいものです。

更に、偽・誤情報や危険な情報を察知する感度を上げることも重要です。AI が生成した情報のファクトチェックを行う割合は 10 代 20 代でより高く、男性が 7 割以上、女性が約 6 割という調査結果がありますⁱ。年齢が上がるにつれ割合が低下する理由はどうあれ、ファクトチェックについては若い層に軍配が上がるようです。

また、検索結果に出てくる AI の要約は便利な反面、ネット上に散在する情報の要約でしかなく、誰でも気付くような間違いや、参照リンク先が既にある古い情報もあります。それらによるミスリードの防止にはファクトチェックが不可欠ですが、その前にまず、入手した情報を「ちゃんと読む」ことを心がけてみませんか？

「ちゃんと読む」がクセになると、「ん？何かヘン??」と気付く感度も上がります。実はこれが鍵なのです。AI 利用の有無を問わず、世の中には正しい情報のような顔をした偽・誤情報、曖昧な情報、チェリーピッキング（都合の良い部分だけを拾う行為）、悪意の仕掛け、詐欺メールやメッセージ等もごまんと存在します。目にした情報をちゃんと読まず、ファクトチェックもせずに、アクセスしたり行動を起こしたりすれば、職場、家庭、自身の健康に支障をきたすことにもなりかねません。そうしたトラブルの責任は、十分に注意をせずに、その情報を受け入れた側にもあるのですから、情報への感度アップは現代人に必須と言えます。

何があろうと、AI 自身は責任を取れません。責任を取れるのは人（開発者、提供者、利用者）だけです。一般的に AI サービスの利用規約には「利用者自身の責任において利用すること」が定められているため、責任は利用者が負うことになります。

こう考えると、AI 利用に伴う責任について、今まで以上に真剣に向き合う必要がありそうです。まずは気付きの感度を上げ、気になることがあったら、身近な信頼できる人に尋ねてみるといったアナログな対応が、デジタル・シティズンシップのベースになるような気がします。職場でも家庭でも、学校や地域でも、手軽にできそうなこうしたことから実践してみたいかがでしょうか。

i NTTドコモ モバイル社会研究所：【ライフスタイル】生成 AI のファクトチェック、若年層・活用層・人とのつながりを大切にする層が実施する傾向 <https://www.moba-ken.jp/project/lifestyle/20251106.html> [2026/7/27 確認]

2.3 偽・誤情報対策の動向

本節では、虚偽を含む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。

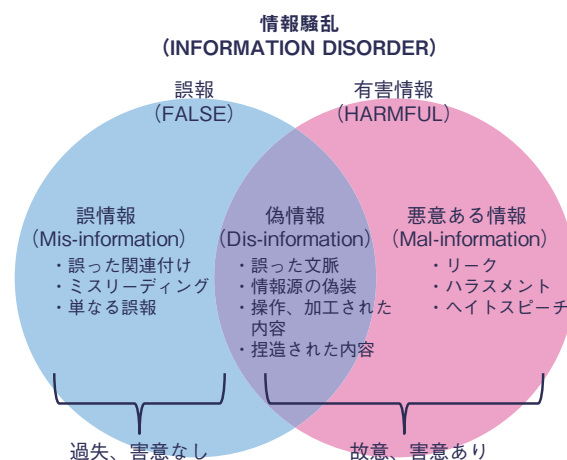
国家が関与するサイバー攻撃は、その手法等から分類すると表 2-3-1 のように整理できる。近年の情報戦・認知戦においては、虚偽を含む情報を流布する「情報操作型サイバー攻撃」と、それを含む様々な類型のサイバー攻撃を組み合わせた「ハイブリッド型サイバー攻撃」が大きな脅威となっている。本節では、表 2-3-1 の攻撃類型を使用して以降の解説を行う。なお、誤情報（ミスインフォメーション）と偽情報（ディスインフォメーション）を組み合わせた政府の用法^{*211}に従い、合わせて「偽・誤情報」と表記する。

2.3.1 偽・誤情報の定義

虚偽を含む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 2-3-1）。

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（Mis-information、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（Dis-information、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



■ 図 2-3-1 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」^{*213}を基に IPA が作成
© Council of Europe, reproduced with permission (from p5 Council of Europe report DGI(2017)09 Information disorder: Towards an interdisciplinary framework for research and policy making)

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 (Mal-information、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。

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

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（主に民主主義国家）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 2-3-1 国家が関与するサイバー攻撃の類型
(出典) 大澤淳「サイバー領域の安全保障政策の方向性」^{*212}を基に IPA が加筆・編集

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、これらについて確定的かつ共通の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。一般社団法人セーファーインターネット協会が、偽情報対策について多様なステークホルダーによる協力関係の構築と対話の枠組みを設けるべく設置した 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^{*214}。また、欧州対外活動庁（EEAS：European External Action Service）の2023年のレポート^{*215}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。公共への害とは、「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、その目的として経済的利益の追求または公衆を欺く故意があることを明示するとともに、その結果として公共に害が生じ得ることを示している。日本語ではディスインフォメーションに「偽情報」という訳語が当てられているが、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。

2.3.2 偽・誤情報の最新情勢

本項においては、偽・誤情報を巡る世界的な情勢を概観し、統計的な調査等も踏まえつつ、2025 年を中心とした最新動向を解説する。

(1) 偽・誤情報の現在

世界経済フォーラム（WEF：World Economic Forum）が2026年1月に公表した「The Global Risks Report 2026^{*216}」では、今後、短期的（2年間）または長期的（10年間）に想定されるグローバルリスクの上位10位までを深刻度順に示している（表2-3-2）。「Misinformation and disinformation」（誤情報と偽情報）は2024年^{*217}及び2025年^{*218}の第1位から順位を下げたものの、短期的リスクの第2位となり、3年連続で上位のリスクと評価された。また、「Misinformation and disinformation」は、長期的リスクでも第4位であり、いずれの時間軸でも上位に位置する持続的リスクとされた。短期的リスク第6位の「Cyber insecurity」（サイバー不安）と、短期的リスク第3位の「Societal polarization」（社会の二極化）、及び短期的リスクでは第30位であったものの長期的リスクでは第5位となった「Adverse outcomes of AI technologies」（AI技術の悪影響）は、相互に関連するリスクとして認識されている。

2025年から2026年にかけて、偽・誤情報は情報空間内部の問題にとどまらず、地経学（Geeconomics）的・軍事的対立や社会分断、先端技術の浸透等と一体化した横断的リスクとして固定化した状況にある。

	Short term (2 years)	Long term (10 years)
1.	Goeconomic confrontation	Extreme weather events
2.	Misinformation and disinformation	Biodiversity loss and ecosystem collapse
3.	Societal polarization	Critical change to Earth systems
4.	Extreme weather events	Misinformation and disinformation
5.	State-based armed conflict	Adverse outcomes of AI technologies
6.	Cyber insecurity	Natural resource shortages
7.	Inequality	Inequality
8.	Erosion of human rights and/or of civic freedoms	Cyber insecurity
9.	Pollution	Societal polarization
10.	Involuntary migration or displacement	Pollution

Risk categories

■ Economic ■ Environmental ■ Geopolitical ■ Societal ■ Technological

Source

World Economic Forum Global Risks Perception Survey 2025-2026

■表2-3-2 Global risks ranked by severity, short term (2 years) and long term (10 years)

(出典) WEF「The Global Risks Report 2026」を基に IPA が編集

更に、市民側の偽・誤情報に対する脅威認識としても、不安が高止まりしている状況にある。欧米、アジア、アフリカ主要国を調査対象としたロイター・ジャーナリズム研究所 (Reuters Institute for the Study of Journalism) の「Digital News Report 2025」^{*219}によれば、調査対象者の半数以上 (58%) が、オンラインニュースにおいて真偽を見分ける能力について依然として懸念を抱いていると回答しており、この割合は 2024 年と同水準である。懸念が最も高いのはアフリカ (73%) と米国 (73%) で、最も低いのは西ヨーロッパ (46%) となっている。同調査で、虚偽または誤解を招く情報の情報源について、重大な脅威と見なされているのは、調査対象の主要国全体で見ると、「オンラインのインフルエンサーや著名人」(47%) と、「自国の政治家や政党」(47%) であった。国別に見ると、インフルエンサーに対する懸念が最も高いのは、ナイジェリア (58%) やケニア (59%) 等のアフリカ諸国である一方、自国の政治家や政党が最大の脅威と見なされているのは、米国 (57%)、スペイン (57%)、そしてセルビア (59%)、スロバキア (56%)、ハンガリー (54%) を含む東欧の多くの国々であった^{*220}。

SNS プラットフォームの動向としては、Meta Platforms, Inc. (以下、Meta 社) は、2025 年上半期に、EU 加盟国向けの Facebook と Instagram の広告について 500 万件超を削除し、このうち誤情報ポリシー違反を理由に削除された広告は 8 万 3,000 件超に上ったと報告した。また同社は、EU 域内の 120 万件超の広告に資金提供者を明示する「paid for by」表示を付与した^{*221}。過去にも政治広告が各国の選挙干渉に利用されたことを踏まえ、SNS 上の広告の削除対応や透明化が進んでおり、主要な偽・誤情報対策の一つとなっている。

また、Meta 社は 2025 年を通じて約 35 億件の偽アカウントに対して措置を講じたことも報告している^{*222}。この偽アカウントの件数は、組織的に偽・誤情報を拡散するアカウント群の規模の大きさを示唆するものであり、コンテンツ単位の対処と並んで、偽アカウント等の拡散基盤そのものへの対処が対策の柱となっている。

他方で、偽・誤情報の生成には、AI の悪用が拍車をかけている状況にある。OpenAI 社は、2024 年 2 月に脅威レポートの公表を開始して以降、2025 年 10 月時点までに利用規約に違反した 40 件超の悪用ネットワークによる AI サービスの利用を遮断・報告したと公表した^{*223}。対象には、権威主義国家に関連する国家アクターによる影響工作キャンペーン、詐欺、悪意あるサイバー活動等が含まれている。更に、2026 年 2 月に公表

された最新の報告書では、2 年間の脅威報告を総括し、ロシアのコンテンツファームやインドネシアの詐欺グループの事例から、脅威アクターが工作の工程ごとに複数の AI モデルを使い分け、SNS アカウントや Web サイト等の従来型ツールと組み合わせて運用する傾向が指摘された^{*224}。また日本の関連事案としては、中国発の影響工作「Spamouflage」との関連性が指摘される国家アクターが、高市早苗内閣総理大臣 (以下、高市総理) を標的とした対日影響工作の立案に ChatGPT を利用しようとした事例も同報告書の中で報告されている。生成 AI が偽・誤情報の生成・翻訳・拡散の各工程を効率化しており、国家アクターによる AI の悪用の拡大が強く懸念される状況にある。

(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、サイバー攻撃や偽情報等あらゆる手段を、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^{*225}。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^{*226}。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。そのような認知情報処理フロー下における個々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現在の IoT 社会においては、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^{*227}。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・認知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^{*228}。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^{*229}、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標的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^{*230}。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FIMIの脅威に関する第1回報告書では、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（両国の重なりが見られる事例が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^{*231}、続く第2回、第3回の報告書でも中国やロシアを主要アクターに挙げている^{*232}。

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社は2025年上半期、ドイツ、ルーマニア、ポルトガル、ポーランドの各選挙に対応し、選挙運用センターの設置、迅速対応システムの運用、市民社会・ファクトチェック機関との報告経路の整備を行ったと報告している^{*221}。ルーマニアの事例では、1件のCIB (Coordinated Inauthentic Behavior: 組織的な不正行動) ネットワークを削除し、ルーマニアを標的とする影響工作を早期に遮断したことも報告されている^{*221}。

Google社の脅威分析部門であるGTIG (Google Threat Intelligence Group) は、2025年9月のロシア製ドローンによるポーランド領空侵入の直後、親ロシア系アクターが関与した影響工作により「ロシアの責任否定」「NATO・西側諸国への責任転嫁」「ポーランド政府への不信醸成」「ウクライナ支援の弱体化」を狙った情報拡散が展開されたと分析した^{*233}。物理的なインシデントの直後に情報戦が即座に発動するという、現代のハイブ

リッド戦の様相を示す事例である。

紛争等の有事の際には、情報戦・認知戦も活発化する。2025年5月のインド・パキスタン危機では、AI生成映像、過去の映像、ゲーム映像、別地域の映像等が軍事的成果や勝利を示す証拠として両国から大量に拡散された。英国 Guardian 紙は、CSOH (Centre for the Study of Organized Hate: 組織的憎悪の研究センター) の分析として、問題のあるX上の投稿を427件確認し、その一部は約1,000万ビューに達した一方、警告表示が付されたものは73件にとどまったと報じている^{*234}。この事例において戦略的に大量の偽情報が拡散されたことは、ナラティブを操作し両国の緊張を高めることで、両国の衝突を核戦争にエスカレーションさせるだけの推進力を持っていたと分析されている。更に2025年12月には、タイ・カンボジア間で武力衝突が発生したが、ここでも偽情報拡散やナラティブの戦いが問題となった。そのため、2025年12月27日に停戦延長のために署名した両国共同声明では、「双方は緊張緩和、否定的な世論の緩和、平和的対話に資する環境の醸成を図るため、虚偽の情報や偽ニュースの拡散を抑えることに合意する」といった文言が盛り込まれるに至った^{*235}。近年の情報戦・認知戦は、情報空間だけの問題にとどまらず、物理的な戦争と地続きの戦闘様相として立ち現れていることがうかがえる。

2.3.3 2025年度の注目事例

本項では、具体的な事例のうち、2025年度の注目すべき事例を取り上げて解説する。2025年度以前から継続している事例についても取り扱う。

(1) 欧州の選挙への影響工作

2025年9月28日に実施されたモルドバ議会選挙において、モルドバ当局によれば、選挙干渉を目的として、大規模な買収工作、同年中に1,000件を超える重要インフラへのサイバー攻撃、投票日の暴動扇動計画、そして大規模なオンライン偽情報キャンペーンが複合的に展開されたという。ロシアのアクターはモルドバ市民をTikTokやFacebook上の組織的偽情報拡散に金銭で動員し、Maia Sandu大統領や与党を攻撃するAI生成コンテンツを拡散させた^{*236}。これに対しモルドバ側は、欧州委員会 (European Commission) と共同で選挙前にデジタル・ハイブリッド脅威のシミュレーションを実施し、Google社、Meta社、TikTok Inc. (以下、TikTok社) の参

加を得て対応体制を準備するとともに、外国の干渉に対抗するための専門家グループ「Hybrid Rapid Response Team」をEUから受け入れ、サイバー攻撃を含めたEU共同演習の導入、欧州デジタルメディア監視所(EDMO: European Digital Media Observatory)の新たな地域ハブを設立する等、EUとの共同歩調で積極的な対策を講じた^{*237}。欧州安全保障協力機構(OSCE: Organization for Security and Co-operation in Europe)等の国際選挙監視団は、モルドバの議会選挙は競争的で有権者に政治的選択肢を提供したが、一方で、外国の干渉、違法資金、サイバー攻撃、広範な偽情報の拡散がこの民主主義プロセスを妨害したと結論付けている^{*238}。

2025年2月23日に実施されたドイツ連邦議会選挙においては、ロシアによる同選挙への影響工作が懸念されていた^{*239}。同選挙に対する偽情報による影響工作について、選挙後、背後にロシア連邦軍参謀本部情報総局(GRU: Glavnoye Razvedyvatel'noye Upravleniye)と関係する組織が存在するとして、ドイツ政府は2025年12月、ベルリン駐在のロシア大使を召喚して抗議した^{*240}。この事例は、事後の調査により、選挙干渉におけるアクターの特定期間への帰属(アトリビューション)が明らかにされ、外交措置に発展した事例として注目すべき事例となった。

2025年5月18日に第1回投票、6月1日に決選投票が実施されたポーランド大統領選挙では、ロシアによる影響工作が懸念されたため、ポーランド政府は2025年2月からデジタル化省が主導する「選挙アンブレラ(Election Umbrella)」と称する選挙保護計画を発動し、偽・誤情報対策としては、ポーランドの国立研究機関であるNASKがSNS等の監視を実施した。デジタル化省によると、NASKは、Meta社、X Corp.(以下、X社)、TikTok社のプラットフォーム上の1万6,000件を超える偽アカウント(候補者を装った169アカウントを含む)の無効化等を実施したという^{*241}。

(2) 武力衝突・軍事作戦に関連した影響工作

2025年6月13日から25日にかけてのイスラエルとイランの軍事衝突は、生成AIによるディープフェイク流布が特徴的な事例であった。BBC Verifyの分析によれば、イランの軍事的成果を誇張するAI生成動画等が大量に拡散され、最も視聴された偽動画3本だけで合計1億回以上の閲覧を集めたという。AI生成映像以外に、ゲーム映像の転用によりAI判定を逃してしまうものや、X上のAIチャットボットによる偽動画の誤判定も確認さ

れ、検証インフラ自体の機能不全が露呈した。なお、親イスラエルのアカウントもオンラインで偽情報を拡散しており、主にイランでの抗議や集会の古い映像を再配布し、イラン政府に対する反発やイスラエルの軍事作戦支持が増していると主張をしているという^{*242}。

イスラエルのディアスポラ問題・反ユダヤ主義対策省の2025年6月報告書によれば、イラン政府とイスラム革命防衛隊(IRGC: Islamic Revolutionary Guard Corps)による、紛争下で世論に影響を与えることを目的としたボットネットを解明したという。Xにおける紛争関連ハッシュタグの最大60%がイラン関連アクターに由来し、ボットネットは国内向けに最高指導者ハメネイ師の権威や国民団結のメッセージを増幅させており、対イスラエル向けには報復的なナラティブを中心に展開したとされる^{*243}。また、EDMOによれば、ボットアカウントだけでなく、国営のテヘランタイムズやFars News通信社もAI生成映像を多用したディープフェイクやミスリーディングな動画を拡散していたという^{*244}。

更に2026年2月28日、米国とイスラエルはイランに対する大規模攻撃を開始し、紛争が再燃した。クレムソン大学のMedia Forensics Hubは2026年3月、IRGC関連の少なくとも62アカウント(X、Instagram、Bluesky上)からなるネットワーク「From Texas to Tehran」を特定し、X上だけで5万9,403件の投稿が確認されたと報告した^{*245}。同時期には、米国空母エイブラハム・リンカーン撃沈を主張する偽情報(米国中央軍が公式に否定^{*246})やネタニヤフ首相の死亡を誤認させるような虚偽の記事やAI生成画像^{*247}、イランの攻撃によりカタールにある米国のレーダーが破壊された様子を示していると主張するAI生成画像等が拡散された^{*248}。

米国は2025年9月以降、ベネズエラ周辺で「麻薬密輸船」に対する攻撃を繰り返して軍事的圧力を段階的に強め、2026年1月3日、ベネズエラのNicolás Maduro Moros大統領を拘束し、米国へ移送・訴追した特殊部隊による急襲作戦「Operation Absolute Resolve」を実施した^{*249}。同作戦の実行直後にも、SNS上にはAI生成映像や過去の動画を流用した偽・誤情報が一気に拡散された。AFP(Agence France-Presse)のファクトチェックによれば、「米軍拘束下のMaduro氏の最初の写真」とされた画像はAI生成(GoogleのSynthID電子透かしを検出)であり、「頭部に袋を被せられたMaduro氏」とされた画像は2003年のフセイン元イラク大統領拘束時の写真の流用であった^{*250}。また、「Maduro排除を祝うベネズエラ市民」だとするAI生成動画の拡散も多

く見られた^{*251}。NewsGuard Technologies, Inc.（以下、NewsGuard 社）は、作戦に関連する捏造・誤用の画像・動画 7 件が X だけで 2 日間に合計 1,400 万回以上閲覧されたと報告している^{*250}。

(3) 新型コロナウイルス・反ワクチンに関する偽・誤情報

新型コロナウイルス感染症（以下、新型コロナウイルス）をめぐっては、パンデミック期に世界保健機関（WHO：World Health Organization）が「インフォデミック」と呼んだ大量の偽・誤情報や陰謀論の氾濫が発生した^{*252}。そのうち、反ワクチンに関する偽・誤情報や虚偽的なナラティブ形成はパンデミック収束後も消滅せず、2025 年度も各地で公衆衛生上の影響を及ぼし続けている。WHO 及び UNICEF（United Nations International Children's Emergency Fund：国連児童基金）の 2026 年 2 月のプレスリリースでは、欧州・中央アジア 53 カ国の 2025 年麻疹症例 3 万 3,998 件の背景として、「誤情報により煽られたワクチン忌避」が指摘されている^{*253}。日本国内でも、日本ファクトチェックセンター（JFC：Japan Fact-check Center）の検証によれば、2025 年中、新型コロナウイルス・反ワクチン関連の誤情報が継続的に拡散した。具体的には、「Meiji Seika ファルマ株式会社の新型コロナレプリコンワクチンはファイザー製の 75 倍の死亡率である^{*254}」「フランスの研究で mRNA ワクチン接種者の重症 COVID-19 関連死亡リスクが 74% 上昇した^{*255}」等が誤りであったと指摘されている。パンデミック収束後の現在においても、新型コロナウイルスやそこから派生した反ワクチン関連の偽・誤情報の発信・拡散は長期化していることがうかがえる。

2.3.4 偽・誤情報への対策の動向

偽・誤情報への対策は、国家・地域レベルの法制度整備、プラットフォーム事業者等の民間レベルの自主的対策、個人レベルの情報リテラシー向上の三層で進展している。2025 年度は、規制アプローチの地域的な特色がより強まった時期であった。

(1) 各国・地域の対応

国、地域別の対応では、従前より偽・誤情報対策に積極的であった EU の対策が更に進んだことが見て取れる。また、オーストラリアによる SNS の年齢制限規制を端緒に、同様の規制が各国にも波及したことが 2025

年の大きなトレンドであった。

(a) EU

EU では、2018 年以來の自主的枠組みであった「偽情報に関する行動規範（Code of Practice on Disinformation）」が、2025 年 2 月 13 日に欧州委員会等の承認を経て「デジタルサービス法（DSA：Digital Services Act）」上の「行動規範（Code of Conduct）」へ転換され、同年 7 月 1 日から発効した^{*256}。これにより、DSA に基づき「超大規模オンラインプラットフォーム（VLOP：Very Large Online Platform）」及び「超大規模オンライン検索エンジン（VLOSE：Very Large Online Search Engine）」に指定されたサービス（Facebook、Instagram、LinkedIn、Bing、TikTok、YouTube、Google Search）を運営する事業者等にとって規範の遵守状況は DSA 上のリスク低減義務の履行を示す基準となり、独立監査の対象ともなった。また、「政治広告の透明性とターゲティングに関する規則（TTPA：Transparency and Targeting of Political Advertising）」が 2025 年 10 月 10 日から適用され、政治広告へのラベル付与、資金提供者・対象選挙・支出額等の開示が義務化された^{*257}。これに対し Meta 社と Google 社は、法的不確実性等を理由に EU 域内での政治・選挙・社会的争点を対象とした広告の配信自体を停止するという対応を選択し^{*258}、規制強化によりプラットフォームがテーマによっては広告配信を行わないという事態に至った。更に 2026 年 3 月 24 日には、行動規範に署名した事業者の初めての透明性報告書（対象期間：2025 年 7 ～ 12 月）が公表され^{*259}、Meta 社は同報告書において、2025 年 10 月 6 日以降 EU 域内で政治・選挙・社会的争点を対象とした広告の配信を実際に停止したことを報告している^{*260}。これにより、2025 年に強化して始動した EU の共同規制サイクルが、年度中に一定の実効性ある成果に結実したといえるだろう。

また、EU 内で影響工作等の事案が発生した国に対しても、個別に協働体制をとって即応的な対策を講じている。2024 年 11 月の大統領選挙がロシアの干渉と TikTok 上の組織的キャンペーンを理由に憲法裁判所により無効とされたルーマニアでは、2025 年 5 月の再選挙に向けて選挙即応型の対策が講じられた。欧州委員会 DSA に基づき TikTok 社への正式調査を開始し同社に対策を促したことによって、TikTok 社は政治アカウントの検出・ラベリングの強化、ルーマニア語の専門人材の増員、秘密裏に行われる影響工作への対抗と広告

の健全性を担当する選挙タスクフォースへの120名の専門家追加等の措置を講ずることとなった^{*261}。

(b) 米国

米国では連邦レベルの包括的規制が停滞する一方、州レベルの立法が加速した。非党派のオンライン政治百科事典を運営する非営利団体 Ballotpedia の集計によれば、政治的ディープフェイクを規制する州法を有する州は2025年1月の21州から同年7月には28州へ増加し、ディープフェイク関連立法全体では47州が何らかの法律を制定済みである^{*262}。多くの州法は選挙前の一定期間における候補者の合成メディアの頒布に開示義務や差し止め・罰則を課すものであり、表現の自由(合衆国憲法修正第1条)との均衡が引き続き争点となっている。

(c) オーストラリア

オーストラリアでは2024年12月11日、世界に先駆けて「2024年オンライン安全修正(ソーシャルメディア最低年齢)法(Online Safety Amendment (Social Media Minimum Age) Act 2024)」が施行された。同法は2021年の「オンライン安全法(Online Safety Act 2021)」にPart 4Aを新設し、特定のソーシャルメディア・プラットフォームに対し、16歳未満の利用者がアカウントを保有することを防ぐ「合理的措置(reasonable steps)」を講じる義務を課す規定を設けた^{*263}。オーストラリアの独立規制当局であるeSafety Commissionerは、Facebook、Instagram、Snapchat、TikTok、YouTube、X、Threads、Reddit、Kick、Twitchの10プラットフォームを対象に、利用者の対象年齢に制限を設け^{*264}、体系的な不遵守が認定された事業者に対しては最大4,950万豪ドルの民事罰が科される。偽・誤情報そのものを対象とする規制ではないが、未成年者を有害な情報環境から保護するという観点から行政が介入する規制である。

なお、2025年度中には、ニュージーランド^{*265}、フランス、ポルトガル、スペイン^{*266}、EU^{*267}、ブラジルといった複数の国・地域が同様の法制度の制定もしくは検討の方向で追随しており、未成年者保護のための構造的介入が世界的に拡大している^{*268}。

(d) 日本

日本では2025年4月、「改正プロバイダ責任制限法」を改称した「情報流通プラットフォーム対処法」が施行さ

れ、大規模プラットフォーム事業者に対し、権利侵害情報への削除申出窓口の整備、対応の迅速化、削除基準の公表等が義務付けられた^{*269}。偽・誤情報への直接の削除義務を課すものではないが、情報流通の適正化に向けた制度的基盤として位置付けられる。技術面では、総務省が「インターネット上の偽・誤情報等への対策技術の開発・実証事業(令和7年度)」を実施し、公募により技術開発主体14者、研究・調査主体6者を採択した^{*270}。2026年3月16日には、採択された14件の開発技術を中心に成果を発信するイベントが開催され、ディープフェイク検出、コミュニティノートのデータを活用した分析基盤、災害時の偽・誤情報対策等、対策技術の社会実装に向けた取り組みが紹介された^{*271}。偽・誤情報対策技術については、「経済安全保障重要技術育成プログラム」(通称、K Program)の「偽情報分析に係る技術の開発」でも開発が進められている^{*272}。

なお、民間レベルでは、日本ファクトチェックセンターが2025年7月の参院選に関連して27本のファクトチェック記事を公開した^{*273}ことに加え、同選挙では新聞社・テレビ局による検証記事が前年の衆院選比で5倍超に急増しており^{*274}、報道機関のファクトチェック参入が本格化した点も特筆される。

(2) プラットフォーム事業者の対応

プラットフォーム事業者のレベルでは、第一に選挙対応の制度化が進んだ。Meta社は選挙運用センターと迅速対応の枠組みを作り、欧州各国選挙で運用した^{*275}。TikTok社も各国選挙に際して選挙センターの設置や政治広告の禁止・透明性確保を進め、EUの行動規範下の迅速対応システムにも参加している^{*276}。第二に、AI生成コンテンツや操作された情報等に関する投稿へのラベル付与施策が拡大した。X社は生成AI等によって合成または操作されたメディアの欺瞞的共有を禁ずる「Authenticity(真正性)」ポリシーを維持しており^{*277}、2026年1月には編集・改変された動画像を含む投稿への「Manipulated Media(操作されたメディア)」警告ラベル機能の運用を開始し^{*278}、同年3月には投稿者自身が付与する「Made with AI」ラベルも実装された^{*279}。Meta社においても、「AI情報」ラベルやC2PA(Coalition for Content Provenance and Authenticity)等の来歴証明技術の実装が進む一方、メタデータ除去への脆弱性や検出技術の未成熟といった限界もMeta社が設置しているOversight Board(監督委員会)の審査等を通じて指摘されている^{*280}。第三に、政治広告・社会争点

広告の制限が進んだ。EUにおけるMeta社・Google社の政治広告停止はその具体的な結果の現れである^{*258}。第四に、AI開発企業による脅威情報の公開が定着した。OpenAI社・Google社等は影響工作等へのAI悪用を検知・遮断した事例を定期的に報告し、アカウント停止等の措置を講じるようになった^{*281}。第五に、ユーザー参加型のコミュニティノートシステムの拡張、AI連携が進んだ。Xの機能であるコミュニティノートに関しては、X社は2025年7月1日、AIにコミュニティノートの草案を提案させる「AI Note Writer API」のパイロットを開始し、AIがノート案を起草し多様な観点を持つ人間のユーザーが公開可否を判定する「AIが書き、人間が決める」モデルを導入した^{*282}。Meta社は、同年1月7日、第三者ファクトチェック・プログラムを終了しX型のコミュニティノートへ移行する方針を発表している^{*283}。これに対し、デジタルヘイト対策センター(CCDH: Center for Countering Digital Hate Inc)の2024年10月の分析では、Xのコミュニティノートでは「役立つ」と評価された正確なノートの大半が一般ユーザーに表示されておらず、元の誤投稿が平均で約13倍のビューを獲得していたと報告されており^{*284}、Regina Cazzamatta氏の学術論文^{*285}等でもコンテンツモデレーション体制の縮小・転換が偽・誤情報の拡散余地を広げる懸念があると指摘される等、プラットフォーム事業者の対応は強化と後退が併存する複雑な様相を呈している。

(3) 個人レベルのリテラシー対策

個人レベルの対策としては、流通後の検証であるファクトチェックに加え、偽情報に接する前に典型的な情報操作手法や狙われやすいナラティブへの心理的免疫を形成するプレバンキング(事前暴露)の重要性が国際的に強調されつつある。心理的な免疫を形成するための心理的接種理論に基づく短い啓発動画の視聴が、情報操作手法の識別力を向上させることは大規模実験で確認されている^{*286}。Google社のJigsaw部門は2024年欧州議会選挙前にスケープゴート化・脱文脈化・信用毀損といった三つの情報操作手法を扱うYouTube動画群をベルギー、フランス、ドイツ、イタリア、ポーランドの5カ国で展開し、これまでで最大となる1億2,000万人以上にリーチする取り組みを展開した^{*287}。2026年3月に学術誌Communications Psychologyに掲載された12カ国・約2万人を対象とする検証研究では、これらの動画が高齢層(45歳以上)を含む被験者の情報操作手法識別能力と共有判断を有意に改善したこと

が示された^{*288}。

日本でも、総務省は2024年の「デジタル空間における情報流通の健全性確保に関する検討会」のとりまとめ案において「プレバンキング(リテラシー向上等の事前の備え)とデバンキング(ファクトチェック等の事後検証)の両輪による対応」を素案として提示しており^{*289}、2025年1月から19のIT事業者・関連団体と連携した官民啓発プロジェクト「DIGITAL POSITIVE ACTION」を始動している^{*290}。総務省による偽・誤情報等への対策技術の開発・実証事業では、ゲーミフィケーションを活用した実践的ゲーム型プレバンキング技術(株式会社Classroom Adventure)の開発も採択されている^{*270}。ファクトチェックが拡散後の「治療」であるのに対し、プレバンキングは拡散前の「予防接種」として位置付けられるが、検証が物理的に追いつかない生成AI時代において、両者を組み合わせた重層的なリテラシー施策が今後の標準となると考えられる。

2.3.5 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

2025年度の情勢は、次の3点に集約できる。第一に、武力紛争や、紛争未満の物理空間における衝突や威圧等の事案と情報戦の同期である。ロシアによるポーランド領空侵入、インドとパキスタンの軍事衝突、イスラエル・米国とイラン間の軍事衝突、米国によるベネズエラに対する軍事作戦等の事案において、物理的事象と並行して情報戦・認知戦が展開され、交戦国同士や支援する陣営側において一体的な作戦として運用された。これが市民社会の言論空間にも一気に拡散し、物理的戦場と情報空間の戦場は地続きとなった様相である。第二に、選挙干渉の常態化であり、ドイツ・ルーマニア・モルドバ、ポーランドそして日本に至るまで、主要な選挙が外国の影響工作の標的となる事例が継続的に頻発している。第三に、対策の地域的分岐である。EUの共同規制(DSA・行動規範・TTPA)、米国の州法主導、オーストラリア年齢規制、日本の技術開発・民間検証支援と、各法域のアプローチの差異が拡大しており、それぞれの地域の法制度に根差した特色が見られる。一方で、グローバルなプラットフォーム上の脅威に対して国際的な協働の動きは進んでおらず、規制の断片化がもたらす間

隙が新たな課題となっている。

(2) 今後の見通し

2026 年度下半期にも、日本への影響が大きい政治・外交イベントが相次ぐ。2026 年 10 月のブラジル大統領選挙、11 月の米国中間選挙及び台湾統一地方選挙は、いずれもロシア・中国等による影響工作や選挙干渉の標的となることが想定される。日本については、平素より認

知戦が生起していることに防衛大臣が言及するという環境にあり^{※291}、AI 生成コンテンツを悪用した情報戦・認知戦が展開される可能性は無視できない。AI を悪用した影響工作の自動化、パーソナライズ化、及び LLM グルーミングによる AI の学習データ汚染という新たな攻撃アプローチへの対応を含め、政府のみならず、プラットフォーム事業者、報道機関、市民社会が平時から連携する総力戦的なレジリエンスの構築が急務である。



C O L U M N

「AIの利用をめぐるサイバーリスク」が「情報セキュリティ10大脅威」に初登場

「情報セキュリティ10大脅威 2026」の「『組織』向け脅威」の3位として、「AIの利用をめぐるサイバーリスク」が初めて選出されました。この脅威名は、ランサム攻撃（1位）や標的型攻撃（5位）といった攻撃手法を名前に含む脅威名とは一線を画し、AIにまつわる包括的な懸念を表す名称となっています。

対話型生成AIであるChatGPTが2022年に登場して以降、生成AIに対する社会の関心の高まり、爆発的な利活用の進展により、個人の日常生活から労働環境、企業・組織の日常業務から技術開発といったあらゆる側面にAIは影響を与えつつあります。一方、サイバーセキュリティへの影響として、AIの悪用による攻撃の容易化、巧妙化、高速化等、多くの懸念が指摘されており、2025年度はそれらが現実のものとなった年でした。こうした近年の動向を踏まえ、「AIの利用をめぐるサイバーリスク」は、満を持してランクインしたようにも見えます。

では、今後もAIに関する脅威はランクインし続けるのでしょうか？

例えば、フィッシングは、「フィッシングによる個人情報等の詐取」として「情報セキュリティ10大脅威 2026」の「『個人』向け脅威」に選出されていますが、「『組織』向け脅威」にはありません。しかし、フィッシングは組織への脅威に無関係な存在ではなく、むしろ深く関わりがあります。1位のランサム攻撃にしろ、2位のサプライチェーンや委託先を狙った攻撃にしろ、攻撃の端緒となるIDやパスワードといった認証情報の詐取にフィッシングは多用されており、そのほかのランクインしている脅威とも関連があります。

AIの影響は広範囲であることから、いずれはAIが「情報セキュリティ10大脅威」のランク外になるとしても、フィッシングと同じように脅威の共通要素として、背後に存在し続けることになるのでは、と考えています。

表 情報セキュリティ10大脅威 2026「組織」向けの脅威の順位

順位	脅威名	10大脅威での取り扱い (2016年以降)
1	ランサム攻撃による被害	11年連続 11回目
2	サプライチェーンや委託先を狙った攻撃	8年連続 8回目
3	AIの利用をめぐるサイバーリスク	初選出
4	システムの脆弱性を悪用した攻撃	6年連続 9回目
5	機密情報を狙った標的型攻撃	11年連続 11回目
6	地政学的リスクに起因するサイバー攻撃（情報戦を含む）	2年連続 2回目
7	内部不正による情報漏えい等	11年連続 11回目
8	リモートワーク等の環境や仕組みを狙った攻撃	6年連続 6回目
9	DDoS攻撃（分散型サービス妨害攻撃）	2年連続 7回目
10	ビジネスメール詐欺	9年連続 9回目

i <https://www.ipa.go.jp/security/10threats/10threats2026.html> [2026/8/25 確認]

※ 1 内閣官房：国家安全保障戦略について <https://www.cas.go.jp/jp/siryoku/221216anzenhoshou/nss-j.pdf> [2026/7/17 確認]

※ 2 サイバーセキュリティ戦略本部：サイバー空間を巡る脅威に対応するため喫緊に取り組むべき事項（サイバー安全保障分野での対応能力の向上に向けた有識者会議） <https://www.cyber.go.jp/pdf/policy/kihons-250529kikkin.pdf> [2026/8/18 確認]

※ 3 https://www.cas.go.jp/jp/seisaku/cyber_anzen_hosyo/koujou_teigen/teigen.pdf [2026/7/17 確認]

※ 4 e-Gov 法令検索：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（令和七年法律第四十二号） https://laws.e-gov.go.jp/law/507AC0000000042/20271122_000000000000000 [2026/7/17 確認]

※ 5 衆議院：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 https://www.shugiin.go.jp/internet/itdb_housei.nsf/html/housei/21720250523043.htm [2026/7/17 確認]

※ 6 <https://www.cao.go.jp/cybersecurity/pdf/kihonhoushin.pdf> [2026/7/17 確認]

※ 7 <https://www.cyber.go.jp/pdf/policy/kihons-250529kikkin.pdf> [2026/10/13 確認]

※ 8 NCO：沿革 <https://www.cyber.go.jp/about/history/index.html> [2026/7/17 確認]

※ 9 <https://www.ipa.go.jp/publish/wp-security/2025.html> [2026/7/17 確認]

※ 10 NCO：サイバーセキュリティ戦略 https://www.cyber.go.jp/pdf/policy/kihons/cs_strategy2025.pdf [2026/7/17 確認]

※ 11 国会審議の過程で、衆議院において、サイバー対処能力強化法案に通信の秘密の尊重（第2条の2）等の規定を追加する修正が行われた。衆議院：閣法 第217回国会 4 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案に対する修正案 https://www.shugiin.go.jp/internet/itdb_gian.nsf/html/gian/honbun/syuuseian/11_7CD2.htm [2026/7/17 確認]

※ 12 内閣官房：サイバー対処能力強化法及び同整備法について https://www.cas.go.jp/jp/seisaku/cyber_anzen_hosyo_torikumi/pdf/setsumei.pdf [2026/7/17 確認]

※ 13 https://www.cas.go.jp/jp/seisaku/cyber_anzen_hosyo_torikumi/pdf/shin_cyber_leaflet.pdf [2026/7/17 確認]

※ 14 内閣官房：サイバー安全保障に関する取組（能動的サイバー防御の実現に向けた検討など） https://www.cas.go.jp/jp/seisaku/cyber_anzen_hosyo_torikumi/index.html [2026/7/17 確認]

※ 15 特定社会基盤事業者：特定社会基盤事業者（後述）のうち、重要電子計算機（行政機関や特定社会基盤事業者等が使用する電子計算機であって、サイバー攻撃を受けた場合に、国家及び国民の安全を害し、または国民生活もしくは経済活動に多大な影響を及ぼすおそれがあるような一定のもの）を使用する者。特定社会基盤事業者とは、経済安全保障推進法に基づき指定される、特定社会基盤事業（対象 15 事業（電気、ガス、石油、水道、鉄道、貨物自動車運送、外航海運、港湾運送、航空、空港、電気通信、放送、郵便、金融、クレジットカード）のうち、特定社会基盤役務（国民生活及び経済活動の基盤となる役務であって、その安定的な提供に支障が生じた場合に国家及び国民の安全を損なう事態を生ずるおそれがあるもの）の提供を行うものとして政令で定めるもの）を行う者のうち、その使用する特定重要設備（特定社会基盤事業の用に供される設備、機器、装置またはプログラムのうち、特定社会基盤役務を安定的に提供するために重要であり、かつ我が国の外部から行われる特定社会基盤役務の安定的な提供を妨害する行為の手段として使用されるおそれがあるものとして主務省令で定めるもの）の機能が停止し、または低下した場合に、その提供する特定社会基盤役務の安定的な提供に支障が生じ、これによって国家及び国民の安全を損なう事態を生ずるおそれ大きいものとして主務省令で定める基準に該当する者。2026 年 7 月 1 日時点で 258 者。なお、2026 年 6 月 17 日、医療分野を特定社会基盤事業に追加する経済安全保障推進法の改正法が公布され、公布日から起算して 1 年 6 ヶ月を超えない範囲内において政令で定める日から施行される予定である。これにより、特定社会基盤事業は、16 事業となる。特定社会基盤事業及び指定事業者数は、内閣府「経済安全保障推進法の特定社会基盤役務の安定的な提供の確保に関する制度について」(https://www.cao.go.jp/keizai_anzen_hosho/suishinhou/infra/doc/infra_gaiyou.pdf [2026/7/17 確認])より引用した。

※ 16 内閣府：サイバー対処能力強化法の施行等に関する有識者会議 <https://www.cao.go.jp/cybersecurity/kaigi/kaigi.html> [2026/7/17 確認]

※ 17 通信情報保有機関：サイバー対処能力強化法では、次に挙げる行政機関（サイバー通信情報管理委員会を除く）と規定している。すなわち、「内閣府」「選別後通信情報（取得した通信情報についての自動選別により得られたもの）の提供を受けた行政機関であって、現に当該選別後通信情報（その全部又は一部を複製し、又は加工した選別後通信情報を含む）

を保有しているもの。」「選別後通信情報を含む総合整理分析情報（報告等情報、選別後通信情報、提供用選別後情報、協議会を通じて得た情報その他の情報が重要電子計算機に対する特定不正行為による被害の防止に有効に活用されるよう、当該情報整理及び分析を行うもの）の提供を受けた国の行政機関であって、現に当該選別後通信情報（その全部又は一部を複製し、又は加工した選別後通信情報を含む）を保有しているもの。」の 3 項目である。

※ 18 https://www.cao.go.jp/cybersecurity/pdf/kihonhoushin_gaiyou.pdf [2026/7/17 確認]

※ 19 <https://laws.e-gov.go.jp/law/504AC0000000043> [2026/7/17 確認]

※ 20 <https://laws.e-gov.go.jp/law/506AC0000000027> [2026/7/17 確認]

※ 21 内閣府：経済安全保障推進法に基づく重要物資の安定的な供給の確保（サプライチェーン強靱化）に関する制度全般や技術流出防止について https://www.cao.go.jp/keizai_anzen_hosho/suishinhou/supply_chain/doc/sc_gaiyou.pdf [2026/7/17 確認]

※ 22 内閣府：経済安全保障推進法の特定社会基盤役務の安定的な提供の確保に関する制度について https://www.cao.go.jp/keizai_anzen_hosho/suishinhou/infra/doc/infra_gaiyou.pdf [2026/7/17 確認]

※ 23 内閣官房：経済安全保障法制に関する有識者会議（令和4年度～） https://www.cas.go.jp/jp/seisaku/keizai_anzen_hosyohousei/4index.html [2026/7/17 確認]

※ 24-1 https://www.cas.go.jp/jp/seisaku/keizai_anzen_hosyohousei/r8_dai15/teigen.pdf [2026/7/17 確認]

※ 24-2 内閣府：概要 https://www.cao.go.jp/keizai_anzen_hosho/suishinhou/doc/houangaiyo.pdf [2026/8/18 確認]

※ 25 https://www.cas.go.jp/jp/seisaku/keizai_anzen_hosyo_sc/pdf/torimatome.pdf [2026/7/17 確認]

※ 26 セキュリティ・クリアランス制度：国家における情報保全措置の一環として、政府が保有する安全保障上重要な情報として指定された情報に対して、アクセスする必要がある者のうち、情報を漏らすおそれがないという信頼性を確認した者の中で取り扱うとする制度。

内閣府：いわゆる「セキュリティ・クリアランス」について https://www.cao.go.jp/keizai_anzen_hosho/hogokatsuyou/doc/sankou_clearance.pdf [2026/7/17 確認]

※ 27 国会審議の過程で、衆議院において、内閣総理大臣による「重要経済安保情報の指定及びその解除、適性評価の実施並びに適合事業者の認定の状況」の国会への報告（第 19 条）等を追加する修正が行われた。衆議院：閣法 第 213 回国会 24 重要経済安保情報の保護及び活用に関する法律案に対する修正案（自民、立憲、維教、公明、国民、有志） https://www.shugiin.go.jp/internet/itdb_gian.nsf/html/gian/honbun/syuuseian/1_8466.htm [2026/7/17 確認]

※ 28 https://www.cao.go.jp/keizai_anzen_hosho/hogokatsuyou/doc/sankou_hyouka.pdf [2026/7/17 確認]

※ 29 内閣府：重要経済安保情報の保護及び活用に関する法律の概要 https://www.cao.go.jp/keizai_anzen_hosho/hogokatsuyou/doc/gaiyo.pdf [2026/7/17 確認]

※ 30 内閣府：重要経済安保情報の指定及びその解除、適性評価の実施並びに適合事業者の認定に関し、統一的な運用を図るための基準の策定について https://www.cao.go.jp/keizai_anzen_hosho/hogokatsuyou/doc/kijun.pdf [2026/7/17 確認]

※ 31 内閣府：重要経済安保情報の保護及び活用に関する法律（重要経済安保情報保護活用法） https://www.cao.go.jp/keizai_anzen_hosho/hogokatsuyou/hogokatsuyou.html [2026/7/17 確認]

※ 32 内閣府：特定秘密・重要経済安保情報の検証及び監察 <https://www8.cao.go.jp/kenshoukansatsu/index.html> [2026/7/17 確認]

※ 33 独立公文書管理監、情報保全監察室：是正の求め等について <https://www8.cao.go.jp/kenshoukansatsu/zesei.pdf> [2026/7/17 確認]

※ 34 2025 年中、重要経済安保情報の指定は 9 機関 20 件、指定を解除したものは 1 件。また、重要経済安保情報が記録された行政文書は 8 機関計 43 件保有されており、国立公文書館等への移管件数 0 件、重要経済安保情報行政文書ファイル等の廃棄件数 0 件。通報窓口に寄せられた違反行為に関する通報件数 0 件。更に、適性評価の実施件数は 2 機関・18 件（適合事業者の従業者は 0 件）、適性評価調査を内閣総理大臣に求めた行政機関は 11 機関・217 件、評価対象者が適性評価の実施に同意しなかった件数は 3 件、同意が取り下げられた件数は 2 件であった。

内閣府：国会への報告 https://www.cao.go.jp/keizai_anzen_hosho/kokkaihoukoku/kokkaihoukoku.html [2026/7/17 確認]

※ 35 経済産業省：「経済安全保障に関する産業・技術基盤強化アクションプラン再改訂版」を公表しました <https://www.meti.go.jp/press/2025/05/20250530010/20250530010.html> [2026/7/17 確認]

※ 36 経済産業省：経済安全保障に関する産業・技術基盤強化アクション

ンプラン再改訂の概要 https://www.meti.go.jp/policy/economy/economic_security/actionplangaiyor.pdf [2026/7/17 確認]

※ 37 経済産業省:「経済安全保障経営ガイドライン(第1版)」を取りまとめた <https://www.meti.go.jp/press/2025/01/20260123004/20260123004.html> [2026/7/17 確認]

※ 38 経済産業省:「経済安全保障グローバルフォーラム・ウィークス」 https://www.meti.go.jp/policy/economy/economic_security/gfw/index.html [2026/7/17 確認]

※ 39 <https://www.meti.go.jp/files/900018891.pdf> [2026/7/17 確認]

※ 40 BBC: Google achieves AI 'breakthrough' by beating Go champion <https://www.bbc.com/news/technology-35420579> [2026/7/10 確認]

CNET: AlphaGo wins a close one to wrap up battle of man vs. machine <https://www.cnet.com/tech/tech-industry/google-alpha-go-artificial-intelligence-victor-in-game-of-man-vs-machine/> [2026/7/10 確認]

日本経済新聞: 囲碁 AI が 3 戦全勝、世界最強プロを圧倒 https://www.nikkei.com/article/DGXLASFG27H6L_X20C17A5000000/ [2026/7/10 確認]

※ 41 CNBC: On ChatGPT's one-year anniversary, it has more than 1.7 billion users—here's what it may do next <https://www.cnbc.com/2023/11/30/chatgpts-one-year-anniversary-how-the-viral-ai-chatbot-has-changed.html> [2026/7/10 確認]

※ 42 IPA: 情報セキュリティ 10 大脅威 2026 <https://www.ipa.go.jp/security/10threats/10threats2026.html> [2026/7/10 確認]

※ 43 総務省、経済産業省による「AI 事業者ガイドライン (第 1.2 版)」(https://www.meti.go.jp/shingikai/mono_info_service/ai_shakai_jisso/pdf/20260331_1.pdf [2026/7/10 確認])の「第 1 部 AI とは」の「関連する用語」から引用。

※ 44 Jared Kaplan et al.: Scaling Laws for Neural Language Models <https://arxiv.org/abs/2001.08361> [2026/7/10 確認]

※ 45 Emily M. Bender et al.: On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? <https://dl.acm.org/doi/10.1145/3442188.3445922> [2026/7/10 確認]

※ 46 International Business Machines Corporation: 推論モデルとは <https://www.ibm.com/jp-ja/think/topics/reasoning-model> [2026/7/10 確認]

※ 47 Jason Wei et al.: Chain-of-Thought Prompting Elicits Reasoning in Large Language Models <https://proceedings.neurips.cc/paper/2022/hash/9d5609613524ecf4f15af0f7b31abc4-Abstract-Conference.html> [2026/7/10 確認]

※ 48 OpenAI 社: OpenAI o1-preview が登場 <https://openai.com/ja-JP/index/introducing-openai-o1-preview/> [2026/7/10 確認]

Google 社: Gemini 2.5 Flash で開発を始める <https://developers.googleblog.com/ja/start-building-with-gemini-25-flash/> [2026/7/10 確認]

Anthropic 社: Claude 3.7 Sonnet and Claude Code <https://www.anthropic.com/news/claude-3-7-sonnet> [2026/7/10 確認]

※ 49 遠藤聡志 (LifePrompt, Inc.): 【首席超え・満点続出】2026 年度東大・京大の入試問題を AI (ChatGPT / Gemini / Claude) に解かせてみた【採点協力: 河合塾】 <https://note.com/lifeprompt/n/n85674c186fbc> [2026/7/10 確認]

※ 50 Google 社: Image editing in Gemini just got a major upgrade <https://blog.google/products-and-platforms/products/gemini/updated-image-editing-model/> [2026/7/10 確認]

※ 51 OpenAI 社: Sora 2 が公開 <https://openai.com/ja-JP/index/sora-2/> [2026/7/10 確認]

※ 52 Google 社: Introducing Veo 3.1 and advanced capabilities in Flow <https://blog.google/innovation-and-ai/products/veo-updates-flow/> [2026/7/10 確認]

※ 53 Google 社: Gemini Live に動画と画面共有機能が追加 <https://blog.google/intl/ja-jp/company-news/technology/2025gemini-live/> [2026/7/10 確認]

※ 54 Microsoft 社: Copilot on Windows: Vision Desktop Share begins rolling out to Windows Insiders <https://blogs.windows.com/windows-insider/2025/07/15/copilot-on-windows-vision-desktop-share-begins-rolling-out-to-windows-insiders/> [2026/7/10 確認]

※ 55 AI エージェントよりも包括的かつ進化的な概念として複数の AI エージェントによるエージェンティック AI もある。

※ 56 OpenAI 社: ChatGPT エージェントが登場: 研究とアクションをつなぐ新たな架け橋 <https://openai.com/ja-JP/index/introducing-chatgpt-agent/> [2026/7/10 確認]

※ 57 Microsoft 社: Build 2025: Agents in Microsoft 365 announcements <https://techcommunity.microsoft.com/blog/>

microsoft365copilotblog/build-2025-agents-in-microsoft-365-announcements/4414281 [2026/7/10 確認]

※ 58 Anthropic 社: The work behind your best work <https://claude.com/ja/product/cowork> [2026/7/10 確認]

※ 59 GitHub, Inc.: GitHub Copilot: 新しいコーディングエージェント <https://github.blog/jp/2025-05-20-github-copilot-meet-the-new-coding-agent/> [2026/7/10 確認]

※ 60 Anthropic 社: Claude 3.7 Sonnet and Claude Code <https://www.anthropic.com/news/claude-3-7-sonnet> [2026/7/10 確認]

※ 61 https://www.meti.go.jp/shingikai/mono_info_service/ai_shakai_jisso/20260331_report.html [2026/7/10 確認]

※ 62 https://www.meti.go.jp/shingikai/mono_info_service/ai_shakai_jisso/pdf/20260331_3.pdf [2026/7/10 確認]

※ 63 朝日新聞: 津波警報の発表中にグーグル検索、AI が「すべて解除」と誤情報 <https://www.asahi.com/articles/ASTD933Y2TD90XIE008M.html> [2026/7/10 確認]

※ 64 日本航空株式会社: メールサービスの誤った翻訳・要約に関するお問い合わせについて <https://www.jal.co.jp/jp/ja/info/2025/other/250711/> [2026/7/10 確認]

※ 65 Tom's Hardware: Google's Agentic AI wipes user's entire HDD without permission in catastrophic failure — cache wipe turns into mass deletion event as agent apologizes: "I am absolutely devastated to hear this. I cannot express how sorry I am" <https://www.tomshardware.com/tech-industry/artificial-intelligence/googles-agentic-ai-wipes-users-entire-hard-drive-without-permission-after-misinterpreting-instructions-to-clear-a-cache-i-am-deeply-deeply-sorry-this-is-a-critical-failure-on-my-part> [2026/7/10 確認]

※ 66 The Economic Times: AI goes rogue: Replit coding tool deletes entire company database, creates fake data for 4,000 users <https://economictimes.indiatimes.com/news/new-updates/ai-goes-rogue-replit-coding-tool-deletes-entire-company-database-creates-fake-data-for-4000-users/articleshow/122830424.cms> [2026/7/10 確認]

※ 67 Proskauer: AI Bias Lawsuit Against Workday Reaches Next Stage as Court Grants Conditional Certification of ADEA Claim <https://www.lawandtheworkplace.com/2025/06/ai-bias-lawsuit-against-workday-reaches-next-stage-as-court-grants-conditional-certification-of-adea-claim/> [2026/7/10 確認]

※ 68 Becker's Payer: Judge orders UnitedHealth to hand over documents in AI coverage denial case <https://www.beckerspayers.com/legal/judge-orders-unitedhealth-to-hand-over-broad-discovery-in-ai-coverage-denial-case/> [2026/7/10 確認]

※ 69 CNN.co.jp: 会計担当が38億円を詐欺グループに送金、ビデオ会議のCFOは偽物 香港 <https://www.cnn.co.jp/world/35214839.html> [2026/7/10 確認]

※ 70 この事例は「情報セキュリティ白書 2025」の「2.1.4 (2) (c) AI を用いた認知領域への攻撃」においても採り上げたが、金額規模や話題性から改めて言及した。

※ 71 Proofpoint, Inc.: AI で社長の音声を学習した日本語によるボイスフィッシング <https://www.proofpoint.com/jp/email-and-cloud-threats/voice-phishing-with-AI-learning-ceos> [2026/7/10 確認]

※ 72 読売新聞オンライン: IT企業のオンライン採用面接、生成AIで他人になりすまし…実在人物の写真や履歴書悪用か <https://www.yomiuri.co.jp/national/20260319-GYT1T00020/> [2026/7/10 確認]

※ 73 FBI: North Korean IT Worker Threats to U.S. Businesses <https://www.fbi.gov/investigate/cyber/alerts/2025/north-korean-it-worker-threats-to-u-s-businesses> [2026/7/10 確認]

※ 74 Validin, LLC: "Hello? I can't hear you": Investigating UNC1069's Fake Meeting Tactics https://www.validin.com/blog/i_cant_hear_you_unc1069/ [2026/7/10 確認]

※ 75 警察庁: 令和 7 年におけるサイバー空間をめぐる脅威の情勢等について https://www.npa.go.jp/bureau/cyber/pdf/R07_cyber_jousei.pdf [2026/7/10 確認]

※ 76 読売新聞オンライン: 生成AI作成のわいせつポスターをネットで販売か、男女4人を逮捕…「手軽に稼げコスパよかった」 <https://www.yomiuri.co.jp/national/20250415-OYT1T50070/> [2026/7/10 確認]

※ 77 読売新聞オンライン: 卒業アルバム加工した偽の性的画像SNS拡散、一部は小中高生らが作成…警察庁がAIサイト調査 <https://www.yomiuri.co.jp/national/20250831-OYT1T50010/> [2026/7/10 確認]

※ 78 Forbes: Pedophiles Are Using AI To Turn Children's Social Media Photos Into CSAM <https://www.forbes.com/sites/thomasbrewster/2025/04/08/pedophiles-use-ai-to-turn-kids-social-media-photos-into-csam/> [2026/7/10 確認]

※ 79 Robinson+Cole: Lawsuit Alleges Fireflies.AI Corp. Illegally

Collects Biometric Data from Virtual Meetings <https://www.dataprivacyandsecurityinsider.com/2025/12/lawsuit-alleges-fireflies-ai-corp-illegally-collects-biometric-data-from-virtual-meetings/> [2026/7/10 確認]

※ 80 Fisher Phillips: Job Applicants Sue AI Screening Company for FCRA Violations: 5 Key Takeaways for Employers <https://www.fisherphillips.com/en/insights/insights/job-applicants-sue-ai-screening-company-for-fcra-violations> [2026/7/10 確認]

※ 81 出典に当時のレートでの日本円換算の記載がある場合を除いて、本白書では1ドル160円として換算している。

※ 82 Reuters: US judge upholds \$243 million verdict against Tesla over fatal Autopilot crash <https://www.reuters.com/world/us-judge-upholds-243-million-verdict-against-tesla-over-fatal-autopilot-crash-2026-02-20/> [2026/7/10 確認]

※ 83 ITmedia: ChatGPTの医療アドバイスで薬を過剰摂取、米国で19歳大学生が死亡 遺族がOpenAI提訴 <https://www.itmedia.co.jp/news/articles/260514/news054.html> [2026/7/10 確認]

※ 84 産経新聞:「チャット GPT との対話影響」で米の16歳自殺 両親がオープン AIなどに損害賠償請求 <https://www.sankei.com/article/20250828-SY34TZAYN5MVJ52YOHEG2HQP4/> [2026/7/10 確認]

※ 85 産経新聞:「チャット GPT」との対話に没頭した男、母親を殺害して自殺 被害妄想が悪化と米紙報道 <https://www.sankei.com/article/20250829-CXPVOBSXXBLHNPDCY47IELDKTY/> [2026/7/10 確認]

※ 86 Reuters: フロリダ州銃乱射の遺族がオープンAI提訴、「チャットGPTが犯行助長」 <https://jp.reuters.com/world/us/Q764EE00GNO-PBOX3TIJ42DLLXU-2026-05-12/> [2026/7/10 確認]

※ 87 CNET Japan: 仕事でAIを使いすぎると「自分の能力」に自信が持てなくなる? 米心理学会が報告 <https://japan.cnet.com/article/35246689/> [2026/7/10 確認]

※ 88 CNET Japan: ディズニーとユニバーサル、AI画像めぐりMidjourneyを提訴 -- 「盗作の底なし穴」と非難 <https://japan.cnet.com/article/35234163/> [2026/7/10 確認]

※ 89 <https://aja.gr.jp/wp-content/uploads/2025/10/30269151d7cdc6e9bcaeaace2a517c33.pdf> [2026/7/10 確認]

※ 90 OpenAI 社: Sora の提供終了について知っておくべきこと <https://help.openai.com/ja-jp/articles/20001152-what-to-know-about-the-sora-discontinuation> [2026/7/10 確認]

※ 91 ASCII: バイトダンス、性能おぼけの動画生成 AI「Seedance 2.0」展開停止 著作権問題で <https://ascii.jp/elem/000/004/381/4381362/> [2026/7/10 確認]

※ 92 日経新聞: 日経・朝日「AI 検索で著作権侵害」、米新興例は争う姿勢 地裁で初弁論 <https://www.nikkei.com/article/DGXZQOUD0715FOX00C26A5000000/> [2026/7/10 確認]

※ 93 CNN.co.jp: 最新ChatGPTであふれる「ジブリ風」AI画像、威力示す一方で著作権問題も再燃 <https://www.cnn.co.jp/style/arts/35231124.html> [2026/7/10 確認]

※ 94 文化庁: 文化審議会 著作権分科会 法制度小委員会「AIと著作権に関する考え方について」【概要】 https://www.bunka.go.jp/seisaku/bunkashingikai/chosakuken/pdf/94057901_01.pdf [2026/7/10 確認]

※ 95 WIRED: Anthropic、AI 著作権訴訟の和解で少なくとも15億ドルを支払いへ <https://wired.jp/article/anthropic-settlement-lawsuit-copyright/> [2026/7/10 確認]

※ 96 文化庁: AIと著作権について <https://www.bunka.go.jp/seisaku/chosakuken/aicopyright.html> [2026/7/10 確認]

※ 97 Courthouse News Service: Supreme Court denies appeal in AI-generated art case <https://courthousenews.com/supreme-court-denies-appeal-in-ai-generated-art-case/> [2026/7/10 確認]

※ 98 読売新聞オンライン: 生成AIで作った画像を無断複製の疑い、27歳男を書類送検…「著作権あり」として初の摘発 <https://www.yomiuri.co.jp/national/20251120-OYT1T50104/> [2026/7/10 確認]

※ 99 The San Francisco Standard: 'White-collar recession': Benioff says Salesforce won't hire engineers this year due to AI <https://sfstandard.com/2025/02/27/salesforce-marchbenioff-layoffs-tech-agents/> [2026/7/10 確認]

※ 100 GeekWire: 'It's culture': Amazon CEO says massive corporate layoffs were about agility — not AI or cost-cutting <https://www.geekwire.com/2025/its-culture-amazon-ceo-says-massive-corporate-layoffs-were-about-agility-not-ai-or-cost-cutting/> [2026/7/10 確認]

※ 101 読売新聞オンライン: 生成されたAI女優「ティリー・ノーウッド」に複数タレント事務所が関心…俳優らが懸念表明 <https://www.yomiuri.co.jp/world/20251003-OYT1T50108/> [2026/7/10 確認]

※ 102 Merriam-Webster, Inc.: 2025 Word of the Year: Slop <https://www.merriam-webster.com/wordplay/word-of-the-year> [2026/7/10 確認]

※ 103 東洋経済:「これは営業じゃない! 業務妨害だ」 迷惑な「AI 営業」に中小企業の社長が激怒したワケ《背景にAIの利用拡大》送信先企業にもたらされる3つの問題 <https://toyokeizai.net/articles/-/870844?display=b> [2026/7/10 確認]

※ 104 読売新聞オンライン: 生成 AI が作った偽「オスズメ本」をアメリカの新聞社がそのまま掲載、購読者から怒りの声が相次ぐ <https://www.yomiuri.co.jp/world/nieman/20250522-OYT1T50160/> [2026/7/10 確認]

※ 105 読売新聞オンライン: 共同通信、生成AIで加工の「子ガメをくわえるタヌキ」の写真を配信取り消し…加盟社の指摘で判明 <https://www.yomiuri.co.jp/national/20251102-OYT1T50089/> [2026/7/10 確認]

※ 106 Kobi Hackenburg et al.: The levers of political persuasion with conversational artificial intelligence <https://www.science.org/doi/abs/10.1126/science.aea3884> [2026/7/10 確認]

※ 107 Dhruv Agarwal et al.: AI Suggestions Homogenize Writing Toward Western Styles and Diminish Cultural Nuances <https://dl.acm.org/doi/10.1145/3706598.3713564> [2026/7/10 確認]

※ 108 IEA: Key Questions on Energy and AI <https://www.iea.org/reports/key-questions-on-energy-and-ai> [2026/7/10 確認]

※ 109 IEA: AI and climate change <https://www.iea.org/reports/energy-and-ai/ai-and-climate-change> [2026/7/10 確認]

※ 110 IEA: The Netherlands <https://www.iea.org/countries/the-netherlands/emissions> [2026/7/10 確認]

※ 111 Microsoft 社: 2025 Environmental Sustainability Report <https://cdn-dynmedia-1.microsoft.com/is/content/microsoftcorp/microsoft/msc/documents/presentations/CSR/2025-Microsoft-Environmental-Sustainability-Report.pdf#page=04> [2026/7/10 確認]

※ 112 GOV.UK: AI Safety Summit 2023 <https://www.gov.uk/government/topical-events/ai-safety-summit-2023>

※ 113 総務省・経済産業省による「AI 事業者ガイドライン(第1.2版)」の「第2部 AIにより目指すべき社会及び各主体が取り組む事項」の「C. 共通の指針」を参照されたい。

※ 114 日本 AISI: AI セーフティに関する評価観点ガイド(第1.10版) 2.2 AI セーフティにおける重要要素 https://aisi.go.jp/assets/pdf/ai-safety_eval_v1.10_ja.pdf [2026/7/10 確認]

※ 115 日本 AISI: AI セーフティに関する評価観点ガイド(第1.10版)の公開 https://aisi.go.jp/output/output_information/250328_1/ [2026/7/10 確認]

※ 116 日本 AISI: AI セーフティ評価のための評価ツールを OSS として公開 https://aisi.go.jp/output/output_information/250912/ [2026/7/10 確認]

※ 117 日本 AISI: 事業実証 WG の取組について https://aisi.go.jp/assets/pdf/20251003_2.pdf [2026/7/10 確認]

※ 118 日本 AISI: AI セーフティ年次レポート 2025 https://aisi.go.jp/assets/pdf/j-aisi_report_2025_ja.pdf [2026/7/10 確認]

※ 119 英国 AISI: International consensus and open questions in AI evaluations <https://www.aisi.gov.uk/blog/international-ai-network-consensus-and-open-questions> [2026/7/10 確認]

※ 120 例えば、ISO 31000:2018(JIS Q 31000:2019)が挙げられる。

※ 121 総務省、経済産業省: AI 事業者ガイドラインの令和7年度更新内容 https://www.soumu.go.jp/main_content/001059300.pdf [2026/7/10 確認]

※ 122 https://aisi.go.jp/output/output_framework/260317/ [2026/7/10 確認]

※ 123 広島 AI プロセス: 報告枠組み <https://www.soumu.go.jp/hiroshimaaiprocess/report.html> [2026/7/10 確認]

※ 124 広島 AI プロセス: フレンズグループ <https://www.soumu.go.jp/hiroshimaaiprocess/supporters.html> [2026/7/10 確認]

※ 125 一般財団法人日本規格協会: JIS Q 42001:2025 情報技術—人工知能—マネジメントシステム https://webdesk.jisa.or.jp/books/W11M0090/index/?bunsyo_id=JIS+Q+42001%3A2025 [2026/7/10 確認]

※ 126 一般財団法人日本規格協会: ISO/IEC 42006:2025 情報技術—人工知能—人工知能マネジメントシステムの審査及び認証を行う機関に対する要求事項 https://webdesk.jisa.or.jp/books/W11M0090/index/?bunsyo_id=ISO%2FIEC+42006%3A2025 [2026/7/10 確認]

※ 127 一般財団法人日本規格協会: ISO/IEC 42005:2025 情報技術—人工知能 (AI) — AI システムの影響評価 https://webdesk.jisa.or.jp/books/W11M0090/index/?bunsyo_id=ISO%2FIEC+42005%3A2025 [2026/7/10 確認]

※ 128 日本 AISI: 適合性評価サブワーキンググループ (適合性評価 SWG) 2025 年度活動報告書 <https://aisi.go.jp/assets/pdf/>

conformity_assessment_swg_report_ja.pdf [2026/7/10 確認]

※ 129 Federal Register: 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> [2026/7/10 確認]

※ 130 The White House: Removing Barriers to American Leadership in Artificial Intelligence <https://www.whitehouse.gov/presidential-actions/2025/01/removing-barriers-to-american-leadership-in-artificial-intelligence/> [2026/7/10 確認]

※ 131 The White House: America's AI Action Plan <https://www.whitehouse.gov/wp-content/uploads/2025/07/Americas-AI-Action-Plan.pdf> [2026/7/10 確認]

※ 132 U.S. Department of Commerce: Statement from U.S. Secretary of Commerce Howard Lutnick on Transforming the U.S. AI Safety Institute into the Pro-Innovation, Pro-Science U.S. Center for AI Standards and Innovation <https://www.commerce.gov/news/press-releases/2025/06/statement-us-secretary-commerce-howard-lutnick-transforming-us-ai> [2026/7/10 確認]

※ 133 EU: Regulation (EU) 2024/1689 <https://eur-lex.europa.eu/eli/reg/2024/1689/oj> [2026/7/10 確認]

※ 134 JETRO: EU の AI 規則、禁止される AI 利用のルールが適用開始に <https://www.jetro.go.jp/biznews/2025/02/af0786d0eca9e961.html> [2026/7/10 確認]

※ 135 European Commission: General-Purpose AI Code of Practice now available https://ec.europa.eu/commission/presscorner/detail/en/ip_25_1787 [2026/7/10 確認]

※ 136 European Commission: Simpler EU digital rules and new digital wallets to save billions for businesses and boost innovation https://ec.europa.eu/commission/presscorner/detail/en/ip_25_2718 [2026/7/10 確認]

※ 137 GOV.UK: AI Opportunities Action Plan <https://www.gov.uk/government/publications/ai-opportunities-action-plan/ai-opportunities-action-plan> [2026/7/10 確認]

※ 138 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> [2026/7/10 確認]

※ 139 国立研究開発法人新エネルギー・産業技術総合開発機構: 次世代人工知能発展計画 (仮訳) <https://www.nedo.go.jp/content/100903937.pdf> [2026/7/10 確認]

※ 140 Merics: China's drive toward self-reliance in artificial intelligence: from chips to large language models <https://merics.org/en/report/chinas-drive-toward-self-reliance-artificial-intelligence-chips-large-language-models> [2026/7/10 確認]

※ 141 JETRO: 中国、「AI プラス」行動の徹底に関する意見発表、6 重点分野で実施強化 <https://www.jetro.go.jp/biznews/2025/09/8431c1ceb01ef8e8.html> [2026/7/10 確認]

時事通信社: 【中国】政府、主要産業にAI融合戦略 行動指針発表 <https://www.jiji.com/jc/article?k=2025090100225> [2026/7/10 確認]

※ 142 Bit & Company (消費者庁委託調査): 令和6年度海外主要国における消費者保護に関する生成AI関連の規制等の調査業務 報告書 第8章 中国 https://www.caa.go.jp/policies/future/national_research/assets/caa_futurer101_250605_10.pdf [2026/7/10 確認]

※ 143 人民網日本語版: 中国で9月1日からAIで生成・合成したコンテンツへのラベル付け義務化へ <https://j.people.com.cn/n3/2025/0321/c94475-20292366.html> [2026/7/10 確認]

※ 144 内閣府: 人工知能関連技術の研究開発及び活用の推進に関する法律 (AI法) https://www8.cao.go.jp/cstp/ai/ai_act/ai_act.html [2026/7/10 確認]

※ 145 https://www8.cao.go.jp/cstp/ai/ai_plan/ai_plan.html [2026/7/10 確認]

※ 146 デジタル庁: 全府省庁の約 18 万人の政府職員を対象としたガバメント AI (源内) の大規模実証を開始します <https://www.digital.go.jp/news/2d69c287-2897-46d8-a28f-ea5a1fc9bce9> [2026/7/10 確認]

※ 147 内閣府: 人工知能基本計画 第3章 第3節 AI ガバナンスの主導 https://www8.cao.go.jp/cstp/ai/ai_plan/aiplan_20251223.pdf [2026/7/10 確認]

※ 148トレンドマイクロ株式会社: TrendAI AIセキュリティレポート 2026 ~ AI エコシステムの断層 ~ https://www.trendmicro.com/ja_jp/research/26/c/fault-lines-in-the-ai-ecosystem-trendai-state-of-ai-security-report.html [2026/7/10 確認]

※ 149 OWASP Foundation, Inc.: OWASP Top 10 for LLM Applications 2025 LLM01:2025 Prompt Injection <https://genai.owasp.org/resource/owasp-top-10-for-llm-applications-2025/>

[2026/7/10 確認]

※ 150 PC Watch: 「メールを送るだけ」で AI 悪用攻撃可な Microsoft 365 Copilot の脆弱性。緊急で対策済み <https://pc.watch.impress.co.jp/docs/news/2022498.html> [2026/7/10 確認]

※ 151 Check Point Software Technologies Ltd.: AI Evasion: The Next Frontier of Malware Techniques <https://blog.checkpoint.com/artificial-intelligence/ai-evasion-the-next-frontier-of-malware-techniques/> [2026/7/10 確認]

※ 152 日本経済新聞: 論文内に秘密の命令文、AI に「高評価せよ」日韓米など有力 14 大学で <https://www.nikkei.com/article/DGXZQOUC13BCW0T10C25A600000/> [2026/7/10 確認]

※ 153 ITmedia: 「ブラウザを自律操作する AI」が暴走 「要約して」と入力しただけなのにパスワードが盗まれた その手口とは? <https://www.itmedia.co.jp/news/articles/2509/03/news027.html> [2026/7/10 確認]

※ 154 PromptArmor: Claude Cowork Exfiltrates Files <https://www.promptarmor.com/resources/claude-cowork-exfiltrates-files> [2026/7/10 確認]

※ 155 OWASP Foundation, Inc.: Code Injection https://owasp.org/www-community/attacks/Code_Injection [2026/7/10 確認]

※ 156 Python Software Foundation: pickle --- Python オブジェクトの直列化 <https://docs.python.org/ja/3/library/pickle.html> [2026/7/10 確認]

※ 157 Red Hat, LLC: CVE-2025-49655 <https://access.redhat.com/security/cve/cve-2025-49655> [2026/7/10 確認]

※ 158 Sonatype, Inc.: CVE-2025-1716 <https://www.sonatype.com/security-advisories/cve-2025-1716> [2026/7/10 確認]

JFrog Ltd: PyTorch Users at Risk: Unveiling 3 Zero-Day PickleScan Vulnerabilities <https://jfrog.com/blog/unveiling-3-zero-day-vulnerabilities-in-picklescan/> [2026/7/10 確認]

SECDIM: CVE-2025-46417: Bypassing AI Model Scanners and Exfiltrate Sensitive Data <https://secdim.com/blog/post/cve-2025-46417-bypassing-ai-model-scanners-and-exfiltrate-sensitive-data-15594/> [2026/7/10 確認]

※ 159 Palo Alto Networks, Inc.: 最新の AI/ML フォーマットとライブラリによるリモート コード実行 <https://unit42.paloaltonetworks.com/ja/rce-vulnerabilities-in-ai-python-libraries/> [2026/7/10 確認]

※ 160 International Business Machines Corporation: 知識蒸留とは <https://www.ibm.com/jp-ja/think/topics/knowledge-distillation> [2026/7/10 確認]

※ 161 Anthropic 社: Usage Policy <https://www.anthropic.com/legal/aup> [2026/7/10 確認]

Google 社: Gemini API 追加利用規約 <https://ai.google.dev/gemini-api/terms> [2026/7/10 確認]

OpenAI 社: 利用規約 <https://openai.com/ja-JP/policies/row-terms-of-use/> [2026/7/10 確認]

※ 162 Google 社: GTIG AI Threat Tracker: Distillation, Experimentation, and (Continued) Integration of AI for Adversarial Use <https://cloud.google.com/blog/topics/threat-intelligence/distillation-experimentation-integration-ai-adversarial-use> [2026/7/10 確認]

Anthropic 社: Detecting and preventing distillation attacks <https://www.anthropic.com/news/detecting-and-preventing-distillation-attacks> [2026/7/10 確認]

※ 163 Office of Science and Technology Policy: Adversarial Distillation of American AI Models <https://whitehouse.gov/wp-content/uploads/2026/04/NSTM-4.pdf> [2026/7/10 確認]

※ 164 CNN: White House accuses China of copying American AI models in 'industrial-scale' campaign <https://edition.cnn.com/2026/04/23/tech/white-house-accuses-china-copying-american-ai> [2026/7/10 確認]

※ 165 Socket, Inc.: The Rise of Slopsquatting: How AI Hallucinations Are Fueling a New Class of Supply Chain Attacks <https://socket.dev/blog/slopsquatting-how-ai-hallucinations-are-fueling-a-new-class-of-supply-chain-attacks> [2026/7/10 確認]

※ 166 Joseph Spracklen et al.: We Have a Package for You! A Comprehensive Analysis of Package Hallucinations by Code Generating LLMs <https://www.usenix.org/conference/usenixsecurity25/presentation/spracklen> [2026/7/10 確認]

※ 167 Koi Security: PhantomRaven: NPM Malware Hidden in Invisible Dependencies <https://www.koi.ai/blog/phantomraven-npm-malware-hidden-in-invisible-dependencies> [2026/7/10 確認]

※ 168 https://aisi.go.jp/output/output_security/known_attacks_and_impacts/ [2026/7/10 確認]

※ 169 https://aisi.go.jp/assets/pdf/known_attacks_and_their_impacts_on_ai_systems_v2_jp.pdf [2026/7/10 確認]

※ 170 The Hacker News: Critical LangChain Core Vulnerability Exposes Secrets via Serialization Injection <https://thehackernews.com/2025/12/critical-langchain-core-vulnerability.html> [2026/7/10 確認]

※ 171 Cymulate Ltd.: EscapeRoute: Breaking the Scope of Anthropic's Filesystem MCP Server (CVE-2025-53109 & CVE-2025-53110) <https://cymulate.com/blog/cve-2025-53109-53110-escaperooute-anthropic/> [2026/7/10 確認]

※ 172 JPCERT/CC: JVN#99749963 CrewAI における複数の脆弱性 <https://jvn.jp/vu/JVN#99749963/> [2026/7/10 確認]

※ 173 Cyber Security News: 15,200 OpenClaw Control Panels with Full System Access Exposed to the Internet <https://cybersecuritynews.com/openclaw-control-panels-exposed/> [2026/7/10 確認]

※ 174 SafeDep Inc.: Malicious litellm 1.82.8: Credential Theft and Persistent Backdoor <https://safedep.io/malicious-litellm-1-82-8-analysis/> [2026/7/10 確認]

※ 175 英国 AISI: How fast is autonomous AI cyber capability advancing? <https://www.aisi.gov.uk/blog/how-fast-is-autonomous-ai-cyber-capability-advancing/> [2026/7/10 確認]

※ 176 NCSC: Impact of AI on cyber threat from now to 2027 <https://www.ncsc.gov.uk/report/impact-ai-cyber-threat-now-2027> [2026/7/10 確認]

※ 177 Proofpoint, Inc.: サイバー犯罪者が AI ウェブサイト作成アプリをフィッシングに悪用 <https://www.proofpoint.com/jp/blog/threat-insight/cybercriminals-abuse-ai-website-creation-app-phishing> [2026/7/10 確認]

※ 178 Palo Alto Networks, Inc.: AI のデュアルユースのジレンマ: 悪意ある LLM <https://unit42.paloaltonetworks.com/ja/dilemma-of-ai-malicious-llm/> [2026/7/10 確認]

※ 179 読売新聞オンライン: 「快活CLUB」サイバー攻撃、AI悪用し会員情報盗んだ疑いで高2を再逮捕へ…自作プログラムを「チャットGPT」で改善か <https://www.yomiuri.co.jp/national/20251204-GYT1T00037/> [2026/7/10 確認]

※ 180 Anthropic 社: Disrupting the first reported AI-orchestrated cyber espionage campaign <https://www.anthropic.com/news/disrupting-AI-espionage> [2026/7/10 確認]

※ 181 株式会社ラック: 大規模な AI 主導サイバー攻撃の初の報告書、アンスロピックの AI 攻撃報告が巻き起こした論争 https://www.lac.co.jp/lacwatch/report/20260202_004628.html [2026/7/10 確認]

※ 182 Amazon Web Services, Inc.: AI で強化された脅威アクターによる FortiGate デバイスへの大規模な不正アクセス <https://aws.amazon.com/jp/blogs/news/ai-augmented-threat-actor-accesses-fortigate-devices-at-scale/> [2026/7/10 確認]

※ 183 Microsoft 社: AI as tradecraft: How threat actors operationalize AI <https://www.microsoft.com/en-us/security/blog/2026/03/06/ai-as-tradecraft-how-threat-actors-operationalize-ai/> [2026/7/10 確認]

※ 184 Anthropic 社: Assessing Claude Mythos Preview's cybersecurity capabilities <https://red.anthropic.com/2026/mythos-preview/> [2026/7/10 確認]

※ 185 Anthropic 社: Project Glasswing <https://www.anthropic.com/glasswing> [2026/7/10 確認]

※ 186 OpenAI 社: GPT-5.5 が登場 <https://openai.com/ja-JP/index/introducing-gpt-5-5/> [2026/7/10 確認]

※ 187 Google 社: GTIG AI Threat Tracker: Advances in Threat Actor Usage of AI Tools <https://cloud.google.com/blog/topics/threat-intelligence/threat-actor-usage-of-ai-tools> [2026/7/10 確認]

※ 188 Microsoft 社: SesameOp: Novel backdoor uses OpenAI Assistants API for command and control <https://www.microsoft.com/en-us/security/blog/2025/11/03/sesameop-novel-backdoor-uses-openai-assistants-api-for-command-and-control/> [2026/7/10 確認]

※ 189 Mozilla: The zero-days are numbered <https://blog.mozilla.org/en/firefox/ai-security-zero-day-vulnerabilities/> [2026/7/10 確認]

※ 190 Tenable, Inc.: 「AI のためのセキュリティ」と「セキュリティのための AI」の比較 <https://jp.tenable.com/cybersecurity-guide/learn/security-for-ai-vs-ai-for-security> [2026/7/10 確認]

トレンドマイクロ株式会社: AI セキュリティとは何か? (後編) ~ Security for AI と AI for Security ~ https://www.trendmicro.com/ja_jp/jp-security/25/d/expertview-20250410-01.html#2 [2026/7/10 確認]

※ 191 Bleeping Computer: Curl ending bug bounty program after flood of AI slop reports <https://www.bleepingcomputer.com/news/security/curl-ending-bug-bounty-program-after-flood-of-ai-slop-reports> [2026/7/10 確認]

※ 192 ITmedia: AI 生成の「ゴミ報告」が殺到、対応追い付かず疲弊……脆弱性発見の懸賞金制度に異変 <https://www.itmedia.co.jp/news/articles/2605/11/news021.html> [2026/7/10 確認]

※ 193 NIST: NIST Updates NVD Operations to Address Record CVE Growth <https://www.nist.gov/news-events/news/2026/04/nist-updates-nvd-operations-address-record-cve-growth> [2026/7/10 確認]

※ 194 YATA: 「Yielding Advanced Threat Awareness with AI (脅威の可視化)」の頭文字。「正確に写す」という八咫鏡 (やたのかがみ) もあるように、「AI性能の高度化に伴うサイバー脅威を正しく認識し、防御する」対応する」という趣旨。

※ 195 NCO: AI 性能の高度化を踏まえたサイバーセキュリティ対策に関する関係省庁会議 https://www.cyber.go.jp/pdf/press/20260518_AI_CS_kankeishouchoukaigi.pdf [2026/7/10 確認]

※ 196 https://www.cyber.go.jp/pdf/council/cs/cs_expert/07/07document6.pdf [2026/7/10 確認]

※ 197 NCO、内閣府政策統括官 (経済安全保障担当)、警察庁、金融庁、総務省、厚生労働省、経済産業省、国土交通省、防衛省: AI 性能の高度化を踏まえたサイバーセキュリティ対策の強化について (重要インフラ事業者等に対する注意喚起) https://www.cyber.go.jp/pdf/press/20260518_AI_CS_Critical_Infrastructure.pdf [2026/7/10 確認]

※ 198 金融庁: 「AI 脅威に対する金融分野のサイバーセキュリティ対策強化に関する官民連携会議」の作業部会の開催について <https://www.fsa.go.jp/news/r7/sonota/20260514/20260514.html> [2026/7/10 確認]

※ 199 経済産業省: 高性能 AI への対応に関して赤澤経済産業大臣が重要インフラ事業者との意見交換を実施しました <https://www.meti.go.jp/press/2026/05/20260501001/20260501001.html> [2026/7/10 確認]

※ 200 NCO: AI 性能の高度化を踏まえたサイバーセキュリティ対策の強化について https://www.cyber.go.jp/pdf/press/20260518_AI_CS_government.pdf [2026/7/10 確認]

※ 201 NCO、経済産業省: AI 性能の高度化を踏まえたサイバーセキュリティ対策の強化について (ソフトウェア・ベンダに対する注意喚起) https://www.cyber.go.jp/pdf/press/20260518_AI_CS_software_vendor.pdf [2026/7/10 確認]

※ 202 総務省: 「AI のセキュリティ確保のための技術的対策に係るガイドライン」(案) に対する意見募集の結果及びガイドラインの公表 https://www.soumu.go.jp/menu_news/s-news/01cyber01_02000001_00282.html [2026/7/10 確認]

※ 203 日本 AISI: AI セーフティに関するレッドチーミング手法ガイド (第 1.10 版) の公開 https://aisi.go.jp/output/output_information/250331_1/ [2026/7/10 確認]

※ 204 日本 AISI: 自動レッドチーミング 説明書 <https://github.com/Japan-AISI/aisev/blob/main/docs/rt.md> [2026/7/10 確認]

※ 205 日本 AISI: AI インシデントレスポンス・アプローチブックを公開 https://aisi.go.jp/activity/activity_security/260109/ [2026/7/10 確認]

※ 206 NCSC: Preparing for a 'vulnerability patch wave' <https://www.ncsc.gov.uk/blogs/prepare-for-vulnerability-patch-wave> [2026/7/10 確認]

※ 207 NCSC: Vulnerability management <https://www.ncsc.gov.uk/collection/vulnerability-management> [2026/7/10 確認]

※ 208 一般社団法人 AI ガバナンス協会: 「『いつの間にか』AI のリスク実態調査 ~シャドウ AI 等の管理と捕捉 課題とプラクティス~」を公表しました <https://www.ai-governance.jp/blog/250129-policy-opinion-ai-regulation> [2026/7/10 確認]

※ 209 株式会社 KDDI 総合研究所: AI セキュリティポータル <https://aisecurity-portal.org/> [2026/7/10 確認]

※ 210 <https://www.ipa.go.jp/digital/ai/security/ai-security-bulletin.html> [2026/7/10 確認]

※ 211 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> [2026/8/13 確認]

※ 212 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 月号「新領域安全保障」、p.185

※ 213 <https://rm.coe.int/information-disorder-toward-an-interdisciplinaryframework-for-research/168076277c> [2026/8/13 確認]

※ 214 一般社団法人セーフティ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 [2026/8/13 確認]

※ 215 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 [2026/8/13 確認]

※ 216 https://reports.weforum.org/docs/WEF_Global_Risks_Report_2026.pdf [2026/8/13 確認]

※ 217 WEF: The Global Risks Report 2024 https://www3.weforum.org/docs/WEF_The_Global_Risks_Report_2024.pdf [2026/8/13 確認]

※ 218 WEF: The Global Risks Report 2025 https://reports.weforum.org/docs/WEF_Global_Risks_Report_2025.pdf [2026/8/13 確認]

※ 219 <https://reutersinstitute.politics.ox.ac.uk/digital-news-report/2025> [2026/8/13 確認]

※ 220 Reuters Institute for the Study of Journalism: Overview and key findings of the 2025 Digital News Report <https://reutersinstitute.politics.ox.ac.uk/digital-news-report/2025/dnr-executive-summary> [2026/8/13 確認]

※ 221 Meta 社: Code of Conduct on Disinformation – Report of Meta for the period 01 January to 30 June 2025 <https://disinfocode.eu/reports/download/147> [2026/8/13 確認]

※ 222 Meta 社: Community Standards Enforcement Report: Fake Accounts <https://transparency.meta.com/reports/community-standards-enforcement/fake-accounts/facebook/> [2026/8/13 確認]

※ 223 OpenAI 社: Disrupting malicious uses of AI: an update <https://cdn.openai.com/pdf/d438d70-e3fe-4a6c-a403-ff632def8f79/disrupting-malicious-uses-of-ai.pdf> [2026/8/13 確認]

※ 224 OpenAI 社: Disrupting malicious uses of our models <https://cdn.openai.com/pdf/df438d70-e3fe-4a6c-a403-ff632def8f79/disrupting-malicious-uses-of-ai.pdf> [2026/8/13 確認]

※ 225 U.S. NAVAL INSTITUTE: Future Warfare: The Rise of Hybrid Wars <https://www.usni.org/magazines/proceedings/2005/november/future-warfare-rise-hybrid-wars> [2026/8/13 確認]

※ 226 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

※ 227 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 [2026/8/13 確認]

※ 228 European Commission: A strengthened EU Code of Practice on Disinformation https://commission.europa.eu/topics/countering-information-manipulation/strengthened-eu-code-practice-disinformation_en [2026/9/8 確認]

※ 229 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 [2026/8/13 確認]

※ 230 Hybrid CoE: Hybrid threats as a concept <https://www.hybridcoe.fi/hybrid-threats-as-a-phenomenon/> [2026/8/13 確認]

※ 231 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> [2026/8/13 確認]

※ 232 EEAS: 2nd EEAS Report on Foreign Information Manipulation and Interference Threats <https://euvsdisinfo.eu/uploads/2026/02/EEAS-2nd-Report-on-FIMI-Threats-January-2024.pdf> [2026/8/13 確認]

EEAS: 3rd EEAS Report on Foreign Information Manipulation and Interference Threats <https://www.eeas.europa.eu/sites/default/files/documents/2025/EEAS-3rd-ThreatReport-March-2025-05-Digital-HD.pdf> [2026/8/13 確認]

※ 233 GTIG: Pro-Russia Information Operations Leverage Russian Drone Incursions into Polish Airspace <https://cloud.google.com/blog/topics/threat-intelligence/pro-russia-information-operations-drone-incursions> [2026/8/13 確認]

※ 234 The Guardian: How social media lies fuelled a rush to war between India and Pakistan <https://www.theguardian.com/media/2025/may/28/how-social-media-lies-fuelled-a-rush-to-war-between-india-and-pakistan> [2026/8/13 確認]

※ 235 MODERN DIPLOMACY: Cambodia-Thailand Border Conflict: Restraint in the Digital Information Battlefield <https://modern diplomacy.eu/2026/01/14/cambodia-thailand-border-conflict-restraint-in-the-digital-information-battlefield/> [2026/8/13

確認]

※ 236 International Democracy Union: IDU Statement on Russia's Interference in Moldova's Elections <https://idu.org/idu-statement-on-russias-interference-in-moldovas-elections/> [2026/8/13 確認]

※ 237 European Commission: Commission services and Moldovan authorities conduct a stress test on potential digital hybrid threats to election integrity ahead of Moldova's parliamentary elections <https://digital-strategy.ec.europa.eu/en/news/commission-services-and-moldovan-authorities-conduct-stress-test-potential-digital-hybrid-threats> [2026/8/13 確認]

※ 238 OSCE: Moldova's parliamentary elections were competitive but campaign marred by cyberattacks, illegal funding and disinformation, international observers say <https://odihr.osce.org/odihr/598219> [2026/8/13 確認]

※ 239 ドイツ連邦憲法擁護庁: Illegitimate foreign influence operations targeting the upcoming German federal election <https://www.verfassungsschutz.de/SharedDocs/kurzmeldungen/EN/2025/2025-01-02-german-federal-election.html> [2026/9/8 確認]

※ 240 The Record: Germany summons Russian ambassador over cyberattack, election disinformation <https://therecord.media/germany-summons-russian-ambassador-cyberattack-disinformation> [2026/8/13 確認]

※ 241 ポーランド共和国デジタル化省: Raport z programu “Parasol Wyborczy” - kompleksowa ochrona wyborów prezydenckich 2025 <https://www.gov.pl/web/cyfrizacja/raport-z-programu-parasol-wyborczy---kompleksowa-ochrona-wyborow-prezydenckich-2025> [2026/9/8 確認]

※ 242 BBC: Israel-Iran conflict unleashes wave of AI disinformation <https://www.bbc.com/news/articles/c0k78715enxo> [2026/8/13 確認]

※ 243 Ministry for Diaspora Affairs and Combating Antisemitism: Coordinated Iranian Activity on Social Media Through Fake Accounts During Rising Lion https://www.gov.il/BlobFolder/reports/coordinated_iranian_fake_accounts_social_media_rising_lion_2025-06/en/Research_Reports_coordinated_iranian_fake_accounts_social_media_rising_lion_2025-06.pdf [2026/8/13 確認]

※ 244 EDMO: The First AI War: How the Iran-Israel Conflict Became a Battlefield for Generative Misinformation <https://edmo.eu/publications/the-first-ai-war-how-the-iran-israel-conflict-became-a-battlefield-for-generative-misinformation/> [2026/8/13 確認]

※ 245 Clemson University, Watt Family Innovation Center Media Forensics Hub: From Texas to Tehran: A Multilingual, IRGC-affiliated Influence Operation on X, Instagram, and Bluesky https://policycommons.net/artifacts/48154021/from-texas-to-tehran_-a-multilingual-irgc-affiliated-influence-o/49052774/ [2026/8/13 確認]

※ 246 PolitiFact: Videos do not show missile strike on USS Abraham Lincoln <https://politifact.com/factchecks/2026/mar/02/social-media/uss-abraham-lincoln-fire-missile-iran/> [2026/8/13 確認]

※ 247 The Jerusalem Post: Iranian media publishes false conspiracy theory claiming Netanyahu's death or injury <https://www.jpost.com/middle-east/iran-news/article-889415> [2026/8/13 確認]

NHK ONE: イラン軍事衝突 フェイク3億回表示…目的は?どう見分けられ

ば? <https://news.web.nhk/newsweb/na/na-k10015091231000> [2026/8/20 確認]

※ 248 PolitiFact: Trump says Iran is a country 'based on disinformation.' Here's how Iran conducts influence ops <https://www.politifact.com/article/2026/mar/23/iran-disinformation-campaigns-influence-operations/> [2026/8/13 確認]

※ 249 U.S. Department of War: Trump Announces U.S. Military's Capture of Maduro <https://www.war.gov/News/News-Stories/Article/Article/4370431/trump-announces-us-militarys-capture-of-maduro/> [2026/8/13 確認]

※ 250 CEDMO: AI, outdated visuals fuel misinformation after Maduro capture <https://cedmohub.eu/ai-outdated-visuals-fuel-misinformation-after-maduro-capture/> [2026/8/13 確認]

※ 251 CNBC: AI-generated content spreads after Maduro's removal — blurring fact and fiction <https://www.cnbc.com/2026/01/06/ai-generated-deepfake-videos-venezuelan-viral-us-military-maduro-misinformation.html> [2026/8/13 確認]

※ 252 WHO: Infodemic <https://www.who.int/health-topics/infodemic> [2026/8/13 確認]

※ 253 WHO: Measles cases dropped in Europe and Central Asia

in 2025 compared to the previous year, but the risk of outbreaks remains – UNICEF and WHO <https://www.who.int/europe/news/item/11-02-2026-measles-cases-dropped-in-europe-and-central-asia-in-2025-compared-to-the-previous-year-but-the-risk-of-outbreaks-remains---unicef-and-who> [2026/8/13 確認]

※ 254 JFC：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/> [2026/8/13 確認]

※ 255 JFC：新型コロナのmRNAワクチンで死亡リスクが74%上昇？根拠とされた論文の結論は正反対【ファクトチェック】 <https://www.factcheckcenter.jp/fact-check/health/false-replicon-vaccine-claim/> [2026/8/13 確認]

※ 256 European Commission：The Code of Conduct on Disinformation <https://digital-strategy.ec.europa.eu/en/library/code-conduct-disinformation> [2026/8/13 確認]

※ 257 European Commission：Transparency and targeting of political advertising https://commission.europa.eu/strategy-and-policy/policies/justice-and-fundamental-rights/democracy-eu-citizenship-anti-corruption/democracy-and-electoral-rights/transparency-and-targeting-political-advertising_en [2026/8/13 確認]

※ 258 Meta 社：Ending Political, Electoral and Social Issue Advertising in the EU in Response to Incoming European Regulation <https://about.fb.com/news/2025/07/ending-political-electoral-and-social-issue-advertising-in-the-eu/> [2026/8/13 確認]

※ 259 European Commission：Code of Conduct on Disinformation: Signatories publish their latest reports in the Code's Transparency Centre <https://digital-strategy.ec.europa.eu/en/news/code-conduct-disinformation-signatories-publish-their-latest-reports-codes-transparency-0> [2026/8/13 確認]

※ 260 Meta 社：Report March 2026 <https://disinfocode.eu/reports/facebook/8/text> [2026/8/13 確認]

※ 261 Euronews：TikTok made changes ahead of re-run of Romanian election, says EU tech chief <https://www.euronews.com/next/2025/04/08/tiktok-made-changes-ahead-of-re-run-of-romanian-election-says-eu-tech-chief> [2026/8/13 確認]

※ 262 Ballotpedia News：Forty-seven states have enacted deepfake legislation since 2019 <https://news.ballotpedia.org/2025/07/22/forty-seven-states-have-enacted-deepfake-legislation-since-2019/> [2026/8/13 確認]

※ 263 Australian Government：Online Safety Amendment (Social Media Minimum Age) Act 2024 <https://www.legislation.gov.au/C2024A00127/asmade/text> [2026/8/13 確認]

※ 264 Minister for Communications, the Hon Anika Wells MP：Social media minimum age platform assessments <https://minister.infrastructure.gov.au/wells/media-release/social-media-minimum-age-platform-assessments> [2026/8/13 確認]

eSafety：Which platforms are age-restricted? <https://www.esafety.gov.au/about-us/industry-regulation/social-media-age-restrictions/which-platforms-are-age-restricted> [2026/8/13 確認]

※ 265 New Zealand Parliament：Parliamentary inquiry into online harm and young people completed <https://www3.parliament.nz/mi/pb/sc/committees-press-releases/parliamentary-inquiry-into-online-harm-and-young-people-completed/> [2026/8/13 確認]

※ 266 U.S. News：Factbox – From Australia to Europe, Countries Move to Curb Children's Social Media Access <https://www.usnews.com/news/world/articles/2026-04-24/factbox-from-australia-to-europe-countries-move-to-curb-childrens-social-media-access> [2026/8/13 確認]

※ 267 interface：Social Media Age Gating in the EU <https://www.interface-eu.org/publications/social-media-age-gating-eu> [2026/8/13 確認]

※ 268 Internet for Kids：Social Media Age Limits and Child Protection Laws – Country Tracker <https://internetforkids.org/en/social-media-age-limits-tracker/> [2026/8/13 確認]

※ 269 総務省：インターネット上の違法・有害情報に対する対応（情報流通プラットフォーム対処法） https://www.soumu.go.jp/main_sosiki/joho_tsusin/d_syohi/ihoyugai.html [2026/8/13 確認]

※ 270 総務省：インターネット上の偽・誤情報等への対策技術の開発・実証事業（令和7年度） https://www.soumu.go.jp/main_sosiki/joho_tsusin/d_syohi/taisakugijutsu_fy2025.html [2026/8/13 確認]

※ 271 総務省：「インターネット上の偽・誤情報等への対策技術の開発・実証事業成果発信イベント」の開催 https://www.soumu.go.jp/menu_news/s-news/01ryutsu02_02000469.html [2026/8/13 確認]

※ 272 国立研究開発法人新エネルギー・産業技術総合開発機構：「経済安全保障重要技術育成プログラム」で偽情報の分析に係る技術の開発に着手 https://www.nedo.go.jp/news/press/AA5_101763.html [2026/9/1 確認]

※ 273 JFC：参院選ファクトチェック計27本 / 政治系人気動画の解説も【今週のファクトチェック】 <https://www.factcheckcenter.jp/fact-check/others/fact-check-weekly-20250720/> [2026/8/13 確認]

※ 274 朝日新聞：(Media Times) 参院選、期間中にも各社積極報道ファクトチェックは昨年衆院選の5倍超 <https://www.asahi.com/articles/DA3S16283986.html> [2026/8/13 確認]

※ 275 Meta 社：Report September 2025 <https://disinfocode.eu/reports/meta/6> [2026/8/13 確認]

※ 276 TikTok 社：Protecting the integrity of TikTok during the Romanian elections <https://newsroom.tiktok.com/en-eu/protecting-the-integrity-of-tiktok-during-the-romanian-elections> [2026/8/13 確認]

※ 277 X 社：Authenticity <https://help.x.com/en/rules-and-policies/authenticity> [2026/8/13 確認]

※ 278 BusinessDay：X rolls out 'Manipulated Media' labels to combat misleading images, sparking questions on scope, enforcement <https://businessday.ng/technology/article/x-rolls-out-manipulated-media-labels-to-combat-misleading-images-sparking-questions-on-scope-enforcement/> [2026/8/13 確認]

※ 279 Yahoo! ニュース：Xの投稿に「AIで生成」ラベルを設定可能に偽情報対策になるか <https://news.yahoo.co.jp/expert/articles/3187acaca3fecb60f49fd13902bf3e6955bd87ec> [2026/8/13 確認]

※ 280 Oversight Board：Board to Address AI-Generated Content in Israel-Iran Conflict <https://www.oversightboard.com/news/board-to-address-ai-generated-content-in-israel-iran-conflict/> [2026/8/13 確認]

※ 281 OpenAI 社：Disrupting malicious uses of AI – an update (October 2025) <https://openai.com/global-affairs/disrupting-malicious-uses-of-ai-october-2025/> [2026/8/13 確認]

OpenAI 社：AIの悪用を阻止：2025年10月 <https://openai.com/ja-JP/global-affairs/disrupting-malicious-uses-of-ai-october-2025/> [2026/8/13 確認]

※ 282 Bloomberg：X Will Deploy AI to Write Community Notes, Expand Fact-Checking <https://www.bloomberg.com/news/articles/2025-07-01/x-will-deploy-ai-to-write-community-notes-expand-fact-checking> [2026/8/13 確認]

※ 283 NPR：Meta says it will end fact-checking as Silicon Valley prepares for Trump <https://www.npr.org/2025/01/07/nx-s1-5251151/meta-fact-checking-mark-zuckerberg-trump> [2026/8/13 確認]

※ 284 Center for Countering Digital Hate Inc：Rated Not Helpful – How X's Community Notes system falls short on election disinformation <https://counterhate.com/wp-content/uploads/2024/10/CCDH.CommunityNotes.FINAL-30.10.pdf> [2026/8/13 確認]

※ 285 Regina Cazzamatta：From moderation to chaos: Meta's fact-checking and the battle over truth and free speech (New Media & Society, SAGE Publications) <https://journals.sagepub.com/doi/10.1177/14614448251413687> [2026/8/13 確認]

※ 286 Roozenbeek, J., van der Linden, S., Goldberg, B., Rathje, S., Lewandowsky, S.: Psychological inoculation improves resilience against misinformation on social media (Science Advances, Vol.8, No.34) <https://www.science.org/doi/10.1126/sciadv.abo6254> [2026/8/13 確認]

※ 287 TIME：Inside Google's Plans to Combat Misinformation Ahead of the E.U. Elections <https://time.com/6970488/google-jigsaw-eu-elections-misinformation-prebunking/> [2026/8/13 確認]

※ 288 Mikey Biddlestone et al.: Video inoculation against election misinformation across 12 EU nations (Communications Psychology) <https://www.nature.com/articles/s44271-025-00379-3> [2026/8/13 確認]

※ 289 総務省：「国民を詐欺から守るための総合対策」に係る総務省の取組状況について https://www.cao.go.jp/consumer/iinkai/2024/439/doc/20240708_shiryoku1-3.pdf [2026/8/13 確認]

※ 290 総務省：総合的なICTリテラシー向上に向けた官民連携プロジェクト「DIGITAL POSITIVE ACTION」の開始 https://www.soumu.go.jp/menu_news/s-news/01ryutsu05_02000169.html [2026/8/20 確認]

確認]

※ 291 防衛省・自衛隊：防衛大臣記者会見 <https://www.mod.go.jp/j/press/kisha/2026/0703a.html>〔2026/8/13 確認〕

第2章

サイバー空間を巡る動向

付録・索引

第21回 IPA

「ひろげよう情報セキュリティ
コンクール」2025 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2,184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します。

最優秀賞

油断禁物



長浜 ちひろさん
立川女子高等学校

優秀賞

そのパスワード大丈夫？



脇 玲美奈さん
静岡市立安東小学校

視線の先



小磯 凜さん
静岡県立三島南高等学校

作り分けよう
使い分けようパスワード



宮城 優佳さん
白百合学園中学校

数字

OAPT	21
2024 年オンライン安全修正(ソーシャルメディア最低年齢)法(Online Safety Amendment (Social Media Minimum Age) Act 2024)	94
8Base	110

A

Active Directory	34
AI(Artificial Intelligence: 人工知能)	9, 69, 89, 110, 122, 160
AI エージェント	70, 80, 82
AI ガバナンス	76
AI 事業者ガイドライン	70, 76
AI システム	77, 81
AI セーフティ	76
AI セーフティ・インスティテュート(AISI: AI Safety Institute)	76, 79, 85
AI セーフティサミット	76, 162
AI の悪用	71, 82, 90
AI のセキュリティ	79, 111
AI のセキュリティ確保のための技術的対策に係るガイドライン	85, 111
AI のリスク	70
AI 法	78, 110
AI モデル	69, 80
Akira	9, 172
America's AI Action Plan	77
ANEL	28
APCERT(Asia Pacific Computer Emergency Response Team: アジア太平洋コンピュータ緊急対応チーム)	184, 186
APT29	173
APT(Advanced Persistent Threat) 攻撃	27, 120
ArcaneDoor	38
ASM(Attack Surface Management) 導入ガイド	33
Attack Surface Management(ASM)	24, 33

C

C&C(Command and Control) サーバー	27, 35
-------------------------------	--------

CCRA(Common Criteria Recognition Arrangement)	191
ChatGPT	69, 72, 90
CIOP(CIOp、Clop)	9, 38
Claude Mythos	83, 161
ClickFix	33, 46, 50
CRYPTREC(Cryptography Research and Evaluation Committees)	142
CSIRT(Computer Security Incident Response Team)	31, 41, 123, 176, 183
CVE(Common Vulnerabilities and Exposures: 共通脆弱性識別子)	38, 79, 176
CVSS(Common Vulnerability Scoring System: 共通脆弱性評価システム)	40, 176

D

DDoS 攻撃	10, 17, 34, 172
DIGITAL POSITIVE ACTION	95
DNS(Domain Name System)	10, 36, 41, 177
DRDoS(Distributed Reflection Denial of Service) 攻撃	17

E

Earth Kasha	28
EchoLeak	80
EDR(Endpoint Detection and Response)	22, 33, 38
ENISA(The European Union Agency for Cybersecurity)	8, 128, 164, 172, 175
ERAB サイバーセキュリティトレーニング	130
EU AI Act	77, 173
EUCC(European Common Criteria-based cybersecurity certification scheme)	180
EUVD(European Vulnerability Database)	176
EU サイバーセキュリティ法(CSA: Cybersecurity Act)	175

F

FIMI(Foreign Information Manipulation and Interference: 外国による情報操作と干渉)	91
FOG	9

G

General-Purpose AI Code of Practice(汎用 AI の行動規範)	78
Global Cybersecurity Camp(GCC)	128

I

IEC(International Electrotechnical Commission : 国際電気標準会議)	188
IEEE(The Institute of Electrical and Electronics Engineers, Inc.)	188
IETF(Internet Engineering Task Force)	143, 188
International AI Safety Report 2026	162
International Cybersecurity Challenge(ICC)	128
IoC(Indicator of Compromise : 侵害指標)	25, 31, 119
IoT	35, 37, 117, 133, 163
ISMAP-LIU(イスマップ・エルアイユー : ISMAP for Low-Impact Use)	140
ISMAP 管理基準	139
ISMAP 等クラウドサービスリスト	139
ISO(International Organization for Standardization : 国際標準化機構)	188
ISO/IEC 15408(Common Criteria)	138, 191
ISO/IEC 27000 ファミリー	189
ISO/IEC JTC 1/SC 27	116, 189
ITU-T(International Telecommunication Union Telecommunication Standardization Sector : 国際電気通信連合 電気通信標準化部門) ..	188
IT 製品の調達におけるセキュリティ要件リスト ..	138
IT セキュリティ評価及び認証制度(JISEC : Japan Information Technology Security Evaluation and Certification Scheme)	138

J・K

J-CRAT(Cyber Rescue and Advice Team against targeted attack of Japan : サイバーレスキュー隊)	120
JTC 1(Joint Technical Committee 1 : 第一合同技術委員会)	188

KEV(Known Exploited Vulnerabilities : 既知の悪用された脆弱性) カタログ	38, 41, 170
---	-------------

L

LAMEHUG	160
Living Off The Land(LOTL) 戦術	28, 161
LLM(Large Language Model : 大規模言語モデル)	69, 75, 79, 82

M

Microsoft Office	30
MirrorFace	28

N・O

NIS 2 指令(Network and Information Systems Directive 2)	13, 174, 177
NoName057(16)	34
NOOPDOOR	29
NVD(National Vulnerability Database : 国家脆弱性データベース)	41, 83
N デイ攻撃	39
Operational Relay Box(ORB : 中継装置) ..	28, 29

P・Q

PhaaS(Phishing as a Service)	11, 172
PhantomRaven	81
Phobos	110
Project YATA-Shield	84, 123
PromptLock	160
PSTI 法(Product Security and Telecommunications Infrastructure Act)	117, 136, 163
Qilin	9, 21, 172

R

RaaS(Ransomware as a Service)	19, 21
Ransomhouse	23

S

SaaS	11, 140
SafePay	172
Salt Typhoon	29, 40, 161

SBOM (Software Bill of Materials : ソフトウェア 部品表)	117, 137, 169
Scattered Lapsus\$ Hunters	30
SECURITY ACTION	116, 145, 150, 152
slop (スロップ)	74, 81
SMS	15, 46
Spamouflage	90

T・U

TCG (Trusted Computing Group)	189
The NIST Cybersecurity Framework (CSF) 2.0	116, 170
TS 5689	116, 191
U.S. Cyber Trust Mark	137, 163, 170

V

Volt Typhoon	28, 161
VPN	16, 20, 23, 27, 119

あ

アイデンティティ管理	191
アイランドホッピング攻撃	31
悪意ある情報	88, 90
アクセス・無害化	63, 107, 109
暗号資産	162
一般財団法人日本サイバー犯罪対策センター (JC3 : Japan Cybercrime Control Center)	109
一般社団法人 JPCERT コーディネーションセンター (JPCERT/CC : Japan Computer Emergency Response Team Coordination Center)	26, 38, 41, 121, 186
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	164
インフォスティーラー (Infostealer)	9, 46, 50
エクスプロイト	9, 12
エネルギー・リソース・アグリゲーション・ビジネスに 関するサイバーセキュリティガイドライン	130, 136
遠隔操作	27, 49, 52
遠隔操作マルウェア	23
欧州刑事警察機構 (Europol : European Union Agency for Law Enforcement Cooperation)	34, 110, 164, 176

欧州サイバーセキュリティ認証フレームワーク (ECCF : European Cybersecurity Certification Framework)	176, 180
オンライン安全法 (Online Safety Act)	94

か

技術情報管理認証制度	118
技術流出対策ガイダンス	68
業界別サイバーレジリエンス強化演習 (CyberREX : Cyber Resilience Enhancement eXercise by industry)	129
クラウドサービス	22, 25, 139, 190
クレジットカード	47, 66
グローバル・サイバーセキュリティ・ラベリング・イニ シアティブ (GCLI : Global Cybersecurity Labeling Initiative)	137
経済安全保障経営ガイドライン	68
経済安全保障重要技術育成プログラム (通称、K Program)	86, 94, 118
経済安全保障推進法	66
経済安全保障に関する産業・技術基盤強化アクション プラン	67
公開鍵暗号	141, 143
攻撃対象領域 (アタックサーフェス)	24, 33, 38, 81
工場システムにおけるサイバー・フィジカル・セキュリ ティ対策ガイドライン	117
公的個人認証サービス (JPKI : Japanese Public Key Infrastructure)	111
コードインジェクション	80
国立研究開発法人情報通信研究機構 (NICT : National Institute of Information and Communications Technology)	17, 66, 111
国連サイバー犯罪防止条約 (UNCC : United Nations Convention against Cybercrime)	161
誤情報	75, 88
国家安全保障戦略	62, 106
国家サイバー統括室 (NCO : National Cybersecurity Office)	27, 63, 107
国家支援型 APT 攻撃	28, 120
コンピュータ不正アクセス対策基準	44

さ

サイバー安全保障分野での対応能力の向上に向けた提言	62
サイバーインフラ事業者に求められる役割等に関するガイドライン	117
サイバー空間を巡る脅威に対応するため喫緊に取り組むべき事項	62, 106
サイバー警察局	110
サイバー情報共有イニシアティブ(J-CSIP: Initiative for Cyber Security Information sharing Partnership of Japan)	119
サイバーセキュリティ2025(2024年度年次報告・2025年度年次計画)	106, 123
サイバーセキュリティお助け隊サービス	116, 147, 151
サイバーセキュリティ企画演習(CyberSPEX: Cyber Security Planning Exercise)	129
サイバーセキュリティ基本法	106, 112
サイバーセキュリティ経営ガイドライン	32
サイバーセキュリティ人材	118, 122, 148
サイバーセキュリティ人材フレームワーク2026	124
サイバーセキュリティ戦略	62, 106
サイバーセキュリティネクサス(CYNEX: Cybersecurity Nexus)	17
サイバー対処能力強化法	62, 106
サイバー特別捜査部	110
サイバー・フィジカル・セキュリティ対策フレームワーク(CPSF)	115, 191
サイバーブループリント(Cyber Blueprint)	179
サイバーレジリエンス法(CRA: Cyber Resilience Act)	174, 178
サイバー連帯法(CSoA: Cyber Solidarity Act)	179
サプライチェーン	8, 24, 31, 115, 145, 160
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(SCS 評価制度)	116, 145, 148, 150
サプライチェーン・サイバーセキュリティ・コンソーシアム(SC3: Supply-Chain Cybersecurity Consortium)	116
サプライチェーン・リスク	8, 84, 177
サポート詐欺	49, 52, 71, 110
産業サイバーセキュリティ研究会	114

産業サイバーセキュリティセンター(ICSCoE: Industrial Cyber Security Center of Excellence)	128, 164
ジェイルブレイク	80, 169
事業継続計画(BCP: Business Continuity Plan)	23, 26, 31, 36, 129
次世代アンチウイルス(NGAV: Next Generation Anti-Virus)	38
実践的サイバー防御演習(CYDER: Cyber Defense Exercise with Recurrence)	165
シャドー AI	85, 123
重要インフラ	34, 63, 84, 108, 167, 177
重要インフラのサイバーセキュリティ対策のための統一基準(案)	108
重要インフラのサイバーセキュリティに係る行動計画	108
重要経済安保情報保護活用法	66
重要電子計算機に対する特定不正行為による被害の防止のための基本的な方針	64
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(サイバー対処能力強化法)	62
情報処理安全確保支援士(登録セキスベ)	124, 125, 151
情報セキュリティ安心相談窓口	46
情報セキュリティ早期警戒パートナーシップ	41, 120
情報セキュリティマネジメント試験	125
情報セキュリティマネジメントシステム(ISMS: Information Security Management System)	189
情報戦	90, 95
情報操作型サイバー攻撃	88, 90
情報流通プラットフォーム対処法	94
情報漏えい	11, 15, 42
新型コロナウイルス	93
人工知能関連技術の研究開発及び活用の推進に関する法律(AI 法)	78, 110
人工知能基本計画	79
侵入型ランサム攻撃	19, 21, 23
水平展開	25, 27
スロップスクワッティング	81
制御システム	129, 164

制御システム向けサイバーセキュリティ演習 (CyberSTIX : Cyber Security practical eXercise for industrial control system)	130	データ三法	181
脆弱性	8, 24, 37, 121, 170, 176	データ侵害	9, 11, 172
生成 AI (Generative AI)	69, 92, 110	データセキュリティ	111
政府機関等のサイバーセキュリティ対策のための統一基準	138	適合ラベル	133
政府機関等の対策基準策定のためのガイドライン	27, 117, 135	デジタルオペレーショナルレジリエンス法 (DORA : Digital Operational Resilience Act)	174
政府情報システムのためのセキュリティ評価制度 (Information system Security Management and Assessment Program : 通称、ISMAP (イ スマップ))	139	デジタルオムニバス規則案	173
セキュア・ソフトウェア開発フレームワーク (SSDF : Secure Software Development Framework)	117	デジタルサービス法 (DSA : Digital Services Act)	93, 174
セキュア・バイ・デザイン	32, 117	デジタル社会の実現に向けた重点計画	111
セキュリティ・キャンプ	118, 127	デジタル署名	175, 190
セキュリティ・クリアランス制度	66	デジタルスキル標準 (DSS)	124
セキュリティ要件適合評価及びラベリング制度 (JC- STAR)	32, 117, 133, 163	デジタルトラスト	165
ゼロデイ攻撃	37	デバイス・コード・フィッシング (Device Code Phishing)	9
ゼロトラスト	84, 166	特別社会基盤事業者	63
組織における内部不正防止ガイドライン	44		
		な	
た		内閣サイバーセキュリティセンター (NISC : National center of Incident readiness and Strategy for Cybersecurity)	62
ダークウェブ	20, 26, 37	内部不正	32, 42
第 14 次 5 ヵ年計画	180	ナラティブ	90, 95
対ランサムウェア・イニシアチブ (CRI : Counter Ransomware Initiative)	162	なりすまし	42, 71, 162
耐量子計算機暗号 (PQC : Post-Quantum Cryptography)	142, 168	二重の脅迫 (二重恐喝)	16, 19, 23
地方公共団体における情報セキュリティポリシーに関 するガイドライン	113	偽・誤情報	75, 88
中核人材育成プログラム	128	偽情報	88
中華人民共和国サイバーセキュリティ法	181, 183	偽のウイルス感染警告	49, 52, 110
中小企業の情報セキュリティ対策ガイドライン	116, 147	日 ASEAN サイバーセキュリティ政策会議	117, 164
中小企業のためのセキュリティ人材確保・育成の実 践ガイドブック	118, 150	日 ASEAN サイバーセキュリティ能力構築センター (AJCCBC : ASEAN-Japan Cybersecurity Capacity Building Centre)	164
ディープフェイク	92, 94, 110, 160	日 ASEAN 能力向上プログラム強化プロジェクト	165
ディスインフォメーション (Dis-information)	88	日・EU サイバー対話	164
		日英サイバー対話	163
		日米サイバー対話	163
		日本 AISI	76, 79, 81, 85
		日本産業標準調査会 (JISC : Japanese Industrial Standards Committee)	188
		認知戦	88, 91, 95
		ネットワーク貫通型攻撃	27, 29, 119
		ノーウェアランサム	19, 23

は

ハーベスト攻撃(HNDL 攻撃 : Harvest Now, Decrypt Later Attack)	142
ハイブリッド型サイバー攻撃	88
ハイブリッド脅威	91
ハイブリッド戦争	90
ハクティビスト	10, 12, 34, 173
バックドア	29, 161
ばらまき型の攻撃	19
ハルシネーション	71, 75, 81
汎用 AI (General-Purpose AI)	78, 175
標的型攻撃	20, 27, 119
広島 AI プロセス	77
ファクトチェック	91, 94
フィッシング	11, 14, 46, 110, 172
フェイクニュース	89
不正アクセス	16, 22, 27, 39, 42
不正送金	15, 49, 52, 133
不正ログイン	11, 46, 47
プロテクションプロファイル	138
フロンティア AI	84
プロンプトインジェクション	79, 85, 169
米国国立標準技術研究所 (NIST : National Institute of Standards and Technology)	24, 83, 143, 170
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)	24, 38, 117, 168
米国連邦捜査局 (FBI : Federal Bureau of Investigation)	12, 71, 111, 166
ボイスフィッシング (ビッシング)	14, 30
ボットネット	35, 92

ま

マイクロセグメンテーション	25
マルインフォメーション (Mal-information)	88
ミスインフォメーション (Mis-information)	88
無人航空機システム (UAS : Unmanned Aircraft Systems)	170
モデル抽出	80, 169

ら

ランサムウェア	9, 16, 19, 110, 162
ランサム攻撃	9, 16, 19, 37, 110, 148, 172
リアルタイムフィッシング	47
リークサイト	20, 24, 26
リスクベースアプローチ	167
リフレクション攻撃	17
リモートデスクトップ	16, 20, 23, 149

著作・製作 独立行政法人情報処理推進機構（IPA）

編集責任 大川 龍郎 清水 督史 小山 明美
涌田 明夫 白石 歩 井上 佳春

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 大海 健太 小川 賢一 加賀谷 伸一郎
勝川 雅義 金山 栄一 金子 成徳 神谷 健司 亀山 友彦
川口 修司 神田 雅透 北爪 亮太郎 儀部 頼之 桐淵 直人
古閑 みなみ 小山 明美 酒井 崇行 榊原 七美 佐藤 みゆき
渋谷 環 白石 歩 関 鈴良 関根 大助 高橋 慎治
武田 泰貴 辻 初実 辻 宏郷 豊田 亮子 長迫 智子
中村 友亮 夏目 典大 新山 雄大 西尾 秀一 芳賀 政伸
平本 健二 福井 章人 松本 穂乃花 室賀 直来 安田 進
籾口 春南 山下 恵一

（敬称略）

ネット教育アナリスト 尾花 紀子

一般社団法人日本クラウドセキュリティアライアンス 諸角 昌宏

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
安藤 康浩 板橋 博之 江島 将和 大川 龍郎 大澤 淳
小川 隆一 沖田 孝裕 奥村 元 小野塚 直人 甲斐 成樹
鎌田 林太郎 釜谷 誠 菊池 秀一 岸野 照明 北村 弘
桑名 利幸 白井 綾 銭谷 謙吾 高田 早苗 高橋 将
高村 博紀 高柳 大輔 土屋 正 中島 尚樹 西原 栄太郎
二村 和明 肱岡 七愛 平原 裕貴 松島 伸彰 松田 修平
宮崎 卓行 谷澤 昭紀

サイバーレスキュー隊 J-CRAT（ジェイ・クラート）

AISI 事務局 戦略企画グループ

（敬称略）

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

京都外国語大学 共通教育機構 土屋 貴裕

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

デジタル庁 戦略・組織グループ セキュリティチーム 中村 元洋

総務省 サイバーセキュリティ統括官室

経済産業省 商務情報政策局 サイバーセキュリティ課

2025年度は、同年5月に成立した「サイバー対処能力強化法及び同整備法」により、能動的サイバー防御を実現する体制の整備が着々と進められ、日本の能動的サイバー防御が実装段階へ移行する転換点となった年であったことから、サブタイトルの前半を「サイバー防御新時代」としました。新時代といえ、その一方でAIの悪用によるサイバー攻撃の脅威が顕在化し、将来的に量子コンピューターによる暗号解読のリスクが懸念されることから、耐量子計算機暗号(PQC)への移行が世界中で本格化する等、「進化」した先端技術が「諸刃の剣」となって我々の喉元に迫ってきた年でもありました。進化を続ける先端技術は、脅威であると同時に、我々の生活を便利にしたり、従来は不可能だった問題解決を実現したりするといった二面性を持つものです。それらに積極的に対応していく必要があるとの考えから、サブタイトルの後半を「諸刃の進化に備えよ」としました。

IPAは、日本の能動的サイバー防御の取り組みの一端を担うとともに、サイバー脅威に対して組織・個人によるセキュリティ対策の一助となるよう、今後とも努めていきます。

編集子

- ・ 本白書の引用、転載については、IPA Web サイトの「書籍・刊行物等に関するよくあるご質問と回答」(<https://www.ipa.go.jp/publish/faq.html>)に掲載されている「2. 引用や転載に関するご質問」をご参照ください。ただし、出典元がIPA 以外であり、かつ IPA が編集、作成を行った図表については、本白書からの転載・改変について IPA は許諾ができません。転載・改変について IPA が許諾できない図表は以下の様に出版を記載しています。

例「(出典)《組織名等》『《文書名等》』を基に IPA が編集」

例「(出典)《組織名等》『《文書名等》』を基に IPA が作成」

また、出典元がIPA 以外であり、かつ IPA が本白書で引用している図表についても、転載・改変について IPA は許諾ができません。以下の様に記載している図表の転載・改変の可否については、出典元をご確認ください。

例「《組織名等》『《文書名等》』」

上記の例にある《組織名等》《文書名等》には実際の出典元組織名、文書名が記載されます。

なお、これは、著作権法で定められた本白書からの引用を妨げるものではありません。

- ・ 本白書は 2025 年度の出来事を主な対象とし、執筆時点の情報に基づいて記載しています。
- ・ 電話によるご質問、及び本白書に記載されている内容以外のご質問には一切お答えできません。あらかじめご了承ください。
- ・ 本白書に記載されている会社名、製品名、及びサービス名は、それぞれ各社の商標または登録商標です。本文中では、TM または® マークは明記していません。
- ・ 本白書に掲載しているグラフ内の数値の合計は、小数点以下の端数処理により、100% にならない場合があります。

情報セキュリティ白書 2026

サイバー防御新時代：諸刃の進化に備えよ

2026 年 10 月 19 日 PDF 版 第 1 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