

# 情報セキュリティ白書

Information Security White Paper

2024

変革の波にひそむ脅威：リスクを見直し対策を



# 「情報セキュリティ白書2024」の刊行にあたって

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

昨今のサイバー空間の動向を振り返ってみると、新型コロナウイルスのパンデミックは収束し、経済・社会活動の回復とともに、働き方改革、デジタル化が大きく進展し、更には生成 AI の登場により変革の兆しが見えます。他方、2022年2月に始まったロシア・ウクライナ戦争の長期化等、現下の厳しい国際情勢下において、重要インフラの機能停止、国民の情報や知的財産の窃取、民主プロセスへの干渉等のサイバー攻撃が顕在化し、サイバー空間が、地政学的緊張を反映した国家間の争いの場の一部ともなっています。今後 AI の悪用によるサイバー攻撃の激化や高度化も懸念されるところです。

国内では、ランサムウェア被害が引き続き多数発生しています。2023年6月の社会保険労務士向けクラウドサービスが被害を受けた事案や、同年7月の港湾コンテナターミナル内のシステム停止をもたらした事案等が発生しました。また、国民情報や知的財産の窃取を目的としたサイバー攻撃も顕在化し、とりわけ、ネットワーク境界の脆弱性を突いた攻撃が多数発生する等、攻撃に一層の巧妙化・高度化が見られます。今後、人手不足解消のための自動化等、デジタルライフラインにおける AI や IoT システムの社会実装が進み、サイバーリスクが、更に増大していくことが予想されます。このようなリスクに対処していくためには、サイバー空間を巡る、変容するリスクを国際的、経済的、地政学的側面から把握・分析し、リスクへの予見性を高めていくこと、そして、サプライチェーンやサイバーやフィジカルが融合した環境を前提として、システムの設計段階から脆弱性を取り除いていく、セキュア・バイ・デザインのアプローチが重要になっています。

各国においては、こうしたサイバー空間を巡る状況変化を踏まえ、セキュリティ対策の見直しが進められています。国内では2023年7月に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群が全面改定、米国でも2024年2月にサイバーセキュリティフレームワーク(CSF)が10年ぶりに大きく改訂され、欧州では2024年の期限に向けて各国がNIS指令及びEUサイバーレジリエンス法案の実装に取り組んでいます。また、AIに関する制度化、ガイドライン等の整備、法制化も進んでいます。2023年12月にはG7において広島AIプロセス包括的政策枠組みが示されました。我が国でも、AIの安全性に対する国際的な関心の高まりを踏まえ、AIの安全性の評価手法の検討等を行う機関として、2024年2月、IPAにAIセーフティ・インスティテュートを設置しました。

本白書は、2023年に生じた事柄を中心に、サイバー空間における脅威や技術の動向、それに対応する内外の政策的対応等について、包括的に記載をしています。本白書が多くの方々にご利用され、サイバーセキュリティに関わる最新状況の把握と、それに伴う脅威やリスクに対する備えを実践するための一助となることを祈念します。

2024年7月

独立行政法人情報処理推進機構(IPA)

理事長 齊藤 裕

序章 2023年度の情報セキュリティの概況	6
第1章 情報セキュリティインシデント・脆弱性の現状と対策	8
1.1 2023年度に観測されたインシデント状況	8
1.1.1 世界における情報セキュリティインシデント状況	8
1.1.2 国内における情報セキュリティインシデント状況	12
1.2 情報セキュリティインシデント別の手口と対策	17
1.2.1 ランサムウェア攻撃	17
1.2.2 標的型攻撃	23
1.2.3 ビジネスメール詐欺(BEC)	28
1.2.4 DDoS攻撃	33
1.2.5 ソフトウェアの脆弱性を悪用した攻撃	36
1.2.6 個人を狙うSMS・メールを悪用した手口	39
1.2.7 個人を狙う様々な騙しと悪用の手口	42
1.2.8 情報漏えいによる被害	48
1.3 情報システムの脆弱性の動向	54
1.3.1 JVN iPediaの登録情報から見る脆弱性の傾向	54
1.3.2 早期警戒パートナーシップの届出状況から見る脆弱性の動向	58
第2章 情報セキュリティを支える基盤の動向	68
2.1 国内の情報セキュリティ政策の状況	68
2.1.1 政府全体の政策動向	68
2.1.2 デジタル庁の政策	74
2.1.3 経済産業省の政策	76
2.1.4 総務省の政策	86
2.1.5 警察によるサイバー空間の安全確保の取り組み	90
2.2 国外の情報セキュリティ政策の状況	97
2.2.1 国際社会と連携した取り組み	97
2.2.2 米国の政策	101
2.2.3 欧州の政策	107
2.2.4 アジア太平洋地域でのCSIRTの動向	112
2.3 情報セキュリティ人材の現状と育成	116
2.3.1 デジタル人材としての情報セキュリティ人材の状況	116
2.3.2 情報セキュリティ人材育成のための国家試験、国家資格制度	119
2.3.3 セキュリティ人材育成のための活動	120

2.4 国際標準化活動	126
2.4.1 様々な標準化団体の活動	126
2.4.2 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)	127
2.4.3 情報通信技術、電気通信に関わるセキュリティ規格の標準化(ITU-T SG17)	135
2.4.4 制御システム関連のセキュリティ規格の標準化(IEC TC 65/WG 10)	137

## 第3章 情報セキュリティ対策強化や取り組みの動向 148

3.1 組織・個人に向けた情報セキュリティ対策の普及活動	148
3.1.1 組織における情報セキュリティの取り組みと支援策	148
3.1.2 情報セキュリティの普及啓発活動	156
3.2 製品・サービス認証制度の動向	159
3.2.1 ITセキュリティ評価及び認証制度	159
3.2.2 暗号モジュール試験及び認証制度	163
3.2.3 政府情報システムのためのセキュリティ評価制度(ISMAP)	163
3.3 暗号技術の動向	167
3.3.1 CRYPTRECの動向	167
3.3.2 暗号関連の技術動向	168
3.4 制御システムのセキュリティ	171
3.4.1 インシデントの発生状況と動向	171
3.4.2 脆弱性及び脅威の動向	173
3.4.3 海外の制御システムのセキュリティ強化の取り組み	175
3.4.4 国内の制御システムのセキュリティ強化の取り組み	177
3.5 IoTのセキュリティ	179
3.5.1 IoTに対するセキュリティ脅威の動向	179
3.5.2 進化を続けるIoTウイルスの動向	183
3.5.3 IoTセキュリティのサプライチェーンとEOLのリスク	186
3.5.4 脆弱なIoT機器のウイルス感染と感染機器悪用の実態	187
3.5.5 各国のセキュリティ対策強化の取り組み	188
3.6 クラウドのセキュリティ	192
3.6.1 クラウドサービスの利用状況	192
3.6.2 クラウドサービスのインシデント事例	193
3.6.3 クラウドサービスのセキュリティの課題と対策	196

第4章 注目のトピック	208
4.1 虚偽を含む情報拡散の脅威と対策の動向	208
4.1.1 虚偽情報とは	208
4.1.2 ディスインフォメーションの生成・拡散の流れ	210
4.1.3 虚偽を含んだ情報生成・拡散の事例	212
4.1.4 虚偽を含んだ情報への対応状況	220
4.1.5 状況のまとめと今後の見通し	222
4.2 AIのセキュリティ	224
4.2.1 本節で対象とするAIのスコープ	224
4.2.2 AIの利用状況と品質特性	224
4.2.3 AIのリスク要因の包括的整理	225
4.2.4 AIのサイバーセキュリティリスク認知状況	227
4.2.5 AIのサイバーセキュリティリスクの分類	230
4.2.6 AIセキュリティ対策の動向	235
4.2.7 まとめ	236
付録 資料	241
資料A 2023年のコンピュータウイルス届出状況	242
資料B 2023年のコンピュータ不正アクセス届出状況	243
資料C ソフトウェア等の脆弱性関連情報に関する届出状況	245
資料D 2023年の情報セキュリティ安心相談窓口の相談状況	248
第19回IPA「ひろげよう情報セキュリティコンクール」2023 受賞作品	250
IPAの便利なツールとコンテンツ	252
索引	257

## コラム

守るだけではない、被害を最小限にするためのセキュリティ対策を	15
情報セキュリティ10大脅威 2024 ～脅威に吞まれる前に十分なセキュリティ対策を～	16
サポート詐欺で人が騙されてしまう心理的要因とその対策	53
デジタル署名が付いたウイルスの広がり	139
「情報セキュリティ監査制度」創設20周年を迎えて	166



# 情報セキュリティ白書

- 序章 2023年度の情報セキュリティの概況
- 第1章 情報セキュリティインシデント・脆弱性の現状と対策
  - 1.1 2023年度に観測されたインシデント状況
  - 1.2 情報セキュリティインシデント別の手口と対策
  - 1.3 情報システムの脆弱性の動向
- 第2章 情報セキュリティを支える基盤の動向
  - 2.1 国内の情報セキュリティ政策の状況
  - 2.2 国外の情報セキュリティ政策の状況
  - 2.3 情報セキュリティ人材の現状と育成
  - 2.4 国際標準化活動
- 第3章 情報セキュリティ対策強化や取り組みの動向
  - 3.1 組織・個人に向けた情報セキュリティ対策の普及活動
  - 3.2 製品・サービス認証制度の動向
  - 3.3 暗号技術の動向
  - 3.4 制御システムのセキュリティ
  - 3.5 IoTのセキュリティ
  - 3.6 クラウドのセキュリティ
- 第4章 注目のトピック
  - 4.1 虚偽を含む情報拡散の脅威と対策の動向
  - 4.2 AIのセキュリティ

# 序章

## 2023年度の情報セキュリティの概況

2023年度は、国内では新型コロナウイルス感染症の5類移行により、停滞していた社会活動や経済活動に活気が戻ってきた。一方で、コロナ禍を一つの契機として業務のデジタル化が進み、事業のIT依存度やシステム・サービス障害による影響が大きくなった。

企業・組織等が受けたサイバー攻撃の件数や被害金額は世界的に増加している。特に、国家の関与が疑われるネットワーク貫通型の攻撃は巧妙かつ執拗で、長期かつ広範囲に及ぶこともあるため深刻な被害を与えている。例えば、「Volt Typhoon」と呼ばれる組織による攻撃は2021年ごろから継続し、2023年5月、2024年2月には複数の国家のセキュリティ関係機関が連名で注意喚起を行っている。また、利用者が多いシステム・サービスの脆弱性への攻撃も続いている。企業向けファイル転送ソフトウェア MOVEit Transfer の脆弱性を狙った攻撃では、2024年3月の時点で、全世界の2,768組織が被害を受けたという。激化するランサムウェア攻撃に対しては、国際協力により摘発や攻撃用ネットワークの破壊も行われている。2024年2月のランサムウェア攻撃グループ「LockBit」の摘発では、約10カ国の捜査当局が連携した。

2023年は、生成AIの利用が急速に進み、悪用や誤用による脅威やリスクが注目され始めた。具体的には選挙等の政治的な宣伝戦、ロシア・ウクライナ戦争やイスラエル・ハマスの武力衝突等において生成AIによる偽・誤情報が拡散しているとの報道が続いた。国内でも偽・誤情報の生成・拡散の事例が確認されている。生成AIは真実でないコンテンツを簡単に生成できるため、偽・誤情報の拡散に注意することが大切である。

国内では、2023年6月に社会保険労務士向けクラウドサービスの事業者がランサムウェア攻撃を受け、約1ヵ月サービスが停止し、約3,400ユーザーの大半に影響が出た。2023年7月には、「LockBit」のランサムウェア攻撃により名古屋港のコンテナターミナル内のシステムが2日半停止し、コンテナの搬出・搬入作業に大きな影響があった。サイバー攻撃によるシステムやサービスの停止により、物流のような社会インフラにも影響が出るこ

とが再認識された。一方で、国内の個人情報漏えい、紛失事故の発生件数、流出した個人情報数は増加傾向にあり、過去最多となった。2023年は内部不正による大量の情報漏えいも報告され、大手通信事業者のグループ企業の内部不正では、2社で合わせて1,500万件を超える顧客情報漏えいが報告された。内部不正は組織の社会的信用を損なう恐れがあり、経営課題として対策に取り組む必要がある。

国外のセキュリティ政策としては、2024年2月、米国NISTがサイバーセキュリティフレームワーク(CSF)2.0版を公開した。10年ぶりとなる大きな改訂で、重要インフラにとどまらないすべての組織におけるサイバーセキュリティ対策の枠組みを示すものとして注目されている。また、2023年12月に米国は「SBOM管理のための推奨事項」を公表した。政府調達において取引先へのSBOM整備の義務化が進められている。欧州では、重要インフラに関し「NIS指令」及び「EUサイバーレジリエンス法案」の実装を中心に取り組んでいる。EU加盟国は2024年10月までに、自国の規定をNIS2指令に準拠させるよう求められており、準備が進められている。

国内のセキュリティ政策としては「サイバーセキュリティ2023」に基づき、対策の強化を進めている。2023年7月には政府機関等のサイバーセキュリティ対策のベースラインとなる統一基準群の全面的な改定がされた。また、同時に「重要インフラのサイバーセキュリティに係る安全基準等策定指針」、更に2024年3月には「重要インフラのサイバーセキュリティに係る行動計画」の改定版を公表し、重要インフラのサイバーセキュリティ確保に向けた取り組みを示した。

2023年度はAIの利用拡大に伴い、AIの安全性に関する政策面の取り組みも各国で進んだ。米国、英国、日本等において、AIの安全性に取り組むAIセーフティインスティテュートが各々設置される等、各国で短期間に法制化やガイドラインの整備、体制強化が進んでいる。日本は、2023年5月に開催されたG7広島サミットにおいて「広島AIプロセス」を発表し、AIの安全な利用に関する国際ルール作りに貢献している。

## 2023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
2023年 4月	● Wi-Fi ルーターで任意のコード実行を可能とする脆弱性が公開され、Mirai の亜種による悪用も観測 (3.5.1)	
5月	● 自動車メーカー子会社のデータがクラウド環境の設定ミスにより公開されていたことを公表 (3.6.2) ● 国家の支援が疑われる攻撃者グループによるゼロデイ脆弱性を悪用した攻撃の観測を発表 (1.2.2)	● G7 広島サミットで官民が連携したサイバー攻撃対策を推進 (2.1.1、2.2.1) ● CISA を含む各国の政府機関「Volt Typhoon」に関する合同のサイバーセキュリティ勧告を発表 (2.2.2)
6月	● 社会保険労務士向けクラウドサービスがランサムウェアによる不正アクセスを受けサービス停止 (1.2.1) ● ファイル転送ソフトウェアに対するゼロデイ攻撃により情報漏えいやランサムウェア被害が発生 (1.2.5)	● 「不正競争防止法等の一部を改正する法律」成立。ビッグデータ等を念頭にした限定提供データと、営業秘密の一体的な情報管理が可能に (2.1.3)
7月	● 名古屋港のコンテナターミナルで利用しているシステムがランサムウェア攻撃を受けて停止 (1.2.1) ● 顧客情報約 596 万件の不正持ち出しを大手通信会社が公表 (1.2.8) ● 国家が支援する攻撃者グループによる、ネットワーク貫通型攻撃による不正アクセスを公表 (1.2.2)	● NISC 「サイバーセキュリティ 2023」、[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群] 改定版、[重要インフラのサイバーセキュリティに係る安全基準等策定指針] 改定版公開 (2.1.1)
8月	● 福島第一原発処理水放出に関する偽・誤情報拡散 (4.1.3)	● 総務省「ICT サイバーセキュリティ総合対策 2023」公表 (2.1.4) ● EU「デジタルサービス法 (Digital Services Act)」発効 (2.2.3)
9月	● 米国フロリダ州の市が、建設業者を装ったビジネスメール詐欺に遭い約 120 万ドルを送金 (1.2.3)	● 警察庁、NISC、米国諸機関は中国を背景とする攻撃グループ「BlackTech」に関する注意喚起を発出 (1.2.2、2.1.5)
10月	● 元派遣社員による顧客情報約 928 万件の不正持ち出しを大手通信会社グループ企業が公表 (1.2.8) ● イスラエル・ハマス間の武力衝突勃発、フェイク画像拡散 (2.2.1、4.1.3)	● 経済産業省、IPA「インド太平洋地域向け日米 EU 産業制御システムサイバーセキュリティウィーク」開催 (2.2.1) ● 米国、AI に関する大統領令 14110 発布 (2.2.2)
11月	● 生成 AI を使用した岸田首相の偽動画拡散 (3.1.2)	● 英国「AI 安全性サミット (AI Safety Summit)」開催 (2.2.1)
12月	● 総合 IT 企業、約 94 万件の個人情報を含むファイルが閲覧可能な状態にあったと公表 (1.2.8、3.6.2) ● 国際刑事警察機構、2023 年 7 月から 12 月にかけて 34 ヶ国が参加した国際的な取り締りを主導 (1.2.3)	● 「広島 AI プロセス包括的政策枠組み」G7 首脳承認 (2.2.1) ● EU サイバーレジリエンス法承認 (2.2.3) ● 米国「SBOM 管理のための推奨事項」公表 (2.2.2)
2024年 1月	● 能登半島地震が発生、SNS で偽・誤情報拡散 (3.1.2、4.1.3) ● 台湾総統選挙に関連する偽・誤情報拡散 (2.2.2、4.1.3) ● 米国大統領選挙の予備選において、Biden 大統領のディープフェイク音声拡散 (4.1.3)	● デジタル庁「政府情報システムにおけるセキュリティ・バイ・デザインガイドライン」改訂 (2.1.2)
2月	● 約 10 ヶ国の捜査当局、LockBit テイクダウンを実施 (2.1.5、2.2.3)	● AISI Japan 設立 (4.1.4)。USAISI 設立 (2.2.2) ● 「Volt Typhoon」に関する再度の合同のサイバーセキュリティ勧告を発表 (2.2.2) ● NIST「サイバーセキュリティフレームワーク (CSF) 2.0 版」公開 (2.2.2)
3月		● NISC「重要インフラのサイバーセキュリティに係る行動計画」改定 (2.1.1) ● IoT 製品のセキュリティラベリング最終取りまとめ公表 (2.1.3、3.2.1、3.5.5) ● 欧州議会「AI 法」承認 (2.2.3)

※ 2023 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。ランサムウェア被害、標的型攻撃、ビジネスメール詐欺、DDoS 攻撃、Web 改ざん、フィッシング等の攻撃や被害は通年で発生している。表中の数字は本白書中に掲載している項目番号である。特に注目されたもののみを挙げた。他のインシデントや手口と対策、及び政策・イベント等については本文を参照していただきたい。



# 第4章

## 注目のトピック

2024 年は、各国の国政選挙が相次ぐことから、ディープフェイクが選挙に影響を与えることが懸念されている。本章では虚偽情報拡散の脅威と対策について取り上げる。

また各国で議論が活発となり、ガイドラインや制度整備が進んでいる AI について、セキュリティリスクの実態と影響、対策の最新動向を解説する。

### 4.1 虚偽を含む情報拡散の脅威と対策の動向

インターネット上の虚偽情報、あるいは真偽不明な情報の生成・拡散（特定の意図による拡散を含む）による社会の混乱や分断、対立は、近年その深刻さを増している。2016 年の米国大統領選挙以降、世界各国で主に国家の支援を受けた情報操作による影響工作や、世論誘導、社会の分断及び混乱を目的とするサイバー攻撃（情報操作型サイバー攻撃）が猛威を振るうこととなった。選挙における世論誘導や中傷、扇動、新型コロナウイルス感染症（以下、新型コロナウイルス）対策に関する混乱や陰謀論の広がり、ロシア・ウクライナ戦争及びイスラエル・ハマスの武力衝突におけるサイバー情報戦、認知戦等の脅威が連続して発生し、虚偽、あるいは真偽不明な情報の生成・拡散にどう対応すべきか、安全保障上の課題にもなっている。

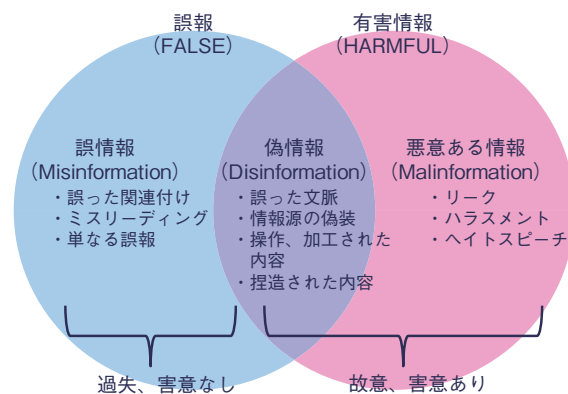
更に近年、「生成 AI (Generative AI)」と呼ばれるコンテンツ生成技術が急速に普及し、事実に見せかけた架空のコンテンツ、あるいは不正確なコンテンツ、著名人のなりすまし画像等が容易に作れる事態となり、生成 AI の利用の在り方の議論も始まっている。本節では、こうした虚偽あるいは真偽不明な情報の生成・拡散について、その脅威と対応の状況を述べる。

#### 4.1.1 虚偽情報とは

「虚偽情報」は、単純な意味では事実と異なる、あるいは不正確な情報を指すと考えられるが、近年、特にインターネット上で意図的に広められる虚偽を含んだ情報について、「デイスインフォメーション (Disinformation)」や「偽情報」「フェイクニュース (Fake news)」といった用語が様々に用いられている。以下ではこのような虚偽を含んだ情報を整理した。

#### (1) 虚偽情報の類型

虚偽を含んだ情報の拡散による社会の混乱（情報騒乱）については、2017 年に欧州評議会 (CoE: Council of Europe) が用語の整理を行っている (図 4-1-1)。この整理による各用語の定義は以下のとおりである。



■ 図 4-1-1 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<sup>\*1)</sup>」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)

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

報で、機密情報や個人情報の暴露を含むことが多い。

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その真偽性にある。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」の語は、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した国際的な定義はなく、特にディスインフォメーションについては定義に多少の揺らぎが見られる。日本国内の Disinformation 対策フォーラムでは、Disinformation を「あらゆる形態における虚偽の、不正確な、または誤解を招くような情報で、設計・表示・宣伝される等を通して、公共に危害が与えられた、又は、与える可能性が高いもの」と定義している<sup>\*2</sup>。また、欧州対外行動庁（EEAS: European External Action Service）の 2023 年のレポート<sup>\*3</sup>では、Disinformation について「経済的利益を得るため、または意図的に公衆を欺くために作成、提示、流布され、公共に損害を与える可能性のある、検証可能な虚偽または誤解を招く情報。公共の損害とは、民主的な政治・政策決定プロセスや市民の健康、環境、安全保障等の公共財に対する脅威を指す。」と定義して目的の一つとして経済的利益に言及するとともに、その意図として公共への害意を明示している。日本語では「偽情報」という訳語があてられているが、ディスインフォメーションは単に虚偽の情報を含むだけではなく、相手の誤解を招くために真の情報も混ぜ合わせて加工や情報操作が行われる点に注意が必要である。

なお情報の「虚偽性」については、以下の類型があると考えられる。

- 内容が事実でない、あるいは不正確なこと：

最も単純な虚偽である。

- 内容を拡大解釈、誇張すること：

事実に基づいていても、誤った解釈とともに拡散されることで誤解を招く。宣伝、他者攻撃等によく用いられる。

- 飛躍した論理で情報を関係させること：

都合よく抽出した事実や虚偽を並べ、仮定に過ぎないストーリーを正しいストーリーに見せる。こうした強い意図に基づくストーリーは、「ナラティブ (Narrative)」と呼ばれる。ナラティブは物語と訳されることもあるが、「多くの物語を含んだ、イデオロギー、理論、または信念に沿った出来事の説明であり、将来の行動への道を指し示すもの」である<sup>\*4</sup>。ナラティブは世界のありよう

を説明し、我々の理解を補い、これから何をすべきかを指し示すことで、行動変容を起こさせる。ナラティブが情報空間で用いられると、そこに含まれる多くの物語に強い感情が呼び起こされ、ナラティブ拡散者の意図する指針に影響を受けて共有行動が促進されることにより、関連の情報がよりいっそう拡散されやすくなる。

- 情報伝達の意図を誤らせること：

情報の本来の意図を錯誤させる。最も端的な例は、宣伝を宣伝と見せずに人を誘導するステルスマーケティング（広告主が自らの広告であることを隠したまま広告を出稿したり、企業からの依頼案件であることを隠した宣伝をインフルエンサー等に依頼したりすること<sup>\*5</sup>）である。

以下では、ディスインフォメーション（偽情報）を中心に、フェイクニュースやミスインフォメーション（誤情報）等も含めた虚偽を含んだ情報について、政府の用法<sup>\*6</sup>と合わせて「偽・誤情報」と表し、解説していく。

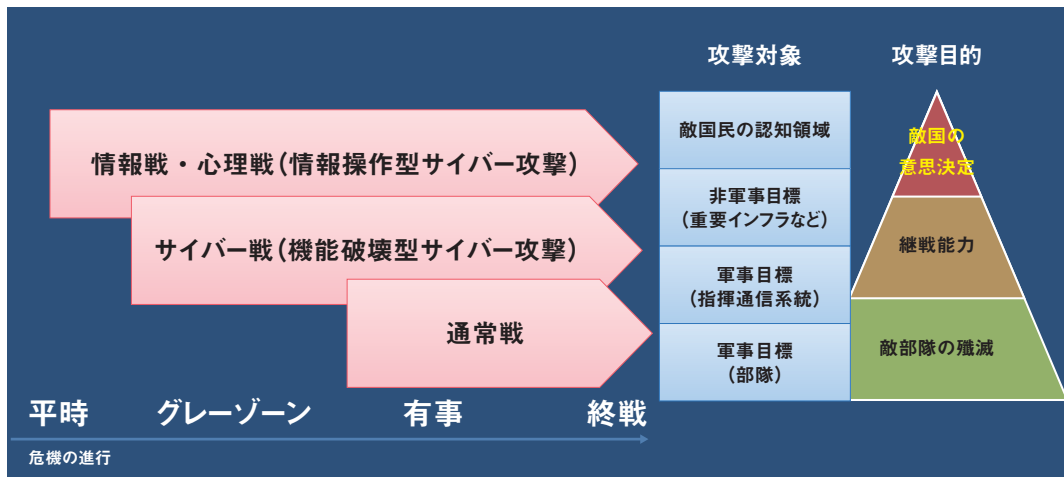
## (2) 偽・誤情報利用による安全保障上の脅威とその背景

近年、ディスインフォメーションを中心とした悪意ある情報操作が安全保障上の問題になっている。

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<sup>\*7</sup>が行われる中でも、国家による情報操作が大きな課題となってきた。ロシアによるハイブリッド戦争は、図 4-1-2（次ページ）で示すように 3 段階で行われている。第 1 段階は、戦闘が始まる前の平時から行われる情報戦・心理戦である。情報戦は、ディスインフォメーションを流布することによって、相手の社会混乱や政府機関の信用失墜を企図する、情報操作型のサイバー攻撃によって行われる<sup>\*8</sup>。

### (a) 情報操作型サイバー攻撃の実態

情報操作型サイバー攻撃はディスインフォメーションの流布だけではない。2016 年の米国大統領選挙では、民主党全国委員会（DNC: Democratic National Committee）等のネットワークがロシアに関係する APT 攻撃グループによりハッキングされ、内部の電子メールが流出し、その内容を利用して民主党や Hillary Clinton 候補を攻撃するディスインフォメーションが流布された<sup>\*10</sup>。2022 年のロシアによるウクライナ侵攻の直前には、ウクライナの国防総省及び国営銀行の Web サイトに対してロシア連邦軍参謀本部情報総局（以下、GRU）により



■ 図 4-1-2 ハイブリッド戦争の様相とサイバー攻撃類型  
(出典)大澤淳「台湾有事とハイブリッド戦争」<sup>9)</sup>

DDoS 攻撃が行われ、Web サイトが接続不可能となるとともに、銀行 ATM が機能していないというディスインフォメーションが流布された<sup>11)</sup>。更にロシアは、その情報戦戦略において、ディスインフォメーションを用いた工作活動と機密情報の漏えいとを組み合わせるとしている<sup>12)</sup>。

このように、情報操作型サイバー攻撃は、情報窃取型や機能破壊型のサイバー攻撃と組み合わせられることで更に大きな影響力を生む。そのため、情報操作型といっても情報戦の文脈にとどまらず、サイバーセキュリティの観点からも警戒と対策が必要である。

情報操作型のサイバー攻撃は、SNS やマイクロターゲティング（マーケティングや選挙活動において、対象となる個人の嗜好や行動を分析してより効果的な戦略を実行する手法）による Web 広告といった新たな IT 基盤を利用し、ディスインフォメーションを累積させることで我々の認知に影響を及ぼし、選挙や政治に関する行動変容を企図していると認識されている。攻撃者が有利なナラティブを意図的に形成することで、影響工作として更に大きな効果を生むといわれる。

#### (b) 国家戦略に利用されるナラティブとその狙い

国家による攻撃の意図のもとで戦略的に利用されるナラティブは「戦略的ナラティブ (Strategic Narrative)」と呼ばれ、これが国家間や組織間で利用されると「ナラティブの戦い (Battle of Narrative)」となる<sup>13)</sup>。この場合の戦略的ナラティブは、主にディスインフォメーションから構築される。それらは事実を含み得るが、「4.1.1 (1) 虚偽情報の類型」で示したように、文脈が歪められたり、悪意を持って操作され本来は表に出るはずのなかった不都合な真実であったり、真の目的を隠蔽するものであ

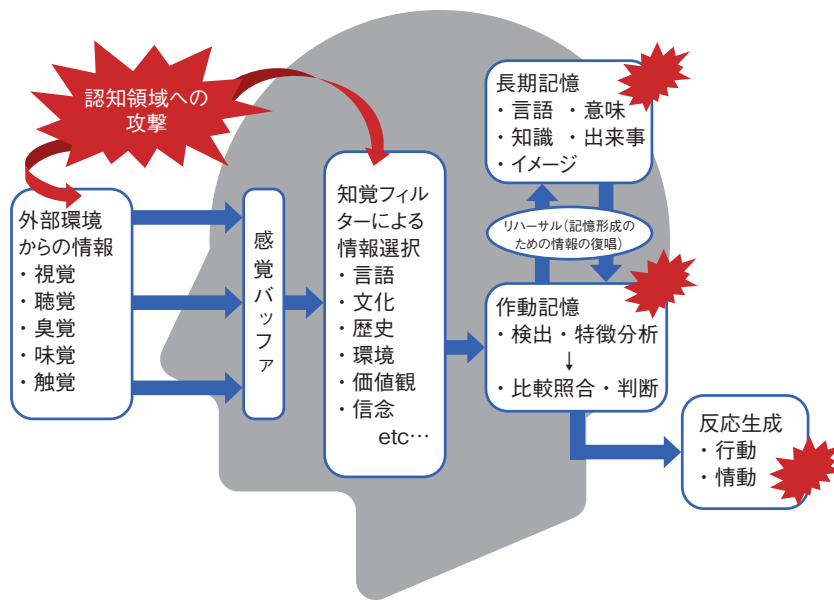
たりする。我々の認知に影響を与えやすい戦略的ナラティブとして、陰謀論が一つの脅威になっている<sup>4)</sup>。

ナラティブの戦いでは、イデオロギー、理論、信念に関する人間の認知情報処理に働きかけ、そこから生成される行動や言動に変容を起こすことを狙いとする。視覚や聴覚等の感覚入力にディスインフォメーションや誤った情報をインプットし、更にナラティブを通じて過去の記憶と紐付けられたワーキングメモリー（作動記憶）にも働きかけ、情報の取捨選択を行う知覚フィルターを変容させる。このフィルターを通じて、認知領域の中でその個人特有の現実の解釈（内部表象）が生まれるため、知覚フィルターを攻撃することで干渉したい事象（政治や選挙等）の解釈を操作しようとする。その結果として、個人の感情や行動に影響を与え、攻撃の所与の目的である戦略的な結果を引き出そうとするという<sup>14)</sup>。このような攻撃者側の戦略は安全保障分野で理論的な研究が積み重ねられており<sup>15)</sup>、このプロセスを示したものが図 4-1-3（次ページ）となる。

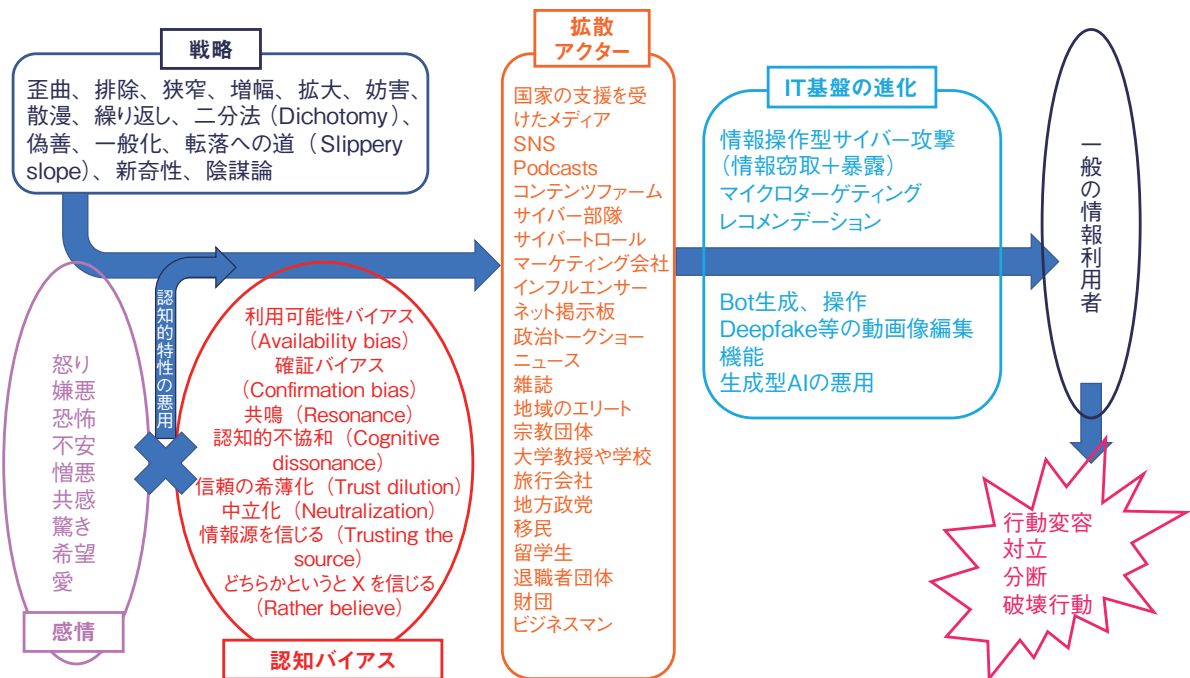
偽・誤情報の利用は、情報操作型サイバー攻撃やナラティブの戦いという形で広がり、安全保障にも影響を与えている。安全保障において、認知領域は陸・海・空・宇宙・サイバー空間に続く新たな第六の戦場とみなされる。現在の国家間の紛争では、認知領域への攻撃は平時有事を問わず行われ、「認知戦」と呼ばれている。

#### 4.1.2 ディスインフォメーションの生成・拡散の流れ

情報操作を狙いとしたディスインフォメーションの生成・拡散の流れについて、整理した結果を図 4-1-4（次ページ）に示す。



■ 図 4-1-3 人の認知処理フローと認知領域への攻撃イメージ  
 (出典)公益財団法人笹川平和財団安全保障研究グループ「“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~”<sup>\*14</sup>」を基に IPA が編集



■ 図 4-1-4 ディスインフォメーションの生成・拡散の流れ  
 (出典)Puma Shen「How China Initiates Information Operations Against Taiwan」<sup>\*16</sup>」を基に IPA が作成

### (1) 生成・拡散の流れ

攻撃主体は、自身の利益や自身の主張の優位性を確立するため、宣伝、あるいはディスインフォメーション等を利用して情報騒乱を引き起こすような戦略に基づいて情報拡散を行う。虚偽のニュースは、真実よりも速く、深く、広く拡散されるという性質が統計的に示されており、組織的なボットよりも、実際には人間が虚偽情報をより多く拡散してしまうことが明らかになっている<sup>\*17</sup>。

情報拡散の際には、自己の優位性を確立するナラティブの形成を目指すとともに、情報拡散に利用しやすい、対象の強い感情を引き起こすナラティブが既に成立している場合は、そのナラティブを拡散に利用する。ナラティブの形成、拡散においては、情報の正確性を歪めたり、より強い感情を惹起したりするような人間の認知バイアスも加味される。強い感情は真偽判断よりも共有意図に大きな影響を与えるが、特に怒りが喚起された場合にこの

傾向が見られる。ディスインフォメーションは強い怒りを喚起することが多いことが先行研究で指摘されており、真偽判断に関わらず、二次的な社会的共有（投稿のシェア、リポスト等）を促進することが示唆されている<sup>\*18</sup>。こうした情報に感情を操作された情報利用者に加え、アクセス数や広告収入増加等の利益（アテンションエコノミー）を意図したインフルエンサー等の第三者、拡散側の国家や組織の支持者が拡散アクターとして増幅、拡散を行う。

また、現在の SNS や Web 検索においては、ユーザーの使用履歴を蓄積し、その傾向や嗜好から各ユーザーにあった広告やリコメンデーションが表示されるシステムが多い。そうした環境においては、ユーザーがより情報を追いたくなるような、好みに合った情報が表示され続け、自分の好みのフィルター（バブル）の中に閉じ込められる「フィルターバブル」という状況に陥りやすい。更には、自分と同じ意見、思想、嗜好の人の記事や書き込みを読む方が心地良いため、気付かぬうちに、自分と同じ意見だけが聞こえてくる環境（エコー：こだま）を無意識的に選択し、多様な情報や自分の考えと異なる情報が目に入らなくなることがある。こうした状況を「エコーチェンバー」という。情報を拡散する側だけでなく、受け手のこうした環境も、情報拡散の一因となる。

こうした一連の流れによって、攻撃主体は、相手の行動変容や対象となる社会の対立、分断を企図している。

## (2) 流れを加速する IT 基盤の進化

近年の IT 基盤の進化により、偽・誤情報の生成・拡散のコストは結果的に大幅に削減されることとなった。偽・誤情報を拡散したい組織及びその支援者は、この基盤を活用して生成・拡散システムを形成している。以下の要素が高度化・自動化したことは、偽・誤情報の生成・拡散システムが拡大する主要な技術的要因となった。

### • 情報窃取・悪意の拡散

サイバー攻撃による不正アクセス等を用いて、評価を貶めようとする組織・個人の情報を窃取する。また大量のボットにより、悪意ある情報拡散を自動化する。

### • コンテンツ生成

生成 AI による「事実とは似て非なるコンテンツ」の生成では、政治家や芸能人のディープフェイク動画拡散等で悪用が懸念され、後述するとおり被害事例も増加している。2020 年以降急速に普及している大規模言語モデル (LLM: Large Language Model) を用いた対話型 AI では、事実とは異なる内容や、文脈と無関係な内容が一見もっともらしい回答として出力され

てしまう現象（「ハルシネーション」）があり、AI から返答を受け取ったユーザーに真偽を判断できる知識がないと、それが正しい情報だと誤認されて拡散されるリスクがある。更には、悪意のある攻撃者がナラティブを利用し真偽を交えたコンテンツを安価に大量生成するといった脅威が考えられる。生成 AI を用いたディスインフォメーション作成に関しては、ディスインフォメーションを含む合計 1 万 7,000 語以上、102 件のブログ記事を 65 分で生成できた上に、40 以上の言語に変換可能な、ワクチンに関するディスインフォメーションを語る医療専門家のディープフェイク動画を 5 分以内に生成できたという実験結果がある<sup>\*19</sup>。

### • 広告関連機能による拡散・増幅

マイクロターゲティング技術の悪用は、個人を特定してその嗜好や思想傾向にカスタマイズした情報生成や情報配信を容易にする。また、検索エンジン等で用いられるレコメンデーションアルゴリズムは、個人だけではなく特定志向を持つグループに彼らが好む情報ばかりを提示し、有害な情報の拡散や特定グループ内での情報の濃縮、同質化を促進する。

## 4.1.3 虚偽を含んだ情報生成・拡散の事例

2023 年度に発生した主な事例を以下に記載する。

### (1) イスラエル・ハマスの武力衝突

本項では、イスラエル・ハマスの武力衝突で見られた偽・誤情報の拡散及びそれらを利用した情報戦の様相について整理する。

#### (a) 事案の内容

2023 年 10 月 7 日、パレスチナのガザ地区を実効支配する武装組織ハマスがイスラエルに対してミサイル攻撃を開始するとともに、イスラエル南部に戦闘員を侵入させ民間人多数を殺傷、拉致するに及んだ。これにより、イスラエルとハマスは戦闘状態となり、これは 2024 年 4 月現在も継続している。

この武力衝突においては、イスラエル政府とハマス双方の公的な発信による熾烈な情報戦が繰り広げられた。一例としては、2023 年 10 月 17 日に起きたガザ地区のアル・アハリ病院の爆破事件が挙げられる。ハマス側はすぐにこれをイスラエルによる故意の空爆であるとして即時に非難し<sup>\*20</sup>、イスラエル側もハマスによる攻撃と断じる発表を行った<sup>\*21</sup>。イスラエル政府の X (旧 Twitter) 公

式アカウントは、これがハマスのロケット弾の誤作動によるものだとするビデオまで公開したが、The New York Times はビデオのタイムスタンプの分析に基づいてその動画がフェイクであることを暴いた<sup>\*22</sup>。米国は各種情報の分析から、これはハマスの誤射であるとしているが<sup>\*23</sup>、反イスラエルのクラスター（組織や団体、個人の集まり）はイスラエルの戦争犯罪の一つだと喧伝し続けている<sup>\*24</sup>。

そのほかに拡散されたミスインフォメーションやディスインフォメーションの例としては、「イスラエルがガザに核爆撃を行った」「イスラエルがガザに白リン弾を使用した」「イスラエルがアイアンビームを実戦投入」「ハマスはイスラエルの子供をさらい檻に閉じ込めている」「パレスチナが死者を捏造している」「パレスチナの病院で被害を訴える医師は女優」「イスラエルはガザ南部の退避地域に空爆を行っておらず民間人を攻撃していない」といった情報がある<sup>\*25</sup>。

こうした情報工作に加えて、イスラエル政府や地元紙の The Jerusalem Post 関連の Web サイトに対して DDoS 攻撃が行われたが、これは情報工作を否定する発表を阻害する狙いがあるとされている<sup>\*26</sup>。一方でイスラエル側も、ハマスの情報戦能力を削減させるために、空爆でガザ地域の携帯電話通信用の電波塔を標的にし、同地域で主力なインターネットサービスプロバイダーへの電力供給を拒否する等、情報インフラ機能を破壊している。こうした攻撃により、2023 年 10 月末までに、ガザ全域のインターネット・トラフィックは 80% 減少した<sup>\*27</sup>。

#### (b) 関係組織

前述のとおり、この武力衝突では、イスラエル政府と武装組織ハマスの情報戦が展開されており、いずれの公的発信も情報戦戦略の一環として解釈する必要がある。

また、この武力衝突を利用して反欧米のナラティブを強化するために、ロシア、中国、イランが国営メディアや政府高官の SNS 発信等で反イスラエル、反米のディスインフォメーションを拡散している<sup>\*24</sup>。

このような活発な情報戦において、公開情報を分析してファクトチェックを行う OSINT（Open Source Intelligence）分析グループが数多く活動している。The New York Times や BBC 等メディアの検証チーム、Bellingcat 等の OSINT 専門グループ、個人や私的なグループで OSINT 分析家を標榜する者等様々だが、こうした OSINT 分析の乱立による競争が生まれるように

なった。SNS の収益化構造も相まって、より早く決定的な分析結果をフォロワーに提供しようとする中で、彼らはかえって、特定の国や勢力を利する政治的なナラティブを強化する一因となってしまっている。このような状況で、OSINT 分析すらも情報戦の兵器として国家等の特定勢力に利用されてしまう（兵器化）構造となったことが指摘されている<sup>\*22</sup>。

#### (c) 手口

情報戦の主な舞台は SNS であった。Thierry Breton 欧州委員（域内市場担当）は、X、Facebook、TikTok、YouTube に対し、今回の武力衝突に関するディスインフォメーションの抑制が十分でないと批判したが、各社は有害なコンテンツに対処する措置を講じたと述べている。2023 年 10 月 7 日以降、イスラエル検察庁のサイバー部門は、ハマスに関連する暴力を扇動する SNS コンテンツに対して削除依頼を出しているが、攻撃開始後 2 週間でその件数は約 4,500 件に上り、その大半は Facebook、TikTok、X に対するものだという<sup>\*28</sup>。

イスラエルのソーシャルメディア分析企業である Cyabra によれば、攻撃開始後の 1 ヶ月で、少なくとも約 40,000 件以上の Bot アカウント及び不正なアカウントを確認したという。また、Facebook、Instagram、TikTok、X でこの武力衝突について投稿したアカウントのおよそ 4 個に 1 個が偽のアカウントであることが攻撃後 1 日で判明したという。更には、アル・アハリ病院での爆発から 24 時間以内に、X に本件を投稿したアカウントの 3 個に 1 個以上が偽のアカウントであった<sup>\*29</sup>。

また、この情報戦を自国に優位なナラティブ形成に利用しようとして、ロシアや中国、イランの政府高官、省庁、大使館といった政府関連アカウントは、SNS 上の発信において、Sputnik や RT（旧称、Russia Today）、Global Times（環球時報）、Tasnim News Agency といった国営（ないしは半国営）メディアからの発信を積極的に引用していた。ロシアにおいては、前述した各社 SNS に加え、Telegram の利用も見られる<sup>\*24</sup>。

この情報戦では、生成 AI によって作成、加工された動画の流布が多く見られた。代表的な事例としては、男性が子供達を瓦礫から救出する画像、瓦礫に押しつぶされ悲鳴を上げる赤ん坊の画像、兵士達がイスラエル国旗を掲げて爆撃地を行進する画像、イスラエル国民がマンションの各部屋から国旗を掲げてイスラエル兵士を歓迎する画像、米国の人気モデルがイスラエル支持を表明する動画等が確認されている<sup>\*30</sup>。

日本ファクトチェックセンター（JFC：Japan Fact-check Center）によると、男性が子供達を瓦礫から救出する画像では、AIが生成した画像に特有な、細部の不自然な描写が多数確認できるという<sup>31</sup>（図4-1-5の丸で囲われた箇所や矢印の箇所）。

小さな子供や赤ん坊の被害を強調して悲壮な感情を扇動する画像は、「#Gaza\_under\_attack」「#Free Palestine」といったハッシュタグとともにパレスチナ擁護側から拡散され、またイスラエル国旗を用いて愛国心を鼓舞する画像は、「#HamasTerrorist」「#IsraelFightsBack」といったハッシュタグとともにイスラエル支持側から拡散された<sup>33</sup>。ただし、「戦争の被害に遭うかわいそうな子供達」というナラティブ自体は、イスラエル・ハマス双方が利用している<sup>34</sup>。

政治的な意図によって作成されたAI画像以外にも、ストックフォトサービスにおいて「中東戦争」とタグ付けされたAI画像が商品化されており、これが今回の武力衝突に関する現実の画像と誤認されて拡散されたり、生成AIによるものと明示せずに使用されたりすることで誤解を招くという問題も指摘されている<sup>35</sup>。

JFCによると、パレスチナ系の米国人モデルであるIsabella Khair Hadidがイスラエルの対応を非難したところ、彼女の過去のスピーチ動画を、生成AIを利用し加工して、イスラエル非難の発言を謝罪しイスラエル支持を表明する動画が捏造され拡散された。この動画では、彼女の音声をAIに学習させ捏造音声を生成するとともに、生成した音声に合わせてリップシンク（唇が連動）



■ 図4-1-5 生成AIによるディープフェイク画像生成・拡散の事例<sup>32</sup>

するように動画が加工されていたという<sup>36</sup>。

#### (d) 影響

この情報戦においては、ハマス側の工作スピードが速く、またイスラエルの非道さを強調するナラティブが功を奏し、イスラエル側は守勢に回っていると評価されている<sup>29</sup>。イスラエルや同国を支援する米国の政府高官には、既に国際世論の支持を失い武力行使の大義を失っているという認識があるという<sup>37</sup>。情報工作の効果について客観的な評価は困難であるが、米国の世論調査では、米国のイスラエル側に立つことを望む人の割合は、2023年10月の43%から同年11月には37%に減少したという<sup>38</sup>。これは、この情報戦の影響の一つと考えられる。

ロシアや中国、イラン等非交戦国による周縁的な情報戦の動きについては、彼らの反欧米のナラティブが欧米の枠を超え、反植民地主義的な不満と呼応するグローバルサウスにおいて共鳴する可能性が高いという指摘がある<sup>24</sup>。この情報戦で、グローバルサウスにおけるウクライナ支持結集の努力や欧米的な自由民主主義の価値観までも毀損されるという影響が広がっているという。

## (2) 福島第一原発処理水放出

本項では、東京電力ホールディング株式会社福島第一原子力発電所（以下、福島第一原発）のALPS（Advanced Liquid Processing System）処理水放出をめぐる偽・誤情報の拡散及び情報戦の様相について整理する。

### (a) 事案の内容

2023年8月24日、福島第一原発のALPS処理水の太平洋への放出が開始された。その際、この処理水を巡って様々なディスインフォメーションやミスインフォメーションが拡散された。事例としては、「魚の大量死は福島第一原発からの処理水の影響」「ALPSを通してストロンチウムを含む放射性物質の約6割が除去されず海に放出される」「日本政府は汚染水を処理せず福島第一原発からそのまま放出」「国際原子力機関（IAEA）の報告書には欠陥がある」「IAEAは日本の計画を支持していない」「汚染水放出は日本の魚介類を汚染し、食用に適さなくなる」「処理水放出後に海面の色が変化する程の汚染があった」といった内容があった<sup>39</sup>。これらは、処理水放出に批判的な立場から発せられたため、処理水（treated wastewater）を核汚染水（nuclear-contaminated water）と呼ぶことによる印象操作も行わ

れた。

また、特に韓国のネットメディアからは、「日本政府がIAEAに対して100万ユーロ以上の政治献金を行った」「IAEAレビュー報告書の結論は最初から絶対安全と決まっている」「IAEAレビューに参加する第三国専門家は飾り物である」「外務省の公電にて『処理水のタンク群の調査の結果、放射能濃度が基準を大幅に超過したため、バラスト水の交換によってALPS処理水の希釈を加速し、安全基準を満たすことが検討されている』といった記載がある」といったディスインフォメーションが拡散され、これらについては、外務省が計2回にわたって公式に反論文書を公表した<sup>\*40</sup>。

また、台湾事実査核中心（台湾ファクトチェックセンター）によると、西村康稔経済産業大臣（当時）が中国による日本産水産物の禁輸撤廃を求めた会見の動画について、簡体字が交じった中国語で、「中国や香港、マカオに輸出予定だった2万匹の魚が台湾に運ばれた」といった虚偽のキャプションが追加されて中国語圏のSNS上で拡散されたという<sup>\*41</sup>。これについては、経済産業省が公式に反論している<sup>\*42</sup>。

#### (b) 関係組織

日本国内では、処理水放出に批判的なSNS上のクラスターにおいて、前述のようなディスインフォメーションが拡散された。更に、中国政府と政府系メディアが結託して、福島第一原発の処理水に関するディスインフォメーションキャンペーンが行われたことが明らかになっている<sup>\*43</sup>。この動きに一部の太平洋島しょ国も同調し、ソロモン諸島政府やフィジーの野党を含む太平洋地域の親中派政治家や活動家には、中国と歩調を合わせて処理水が海に放出されたことを非難する動きがあった。また、前述のとおり韓国メディアからもディスインフォメーション拡散の動きがあった。

#### (c) 手口

主にSNSを中心にディスインフォメーションが拡散されたが、特に組織的な中国の反処理水キャンペーンは以下のようなものであった<sup>\*43</sup>。

- 2023年1月から8月にかけて、環球時報は福島第一原発の排水放出に関する約126本の英文記事を掲載した。同時に、人民日報は福島第一原発の排水に関する記事を英語で約74本、日本語で約60本掲載した。
- 国営テレビである中国中央電視台（CCTV:China

Central Television）、国営ラジオである中央人民廣播電台（CNR:China National Radio）、海外向けのラジオ放送である中国国際廣播電台（CRI:China Radio International）を含む中国の国営メディアが、英語、ドイツ語、ポルトガル語、クメール語を含む複数の言語で、廃水放出がもたらすリスクに関して少なくとも22の有料広告をFacebookやInstagram等のSNSに掲載した。広告は、世界中の少なくとも1,000以上のMeta Platforms, Inc.（以下、Meta社）の広告ライブラリを利用する各種SNSページで掲載された。

- 同年8月24日、福島第一原発事故に関する以下のような北京語のハッシュタグが、中国のSNSである微博（Weibo）のトレンドの上位を占めた。「日本核汚染水排海正式开始（翻訳：日本の核汚染水排出が正式に始まる）- 24億回読了」「日本将用700億日元处理核汚染水负面信息（翻訳：日本は核汚染水に関するネガティブな情報への対応に700億円を使う）- 4億3000万回読了」「中国日料店会大批量倒闭吗（翻訳：中国の日本食レストランは大量に廃業しているのか？）- 3億2,000万回読了」「受日本核污水影响最大的省份（翻訳：日本の原子力汚水によって最も影響を受けた地方）- 1億3千万回読了」等。
- 2023年1月1日から2023年8月25日の間に、中国の国営メディア、政府関係者、親中派インフルエンサーによる「フクシマ」に言及した投稿がWeibo、Facebook、X等のSNSにおいて1,509%増加した。

また、JFCによると、海水面の変色等、中国の動画サイトで拡散されていたトピックが日本語に訳されてSNS上の日本語クラスターに拡散されていた例も見られたという<sup>\*44</sup>（次ページ図4-1-6）。

#### (d) 影響

これらのディスインフォメーションは、米国の同盟国である日本の失敗を強調し、周辺国に害を与えようと思わせることによって、日本を世界から孤立させるとともに、日本国内の世論の分断をも扇動したと見られる。また、現在の世界秩序維持における国際機関の無力さも喧伝するナラティブの拡散を意図したと分析されている<sup>\*43</sup>。

### (3) 台湾総統選挙、立法委員選挙

本項では、台湾総統選挙における偽・誤情報の拡散と情報戦の様相について整理する。





■ 図 4-1-6 偽・誤情報拡散の事例<sup>\*45</sup>

### (a) 事案の内容

台湾の選挙においては、影響工作のためのサイバー攻撃を含む選挙介入が確認されている<sup>\*46</sup>。2016年1月16日の総統選挙では、政府関係者、台湾独立運動家に向けて、ハッカー集団「APT12」によるフィッシング攻撃が行われた。また、2018年の統一地方選挙では、民進党が大敗を喫したが、これについて、民進党の報道官が中国による介入があったとコメントしている。2020年1月11日の総統選挙においても、米国のシンクタンクである戦略国際問題研究所（CSIS: Center for Strategic and International Studies）の分析によれば、2019年から中国の介入があったとされる。この分析によると、①資金供与によって親中の候補に有利となる調査結果を台湾の報道機関と世論調査会社で作成・公表させる、②コメントの投稿で報酬を受け取れる「五毛党」を組織し、Facebook等のSNSで反中候補者を攻撃し、親中コメントを投稿させる、といった手法が使われているとのことである<sup>\*47</sup>。

2024年1月の総統選挙に向けても、サイバー攻撃<sup>\*48</sup>を含む選挙介入が観測された。この影響工作活動においては、軍事的圧力、経済威圧行動も組み合わせられ、報道機関やソーシャルメディア等を通じたディスインフォ

メーションの拡散が行われたとされている<sup>\*49</sup>。

2023年6月に実施された台湾側の軍事演習においては、蔡英文総統（当時）が有事の際に台湾を脱出する練習をしている、といったディスインフォメーションが流布された<sup>\*50</sup>。2023年8月に中国が台湾周辺海域で軍事演習を行った際には、航空機や艦船による中間線への接近頻度を増すことで、単純に軍事的緊張を高めるだけでなく、「戦争の恐怖」を作り出すことによる認知戦の要素もあったと指摘されている。世論の反戦意識を高め、中国が台湾内の特定のメディアや政治団体と協力し、台湾の選挙への実質的な介入を行っていると安全保障関係者は分析している<sup>\*51</sup>。また、脱走兵が出たというディスインフォメーションの流布や、海事を司る女神である媽祖の庇護は中国にあるといった映画を製作することで、中国に有利なナラティブの強化を行うといった活動も見られた<sup>\*52</sup>。これに対して台湾国防部は即座にプレスリリースで反論し、台湾市民に対して台湾側に有利なナラティブの擁護に務めた。軍事的アプローチにとどまらないこうした台湾国防部の対策行動は、多角的な認知戦対策として有効であると考えられる。

ディープフェイクの事例としては、民進党候補者の頼清徳氏が暗号資産への投資を推奨する動画、同氏が野党の主張を擁護する動画、米国下院議員 Rob Wittman 氏が、2024年1月13日の総統選挙で与党・民進党の候補者が勝利した場合、台湾への軍事支援を支持するというフェイク動画等がSNS上で確認された<sup>\*53</sup>。また、音声によるディープフェイクとして、台湾民衆党候補者の柯文哲氏のものとする音声ファイルがメディア関係者にメールで拡散された。その内容は外遊中の頼氏が米国で金を払って支援者を集めている、というものであった<sup>\*54</sup>。こうした活動と並行して、台湾の分断を促進するディスインフォメーションが流布されている。

### (b) 関係組織、手口

中国は、情報戦の一環として、この10年間に国営メディアに多額の投資を行ってきた。現在、中国国営メディアは少なくとも12カ国語で配信し、世界中の視聴者に到達している<sup>\*55</sup>。主なメディアはCCTV、CNR、CRI、国際ニュース放送を行う中国国際電視台（CGTN: China Global Television Network）、新華社通信、中国通信社等である。例えば、CGTN発行の記事は日本のYahoo! ニュースにおいて日本語で配信されており<sup>\*56</sup>、ニュースサイトやまとめサイト等を閲覧する際には記事の配信元の確認が必須である。

情報戦の拡散アクターであるインフルエンサーに関しては、中国は Facebook グループや配信者に外貨を支払って記事を拡散させる動きがあることや、国の支援がなくとも、親中派のメッセージを発する台湾国内の配信者には中国の愛国者達から寄付が集まりやすい構造があることが挙げられる<sup>\*57</sup>。台湾では、親中派の記事を広めるだけで、毎月1,500米ドルをも稼ぐことができる Facebook ページが確認されている。寄付については、2019年に台湾でネット上の寄付を受けた YouTuber トップ10のうち、7人が親中メッセージを拡散していた。トップの YouTuber は7万人の登録者しか得ていないが、年間100万台ドルの寄付を集めていたという。

#### (c) 影響

台湾政府は、SNSだけでなく、テレビ等のメディアについても監視を徹底した。総統・副総統候補による政見発表会やテレビ討論会では各省庁がチェック体制を敷き、誤った情報があれば即座に報道資料を発表して訂正した。また、国家安全法や反浸透法に基づき、中国による選挙への介入について検察は捜査を開始している<sup>\*59</sup>。

2024年1月の台湾総統選の選挙結果としては、与党・民進党の頼清徳氏が当選したものの、頼氏の得票率は40.05%と伸び悩み、新興政党・台湾民衆党の柯文哲氏に票が流れた。これは与党とは対中関係で立場を異にする野党の伸長を示しているとする見方もある<sup>\*60</sup>。中国による認知戦の成果の客観的な測定は難しいが、前述した多様な工作による認知戦の影響もあったといえるだろう。

### (4) 令和6年能登半島地震

本項では、2024年に発生した能登半島地震における、偽・誤情報の拡散状況について整理する。

#### (a) 事案の内容

2024年1月1日、能登半島で最大震度7を記録する大地震が発生した。この震災の混乱に伴い、多くの偽・誤情報が拡散され、以下のような事例が確認されている<sup>\*61</sup>。

- 2011年の東日本大震災の津波の映像を使って、まるで能登半島地震の被害のように誤認させるもの
- 被災地の住所を転記して、まるで自分が被害に遭っているかのように誤解させる偽の救助要請
- PayPay等を経由した虚偽の寄付募集
- 全国から能登半島に盗賊団が大集結中といった根拠

不明の犯罪情報

- 志賀原子力発電所で放射性物質を含む水が2基で約420リットル漏えい中といった原発事故を誤認させるもの
- 人工地震等の陰謀論

#### (b) 関係組織、手口

この地震においては、詐欺目的やいたずら目的の個人の情報発信に加えて、X特有の問題として、いわゆる「インプレゾンビ」による情報の混乱が問題となった。Xは2023年から、課金しているユーザーが一定の「インプレッション」（投稿されたポストが表示された回数）を獲得すると、収益が得られる仕組みを導入している。そのため、こうした災害時に話題になりやすいトピックで投稿を行うことでインプレッションを稼ぎ、増収を狙うアカウントが多数見られた。前述の事例のうち、偽の救助要請を投稿したアカウントの中には、Xでインプレッションを獲得すると収益が得られる仕組みやその方法を教える動画を公開しているケースや、こうした偽・誤情報の投稿で収益を上げたことを報告しているケースも確認された<sup>\*62</sup>。こうしたインプレッション稼ぎのアカウントには、南アジアや中東地域のユーザーも多く、日本語の投稿のコピーアンドペーストだけでなく、インプレッションを集めている投稿にアラビア語でのリプライ（返信）を行うことで主投稿の閲覧者からのインプレッションを稼ぐ事例等も見られる。

#### (c) 影響

地震発生翌日、情報騒乱の状況を危惧し、岸田首相自身が「虚偽情報の流布は許されない、こうした行為は謹んでほしい」と国民に呼びかけた<sup>\*63</sup>。併せて、総務省はSNS等の4事業者に適切な対応を求めた<sup>\*64</sup>。総務省によると、Meta社とLINEヤフー株式会社（以下、LINEヤフー社）では偽情報であることが明らかな規約違反投稿を削除し、X Corp.（以下、X社）はQRコードで寄付等の支援を求める疑わしいアカウントを凍結、Google LLC（以下、Google社）はYouTubeのモニタリング体制を整えるといった対応を取った。更に、総務省の「デジタル空間における情報流通の健全性確保の在り方に関する検討会」において、2024年1月19日の会合で、偽情報に関する新たな作業部会の設置が決定された<sup>\*65</sup>。この検討会を通じて、Google社やLINEヤフー社といった主要プラットフォーム事業者にはヒアリングが行われ、同年5月15日に暫定版の報告書が公表された<sup>\*66</sup>。

## (5) 2024 年の各国の国政選挙

本項では、2024 年の各国国政選挙に向けて増加している偽・誤情報の拡散について、特に動画や音声を用いた事例を中心に整理する。

### (a) 事案の内容、手口

2024 年は選挙イヤーと言われており、米国大統領選挙を始め、各国で国政レベルの選挙が実施される。2024 年は生成 AI が広く普及してから初めての選挙イヤーでもあり、以下のような政治家や選挙に関係する著名人のディープフェイク動画、音声等が数多く拡散している。

- 米国の Joe Biden 大統領が「徴兵法を発動する」と発言して、徴兵への協力を呼びかける動画<sup>\*67</sup>  
TikTok では 20 万以上の「いいね」と 1 万 1,000 以上のコメントを集め、その後、他のプラットフォームにも拡散した。
- ニューハンプシャー州の有権者に対する、予備選挙への参加を思いとどまらせる Biden 大統領の音声によるロボコール(自動の電話メッセージ)<sup>\*68</sup>
- トランスジェンダーに対して、あなたは女性にはなれないと非難する Biden 大統領の動画<sup>\*69</sup>
- インドネシアの Joko Widodo 大統領が中国語で演説する動画<sup>\*70</sup>
- ウクライナ支援に反対するメッセージを伴うセレブリティ達の画像  
歌手の Taylor Swift、サッカー選手の Cristiano Ronaldo といった有名人が利用され、Facebook 上で 560 件の広告掲載がなされたことで、およそ 760 万人の目に触れたと考えられている<sup>\*71</sup>。

また、選挙の当事者である政党同士が相手の党や候補者への攻撃のために生成 AI による動画像を利用する例も増えている。共和党全国委員会(RNC: Republican National Committee) は、Biden 大統領の再選による終末的なシナリオを描いたディープフェイク動画を作成し、投稿後に広く拡散された。ポーランドでは、主要野党・市民プラットフォーム(PO: Platforma Obywatelska) が、現与党である法と正義(PiS: Prawo i Sprawiedliwość) の内政を批判する Mateusz Jakub Morawiecki 首相のディープフェイク音声を公開した。スロバキアでは、進歩党の党首が選挙の不正操作を画策しているとするディープフェイク音声が、投票所の開場数日前に拡散された<sup>\*67</sup>。

### (b) 関係組織

前述した国内当事者同士の関与や、政治的な意図を持つ個人や私的な集団以外に、外国から他国の選挙に干渉しようとする勢力としては、ロシアの Doppelgänger (ドッペルゲンガー) が挙げられる。彼らはロシアの GRU と関係があるとされ、米国に焦点を当てたディスインフォメーション工作を行っている<sup>\*72</sup>。2024 年の米国大統領選挙を前に、社会的・政治的分裂を利用することを目的とし、反 LGBTQ+ 感情を煽り、米国の軍事力を批判し、米国のウクライナ支援をめぐる政治的分裂を増幅させようとしている。前述したセレブリティを利用した反ウクライナ支援キャンペーン等もドッペルゲンガーによるもので<sup>\*71</sup>、彼らの近年の手法として AI を活用していることが特徴的であるとされる。

### (c) 影響

各国でこのようなディープフェイクが蔓延していることは、選挙の信頼性を毀損するものであり、民主主義への脅威がより高まっている状況といえる。

このような状況に対し、米国では、超党派の下院議員グループにより、AI 詐欺禁止法案 (No Artificial Intelligence Fake Replicas And Unauthorized Duplications Act: 通称、No AI FRAUD Act) が 2024 年 1 月に提出され、ディープフェイクへの法規制が議論され始めている<sup>\*73</sup>。

## (6) 新型コロナウイルス関連

本項では、新型コロナウイルス関連の偽・誤情報の拡散状況について整理する。

### (a) 事案の内容

2020 年 1 月に新型コロナウイルスが中国で確認され、世界に感染拡大した。それ以来、新型コロナウイルス関連の偽・誤情報や有害情報が SNS、報道等で拡散し続けた。現在は新型コロナウイルス自体が感染収束傾向にあるため、情報騒乱も沈静化してきてはいるものの、2023 年度もワクチンを中心とした虚偽情報が確認されている。

主な虚偽情報、悪意ある情報の類型は以下のような。

- 発生源に関する不確実情報  
発生源は中国武漢市の市場あるいはウイルス研究所である可能性が高いとの西側メディアの報道が 2023 年時点も継続している<sup>\*74</sup>。中国政府は研究所流出

説を強く否定している。

- 感染対処法に関するデマ  
発生直後、根拠不明の多くのコロナ予防や諸症状への対処法が拡散した。
- 対策に関する詐欺情報  
発生直後から、世界各国で詐欺情報が横行した。具体的には、関係省庁の注意喚起を装うフィッシングメール、マスクや新薬、ワクチン接種等に関する金銭詐欺、給付金支給に関する個人情報詐取等が確認された<sup>\*75</sup>。
- ワクチン接種に関する不正確な主張  
新型コロナウイルスワクチン接種に反対する人々は、医学的合理性が確認されていない「ワクチンによって死亡者が出た」「ワクチンは人口削減のため」といった言説を拡散した<sup>\*76</sup>。

#### (b) 関係組織

上記の不正確な主張について、反対派の人々の言説は「コロナ禍はディープステート（闇の政府）によるもの」「ワクチンは陰謀」といった陰謀論言説と親和性が高いことが明らかになっており<sup>\*77</sup>、こうしたナラティブに共鳴したグループは、SNS等で偽・誤情報の増幅・拡散を行ったと思われる。

#### (c) 手口

多くはSNSによる拡散であり、状況に応じて自然発生的に生じたと考えられる。ただし、ヘイトスピーチの拡散では自動で投稿するボットの利用が確認され<sup>\*78</sup>、共鳴するグループが組織的に拡散したケースがあることがうかがわれる。

#### (d) 影響

世界同時的な感染症による社会不安の発生に伴い、偽・誤情報・不確実情報、陰謀論の同時蔓延も生じ、この状況をWHOは「インフォデミック (Infodemic)」という言葉で表した<sup>\*79</sup>。中国、ロシアによる新型コロナウイルス関連の情報工作は国際関係にも影響を与え、安全保障上の脅威にもなったことが米国国務省のレポート等で指摘されている<sup>\*80</sup>。

### (7) ロシア・ウクライナ戦争

本項では、ロシア・ウクライナ戦争における偽・誤情報の拡散状況及び情報戦の様相について整理する。

#### (a) 事案の内容

2022年2月24日のロシアによるウクライナ侵攻において、ロシアはその侵攻前の段階から、情報窃取や機能破壊型のサイバー攻撃を組み合わせた情報戦を展開していた。併せて、ロシアの侵攻を正当化し、ウクライナ支援を行う欧米の分断を図るナラティブを対ウクライナ及び対国際世論への影響を目論んでSNS上で拡散しており、こうした情報工作の動きは2023年度も継続した。これらは、ロシアの安全保障政策に基づくサイバー情報戦の一環であり、ウクライナ及び西側諸国を敵視したナラティブの拡散は2014年のクリミア併合時点から継続している<sup>\*81</sup>。2024年3月22日にモスクワ近郊のクロクス・シティー・ホールが襲撃され、145人が死亡したテロ事件においては、イスラム国 (IS: Islamic State) が犯行声明を発表しているにもかかわらず、ウクライナと西側諸国の関与があったとロシア当局は主張している<sup>\*82</sup>。

#### (b) 関係組織

ロシアのサイバー情報戦の主体は政府、ロシア軍、ロシア連邦保安庁や対外情報庁等の情報機関、RTやSputnik等の親ロシア系報道機関、Internet Research Agency等の情報操作企業、親ロシア系ハッカーとされる。親ロシア系の第三国からの拡散もあるという<sup>\*83</sup>。

#### (c) 手口

ウクライナ侵攻において、ロシアが工作を行っている類型は以下の三つに大別できる。

- 国際世論をターゲットとしたサイバー影響工作  
Microsoft Corporation (以下、Microsoft社) の調査によると、ロシアは国内・ウクライナ・西側諸国・非同盟国それぞれに、SNS等に事前に配置した宣伝メッセージを一斉に拡散、自国、ウクライナの親ロシア派、非同盟国からの支持強化、及び西側諸国の分断誘発を図った<sup>\*84</sup>。
- ウクライナ政府の評価を毀損するディスインフォメーション  
「ウクライナ政府はネオナチ」「ゼレンスキー大統領が首都キーウを脱出」「プチャの虐殺は米英の陰謀」といったディスインフォメーションや、Volodymyr Zelenskyy 大統領が降伏を呼びかけるディープフェイク動画等が拡散された<sup>\*85</sup>。
- 偽旗作戦 (False Flag Operation)  
侵攻以前から「ウクライナ東部のロシア系住民が迫害された」「親ロシア勢力が攻撃された」等の情報が

SNS等で拡散した<sup>\*86</sup>。米国はこれを、侵攻を正当化するロシアの「偽旗作戦」と断じ、侵攻直前に Biden 大統領が公表するという措置を取った<sup>\*87</sup>。

また、SNSにおける活発な工作のうち、ロシアにおいて特徴的だったのは「Telegram」(テレグラム)の主戦場化である。

Telegramはロシアで誕生した守秘性の高いSNSであり、ロシア・ウクライナ双方で多くの人が利用している。2013年のサービス開始以来、監視等の機能がないために極右・過激主義・反体制集団等が宣伝を行っていたが、政府の規制は免れていた<sup>\*88</sup>。

2022年以降、Telegramはロシア、ウクライナ双方の政府及び政府支援グループが情報工作とプロパガンダを行う主戦場となった。ロシア国民は、他のSNSと異なりTelegram上では、自国に加え、ウクライナ及び西側諸国の拡散情報を自由に見ることができる<sup>\*89</sup>。これにはロシア政府にTelegram規制が自国に有利でない、という判断があると思われる。Telegram上では相手国政府にサイバー攻撃を行う「サイバー義勇軍」の勧誘も行われたという。

ウクライナでも、Telegramは戦時の重要な情報源となっている。多くのウクライナ人が国外に避難しており、テレビ等の主要なウクライナメディアを見ることができない状況で、Telegramのいくつかのチャンネルはマスメディアと同等の働きをしている。ロシア・ウクライナ戦争開戦前の2021年には、ウクライナでのTelegram使用率は20%前後であったが、2023年の調査では成人の72%がTelegramを使用しているという、急成長を示す結果となった<sup>\*90</sup>。しかし、Telegramはロシア発の偽・誤情報の温床となっており、2024年3月にはTelegramを規制する法案がウクライナに提出されている<sup>\*91</sup>。しかし、その規制事項は実質的な禁止措置に等しい内容であるため、言論の自由の観点から大きな反発にあい議論となっている。

#### (d) 影響

ロシア政府の情報戦は、自国民の支持強化には成功した。肝心の対ウクライナについて、侵攻当初、ロシア政府は大半のウクライナ国民が支持または無関心の態度を取ると想定したと見られるが、彼らは反ロシアに回った。西側諸国の分断に対しても、ロシア政府の思惑は外れた。これらの誤算については、ウクライナ政府のSNS等による情報戦が巧みであったこと、2014年のクリ

ミア侵攻の成功でロシアに楽観が生まれたこと等が指摘されている<sup>\*92</sup>。一方で、西側諸国や中国・ロシア陣営に与しないグローバルサウス諸国の支持については一定の効果があったと考えられている。

### 4.1.4 虚偽を含んだ情報への対応状況

日本では、2022年12月に公表された国家安全保障戦略<sup>\*93</sup>において、「偽情報等の拡散を含め、認知領域における情報戦への対応能力を強化する。その観点から、外国による偽情報等に関する情報の集約・分析、対外発信の強化、政府外の機関との連携の強化等のための新たな体制を政府内に整備する」とされた。これに伴い、陸上自衛隊には認知戦対処専門部隊<sup>\*94</sup>が、海上自衛隊には電子戦や偽情報対策を担う部隊<sup>\*95</sup>が新設されることとなった<sup>\*96</sup>。

#### (1) 日本政府の対応

2023年4月には、内閣官房を中心に政府の偽情報対応のための体制整備が進められることが明らかとなった<sup>\*97</sup>。具体的には、内閣情報調査室の内閣情報集約センターにおいて公開情報の収集・集約・分析を行い、その一環として、内閣情報官のもとで、外国からの偽情報等の収集・集約・分析を実施することとなった。併せて、偽情報等に対する対外発信も強化し、内閣広報官のもとで官邸国際広報室が、国家安全保障局、外務省、防衛省を含む関係省庁と連携して発信を行っていくとされた。

各省の対応としては、防衛省は2022年4月時点で既に新たな役職として、国際情勢を分析し、偽情報対応等に当たる「グローバル戦略情報官」を設置していた<sup>\*98</sup>。AIを使用して公開情報やSNSの投稿を自動収集し、諸外国の意図や影響分析を行っている。外務省もAIを使った偽情報対策に乗り出しており、2024年度に向けた概算要求においては、「認知領域における情報戦に係る本省モニタリング・分析・発信強化」と「国際情勢分析能力強化のためのAI活用」といった事業を盛り込み、合わせて12.6億円を計上している<sup>\*99</sup>。また、総務省は2023年10月に「デジタル空間における情報流通の健全性確保の在り方に関する検討会」を立ち上げ、偽・誤情報対策を含め、デジタル空間における情報流通の健全性確保に向けた今後の対応方針と具体的な方策について検討を進めている<sup>\*100</sup>。

安全保障上の脅威となる国家レベルの情報操作や偽

情報の拡散について、一国のみでの対処では十分でないことから、国際協力が進んでいる。2023年4月29～30日に開催されたG7 デジタル・技術大臣会合においては、「人権、特に表現の自由に対する権利を尊重しつつ、オンラインの情報操作や干渉、偽情報に対処するために、ソーシャルメディアプラットフォーム、市民社会、インターネット技術コミュニティ、学術界を含む幅広いステークホルダーがとる行動の重要性」を認識したとする閣僚宣言が発表された。宣言では、「オンラインの偽情報に対処するための様々なステークホルダーによる既存のプラクティスを『偽情報対策既存プラクティス集 (EPaD)』として収集・編集することに協力し、そしてこの報告書を京都で開催される国連 IGF2023 で公表・発表する」ことも宣言された<sup>\*101</sup>。この宣言に従い、2023年10月8日～12日に国連主催で開催された「インターネット・ガバナンス・フォーラム京都 2023 (IGF 京都 2023)」において、同プラクティス集が公表された<sup>\*102</sup>。

また、2023年12月に日本と米国は協力文書に署名し、外国から情報操作があった場合の検知や情報の交換、収集情報の分析等に連携して対処していくことを確認した<sup>\*103</sup>。

SNS等を運営する企業(プラットフォーム)への対処としては、2024年1月に、プロバイダー責任制限法の改正案が提出された<sup>\*104</sup>。これは、SNSを運営する企業に対し、不適切な投稿の削除の申請があった場合に迅速な対応や削除基準の公表等を義務付けるものである。主な狙いはSNS上の誹謗中傷対応を想定したものであるが、前述のとおりディスインフォメーションのオペレーションにはヘイトスピーチや個人情報のリークによる中傷等も組み合わされることから、こうしたプラットフォーム規制は情報戦対応の第一歩ともなる。

更に、能登半島地震をめぐる偽・誤情報の氾濫を契機として、SNSのプラットフォームが偽・誤情報を判別しやすくするため、政府は「発信者の実在性と信頼性を確保する技術」の開発を支援する方向性を打ち出した<sup>\*105</sup>。具体的には、コンテンツの発信者情報を電子的に付与する「オリジネーター・プロフィール (OP: Originator Profile)」(インターネット上の記事や広告に、第三者機関が認証した発信者情報を電子的に付与し、利用者が信頼性を確認できるようにする技術)の開発や、生成AIの技術で作成したディープフェイク動画を判別できる技術開発を支援することを対策に盛り込んだ。2011年の東日本大震災や2016年の熊本地震等、過去の災害で流布された真偽の判別が難しい情報の特徴を分析し、

プラットフォームと共有する考えも示している。

## (2) 国内のファクトチェックの状況

ファクトチェック関連の動向としては、JFCが、LINE ボットによるファクトチェックの提供サービスを開始した<sup>\*106</sup>。このボットはJFCのデータベースと接続されており、LINE アプリでJFCを友達に追加して質問を送ると、ボットがデータベースを参照して関連すると思われる回答を返す。また、データベース上にファクトチェック情報がない場合は、JFCが調査を行い、関連するファクトチェック記事が公開された時点で、質問者に自動的に記事が送信される仕組みとなっている。LINEという身近なアプリを使用することで、ファクトチェックが手軽に行えるようになり、ファクトチェックという習慣がよりいっそう国民に浸透することが期待される。

「4.1.3 虚偽を含んだ情報生成・拡散の事例」で言及したように、近年のディスインフォメーションには生成AIによって作成された動画像が利用されることも多く、AI利用のガバナンスの在り方が問われている。

## (3) 海外の対応

米国では、2023年10月30日にBiden大統領によってAIの安全性に関する大統領令が公表された<sup>\*107</sup>。この大統領令では、AI開発者に対しては開発情報を連邦政府と共有することや、国家安全保障等に重大なリスクをもたらすAIの開発を行う企業には開発の過程で政府への通知を求める等、安全性に関わる諸基準や共有方法等を定めている。この大統領令の目的の項では、ディスインフォメーションのリスク低減も記載されている。

2023年11月1日にはAIセーフティサミットが英国で開催され、各国政府関係者やAI関連企業が参加した。このサミットでは、参加国29カ国が署名したブレッチリー宣言が公表され<sup>\*108</sup>、AIのリスクにどう取り組むか、世界的な合意に達することを目標に掲げている。併せて、米国及び英国では、政府におけるAIの信頼性、安全性を確保するための専門機関としてAIセーフティ・インスティテュート<sup>\*109</sup>の設置も発表した。このような世界的な流れを受けて、我が国でもAIの安全性の評価手法の研究や規格作成等を行うため、2024年2月14日に「AIセーフティ・インスティテュート (AISI: AI Safety Institute)<sup>\*110</sup>」をIPA内に設置した。

更に2024年2月には、生成AIを悪用したディープフェイクによる動画・音声等が各国選挙に影響を及ぼすことを防ぐため、世界の主要IT企業20社が協業体制を

取ることと合意した<sup>\*111</sup>。OpenAI 社や Microsoft 社、Google 社、Meta 社、X 社といった AI 関連企業、プラットフォーム企業が参画する。動画の出所を明示する「電子透かし」や、SNS 上で偽動画を検出する技術の開発、向上を目指して、この 20 社は協働していくこととなった。

AI のセキュリティについては、「4.2 AI のセキュリティ」も参照されたい。

#### 4.1.5 状況のまとめと今後の見通し

本項では、これまで述べてきた虚偽を含んだ情報の生成・拡散、悪用状況をまとめるとともに、今後の見通しについて整理する。

##### (1) 状況のまとめ

偽・誤情報の生成・拡散は、国家レベルの組織的・政治的なもの、陰謀論・差別・偏見等、社会に根強くあるナラティブの形を借りて虚偽のストーリーが作られるもの、災害・パンデミック・金融不安等の社会不安を契機とする突発的なもの、更にこれらの組み合わせや、これらによって経済的利益を得ようとするもの等が確認できる。そして、近年の IT 基盤の発達がこれらの拡散を容易にしまい、SNS、ターゲティング、レコメンデーション等による情報同質化（フィルターバブル）と増幅（エコーチェンバー）が懸念される拡散の脅威が増大している。

IT サービス提供者の側では、過度のビジネス重視による不正コンテンツや不正アカウントの放置が問題化し、ファクトチェック強化や規制強化が進んでいるものの、生成 AI 等による真偽不明なコンテンツの急増には法整備を始めとした対策が追いついていない。

アクセス数・広告収入増加等を目当てに虚偽と思われる情報を拡散した結果、一定の割合でそれを信じる人が現れたと報告されている<sup>\*112</sup>。また、震災等の特異な状況では、使命感や正義感等の強い感情に駆られて、悪意はなくても不確実情報を拡散してしまうユーザーも多く、IT サービス利用者の側にも、不確実情報の拡散に対する意識向上が求められる。総務省ではデジタル空間の健全性確保の観点から啓発を始めているが、国外からの情報操作に対抗するリテラシー醸成については、安全保障を意識した官民の教育プログラムが十分に整備されていない。

これらの偽・誤情報に関連する社会問題については、情報社会のトラスト形成の観点から課題解決を目指す研究開発プロジェクトが、国立研究開発法人科学技術振

興機構（JST：Japan Science and Technology Agency）内の社会技術研究開発センター（RISTEX：Research Institute of Science and Technology for Society）において 2023 年から開始されている<sup>\*113</sup>。「SDGs の達成に向けた共創的研究開発プログラム（情報社会における社会的側面からのトラスト形成）」と題した同プロジェクトにおける採択研究課題について、開発成果の社会還元、社会実装が期待される。更に、安全保障及び情報リテラシー両面での対策が急務である。

##### (2) 今後の見通し

偽・誤情報の生成・拡散は、社会の分断や対立を求める力が働く限り、今後も継続すると思われる。検討されている対応策を整理すると、以下のようになる。

- 情報操作型サイバー攻撃への対処  
情報騒乱を用いた情報戦に対しては、情報操作型サイバー攻撃を防ぐサイバーセキュリティの文脈での対応が求められるとの指摘もある<sup>\*114</sup>。
- ファクトチェック機能強化  
官民を通じた多角的なファクトチェック体制の構築、急増するコンテンツのチェック自動化等の技術支援等が必要と思われる。また、虚偽情報を常習的に拡散するユーザーの把握や「この話題は虚偽情報が確認されている」というリスク情報の開示も、利用者への注意喚起のために重要である。
- プラットフォーマー規制  
憲法で保障されている表現の自由に十分配慮しながら、犯罪目的・武力行使正当化・差別等の偽情報拡散防止についてはプラットフォームの規制を求める指摘もある<sup>\*114</sup>。
- 利用者のリテラシー向上  
公教育におけるサイバーリテラシー教育の導入、社会人等が積極的に参加できるワークショップの実施等、国民のリテラシーを向上させる教育プログラムの拡充が必要である。その中では、以下のような情報を周知することは意味があると考えられる。
  - エコーチェンバー、フィルターバブル、マイクロターゲティングといった偽・誤情報拡散の仕組み
  - 国家支援勢力による情報戦、認知戦の脅威
  - ファクトチェックの重要性と情報ソースの多様化の方法
  - ナラティブの影響力とナラティブの戦いの脅威情報ソースの多様化については、例えばラテラルリー

ディング（「横読み」ともいう。ブラウザのタブを複数開いて、当該の情報について検索したり、関連する公的機関や主要メディアの発信と比較したり、検証ツールを活用したり、といった複数の情報源にあたること<sup>\*115</sup>）という手法がある。

- ナラティブに基づく拡散対応

情報戦においては、過去の事例やその国の歴史・文化等から利用されやすいナラティブがある程度予測できるため、自国への攻撃に利用されやすいナラティブを事前に把握しておき、反論を用意しそれらを国民に周知徹底しておくといった、政府レベルでの対応策が重要である。これには民間のメディアにも協力を仰ぎ、官民で多角的な情報発信をすることが求められる。

- 生成 AI の利用ルール策定

現在は人権への配慮が注目されているが、生成 AI による虚偽情報を増やさないためにも、利用ルールの早期の策定が望まれる。生成 AI は活用するメリットも大きいため、一律規制ではない対応が重要である。

IT 基盤のサイバーセキュリティ確保が重要な対策であることは言うまでもないが、偽・誤情報を拡散する攻撃主体が AI 技術を悪用した場合、この対処は非常に難しくなるため、特に安全保障の分野では喫緊の課題として警鐘が鳴らされている<sup>\*116</sup>。AI 技術の悪用を防止し、攻撃主体の活動をいかに抑止するかが重要課題となる。



## 4.2 AIのセキュリティ

人工知能 (AI: Artificial Intelligence) は、2010 年代のディープラーニング (深層学習) を代表とする機械学習技術の革新により、システム自動制御、画像診断・自動走行等の分野への応用が進んだ<sup>\*117</sup>。更に 2022 年以降、生成 AI と呼ばれる技術・サービスが急激に普及し、AI の専門知識がなくてもデジタルコンテンツを飛躍的に容易に生成することが可能となった。こうした応用が日常生活や業務・産業の革新を生み出すとの期待も高まっている。

一方で、AI がもたらすリスク (判定や処理の安全性、公平性、遵法性、プライバシー保護等) にも懸念があり、AI のサイバーセキュリティリスクもその中に含まれる。2022 年以降の生成 AI の技術進化と普及のスピードは、それらのリスクがどの程度なのか、どう対処すればいいかの検討が追いつかない、という状況を生み出しつつあり、法制化による AI 規制 (例えば欧州の AI 法<sup>\*118</sup> (「2.2.3 (2) (e) AI 法の成立と実装準備」参照))、あるいは政府による AI ガバナンスの枠組みの議論も始まっている (例えば広島 AI プロセス<sup>\*119</sup> (「2.2.1 (1) (b) AI の軍事利用や安全性に関する国際的な議論」参照))。

「情報セキュリティ白書 2019<sup>\*120</sup>」に「3.5 AI のトラストとセキュリティ」を掲載したが、本節では上記の状況を鑑み、2024 年 4 月時点における AI のサイバーセキュリティリスクの実態と影響、対策の最新動向について解説する。

### 4.2.1 本節で対象とするAIのスコープ

本節で対象とする AI のスコープを説明する。まず AI の定義については、AI を「機械学習に基づく分類・生成機能を備えたソフトウェア」、AI システムを「AI の判定・推論に基づき動作・処理を行うシステム」と規定する。このため従来のエキスパートシステム等の機械学習を伴わない AI は除外される。

AI とセキュリティの関係については、AI のセキュリティ (Security for AI) と AI によるセキュリティ (AI for Security) の二つのトピックがある。本節では AI のセキュリティ、特に「AI の利用に伴うセキュリティ脅威の現状と対応」に議論を絞る。セキュリティ脅威には、① AI システムに対するセキュリティ脅威、② 攻撃者の AI 悪用による他システムへのセキュリティ脅威のいずれも含める。②の脅威は詳述する。またプライバシー侵害に関しては、

サイバーセキュリティインシデントの一環として含めるものとする。

また AI の応用範囲について、本節では軍事利用の詳細な分析はしないものとする。

以下ではまず、2024 年 4 月時点の AI の利用状況を概観した後、そのリスク要因を整理する。次いでサイバーセキュリティリスクの要因を細分化し、その特性と脅威について述べる。最後に、現在取り組まれているベンダー側の AI ガバナンスを主体とするリスクマネジメントの取り組みについても紹介する。

### 4.2.2 AIの利用状況と品質特性

本項では生成 AI を中心として、AI の利用状況、及び品質特性、それらに伴うリスクを概観する。

#### (1) 生成 AI の急速な普及

2022 年 11 月、非営利法人 OpenAI, Inc. が大規模言語モデル<sup>\*121</sup> (LLM: Large Language Model) を用いたチャットサービスである ChatGPT をリリース<sup>\*122</sup>した。ChatGPT は対話形式で幅広い質問に回答できることから急激に普及した。主要 IT ベンダー、スタートアップ各社から、様々な用途に利用できる基盤モデル<sup>\*123</sup> (Foundation model) に基づく多様な生成 AI が相次いで公開されていった。

生成 AI は文章、画像、プログラム等を生成できる AI モデルに基づく AI の総称で、汎用の基盤モデル、あるいは特定用途向けの AI モデルに基づき、コンテンツを生成する。技術面では深層学習 (Deep learning) が用いられ、従来の GAN<sup>\*124</sup> 等とは異なり、Diffusion Model、Transformer 等の手法が日進月歩で開発されている。プログラム、質問回答や文書生成等を支援する ChatGPT、画像生成の Stable Diffusion、音声生成の WaveNet、ElevenLabs 等により、AI スキルのない利用者が、自然言語 (プロンプト) で完成度の高いコンテンツをノーコードやローコードで生成できる。生成 AI によるソフトウェアコード生成も注目され、複雑なコードの作成はまだ難度が高いが、試験的なコード生成等から利用が始まっている。

生成 AI 自身の基盤モデルやチューニングモデルの開発では、Meta 社等が基盤モデルを公開する<sup>\*125</sup>等、

オープンソース化、オープンソースコミュニティによるフィンチューニングが進展している。このように、AIの開発・利用いずれにおいても、一般利用者が容易にアクセスできる状況は「AIの民主化」と表現されることがある<sup>\*126</sup>。AIの民主化は言うまでもなくAI普及の原動力の一つであるが、同時にAIのリスクの統制が難しくなるという問題をはらむ。

## (2) AI品質の特性と新たなリスク

本項では、機械学習を利用したAIシステムの品質面での特性とそれに起因するリスクについて述べる。

機械学習を利用したAIシステムの品質の最大の特性は、それがAIのアルゴリズムの性能だけでなく学習の質・量によって決定される点にある。学習データが量的に不十分である、網羅的でなく偏りがある、ノイズが混入している等の要因でAIの分類精度は劣化してしまう。そして、必要な学習データをすべて用意し、完全にノイズを排除して学習を行うことは現実的ではない。

もう一つの重要な品質特性は、AIの動作を人間が理解することは事実上不可能であり、どんな入力かどのような誤りを生じさせるかも完全には知り得ない、という点である。人間の認知行動からはまったく予期できない間違いが生じ得ることは、例えば画像認識についてよく知られている。

以上の議論から、AIの分類・判定には誤りが含まれる可能性があること、どんな入力に対してどのような誤りが発生するか事前に予見することが難しいこと、がAI固有の品質特性として挙げられる。これらの特性は、AIの判定に基づく処理の安全性、あるいはサイバーセキュリティ上の新たなリスクとなり得る。リスクを増大する悪意のデータ入力・訓練データ改ざん、あるいは不適切なAIの利用によりリスクが増大する可能性は、例えばディープフェイク等の深層学習<sup>\*127</sup>技術を用いた生成AIツールの登場・利用の一般化により、急速に深刻化している<sup>\*128</sup>。以下の各項において、これらのリスクの状況を詳細に見ていく。

### 4.2.3 AIのリスク要因の包括的整理

AIの急速な利用拡大の一方で、AIがもたらすリスクが懸念され、多くの議論やリスク低減施策の検討が進んでいる<sup>\*129</sup>。2010年代より、技術者、研究者、法律家等の専門家により、技術・倫理・制度等の様々な観点から、AIの社会実装に関わるリスクの議論が活発化し<sup>\*130</sup>、

国際標準化、学会や民間によるガイドライン公開、更には2020年以降の欧州AI法の議論を始めとするAIリスクの政府による統制や法規制についての議論が急速に進んだ。

AIリスクの中では、安全（セーフティ）に関するリスクが目ざされている。ここでのセーフティリスクは、AIシステムの誤作動や戦争・テロリズム・犯罪等へのAI悪用による身体的・物理的被害に加え、フェイクコンテンツ・情報暴露等による人権侵害・民主主義への脅威を含む点が重要である。更に、データの知的財産権保護、プライバシー保護、AIシステム・データのセキュリティ、動作に関する説明性・追跡性・透明性の不備・不全等がリスク要因とされている。これらのリスクを包括的に統制する議論が官民で行われている<sup>\*131</sup>。

## (1) 国際標準によるリスク対応の枠組み

2023年12月18日、AIマネジメントに関する国際標準ISO/IEC 42001:2023が発行された<sup>\*132</sup>。同標準は、ISO9001品質マネジメントシステム(QMS)規格やISO/IEC27001情報セキュリティマネジメントシステム(ISMS)規格等と同様のアプローチを採用しつつ、AI固有のリスク管理策として、AIの特質からくる信頼性や透明性、説明責任等におけるリスクの軽減や利用における公平性・プライバシーへの配慮を要求するものである<sup>\*133</sup>。適用対象はAIシステムを開発・提供・利用する組織すべてにわたる。PDCAサイクルに基づくリスク対応に透明性や公平性、プライバシー等のAIの信頼(トラスト)に関わるリスクを組み込んだもので、ISMSの既存の規格と親和性の高い形式である。

## (2) ガイドライン等によるリスク対応の枠組み

2023年1月26日、米国国立標準技術研究所(NIST: National Institute of Standards and Technology)はAIのTrustworthinessを確保するためのリスクマネジメントの枠組みのガイドラインとして「Artificial Intelligence Risk Management Framework (AI RMF 1.0)」(以下、AI RMF)を公開した<sup>\*134</sup>。更に2024年4月29日、AI RMFに基づき、生成AI固有、または生成AIにより増幅されるリスク要因を記載したガイドラインである「NIST AI RMF: Generative AI Profile」(以下、Generative AI Profile)のドラフトを公開した<sup>\*135</sup>。更に2024年5月3日、経済協力開発機構(OECD: Organisation for Economic Co-operation and Development)は、生成AI等の技術進展に対応するため、人権・民主的価値

を尊重し、革新的かつ信頼できる AI 利用の原則として「OECD AI Principles」を改訂した<sup>\*136</sup>。「OECD AI Principles」は、米国、欧州、アジア等で策定される AI 関連ガイドラインの基底となるものとして参照されている。

以下では、上記のうち NIST の二つのガイドライン、及び日本政府の「AI 事業者ガイドライン (第 1.0 版)」に基づいて AI リスク要因と統制を整理する。

#### (a) AI RMF のリスク要因整理

AI RMF では、AI の Trustworthiness を担保するために統制すべきリスク要因として、以下を挙げている。

##### ① 妥当性確認・信頼性 (Valid and Reliable)

AI システムが期待した性能・機能で動作し、その妥当性を客観的に確認できること。なおここでの「信頼性」は、一定の期間・条件のもとで期待した正しさ(精度)で動作することをいう。

##### ② 安全 (Safe)

AI システムの動作により人の身体・健康・資産・環境を危険に晒されないこと。AI システムがサイバーフィジカルシステムである場合、サイバー攻撃が身体的な安全を脅かす等、③のセキュリティと重なりが生じ得る。

##### ③ セキュリティ・レジリエンス (Secure and Resilient)

サイバー攻撃に対して防御・対応すること (セキュリティ)、想定外・予期しないイベントに対して安全に動作すること (レジリエンス)。予期しないイベントには、AI に誤判定を起こさせる「敵対的サンプル」の入力、訓練データにノイズを加えて判定精度を劣化させたり、意図的な誤判定を誘導したりする「データポイズニング」等の攻撃が含まれる。レジリエンスの確保はセキュリティ対策の一環であると同時に、安全性対策とも重なる。

##### ④ アカウンタビリティ・透明性 (Accountable and Transparent)

AI システムに関する情報を開示すること (透明性)、また透明性の担保によって AI システムの出力に関する説明責任を果たすこと (アカウンタビリティ)。透明性には、訓練データの出自に関する追跡可能性も含む。

##### ⑤ 説明可能性と解釈可能性 (Explainable and Interpretable)

AI システムの動作の仕組み、AI システム出力の意味を説明し、利用者の「AI が何をしているのか」に関する理解の不足に起因するリスクを低減すること。

##### ⑥ プライバシー強化 (Privacy-enhanced)

AI システムが扱う個人情報の匿名化、頑健な ID 管

理、人権保護、情報開示制限等を行うこと。

##### ⑦ 公平性 - 有害なバイアスのマネジメント (Fair-with harmful bias managed)

AI システムの判定が有害なバイアス・差別等を含まず平等・公正であること。公平性は複雑であり、例えば人口学的な分布に基づく判定は、障害を持った人を他の人と平等な条件で必ずしも判定しない等の難しさがある。

以上で見たように、AI のリスク要因はセキュリティにとどまらず多様である。上記分類では、②安全は③セキュリティと同じレベルのカテゴリとなっているが、特に③のレジリエンスに関わるリスクでは重なりが生じ得る。また、次の「4.2.3 (2) (b) Generative AI Profile のリスク要因整理」で述べるように、生成 AI についてはプロンプトの脆弱性を突いた攻撃 (悪意の質問) や AI の悪用により、新たなリスクが高まる可能性がある。

#### (b) Generative AI Profile のリスク要因整理

Generative AI Profile は AI RMF の実装について、事例を生成 AI に特化し、分野を横断する共通プロファイルとして記載しており、生成 AI 固有、または生成 AI で増大するリスク要因を以下のようにまとめている。

- ① 化学・生物・放射性・核関連兵器に関する情報へのアクセス容易化
- ② 自信を持った表現による誤り・虚偽の記載 (ハルシネーション)
- ③ 危険・暴力的で犯罪や不正につながる推奨
- ④ プライバシー漏えい (生体情報・位置情報・個人識別情報等)
- ⑤ 環境 (生成 AI 訓練用の資源増大による環境劣化)
- ⑥ 人間と AI の構成 (設定・インタラクション不備による不適切な AI 利用)
- ⑦ 情報の一貫性 (真偽不明の情報・誤情報・虚偽情報の拡散)
- ⑧ 情報セキュリティ (サイバー攻撃、ウイルス<sup>\*137</sup> 生成、フィッシング等の容易化)
- ⑨ 知的財産 (権利処理のないコンテンツ利用、営業秘密の開示)
- ⑩ わいせつ・権利を侵害するコンテンツ
- ⑪ 中毒性・ヘイト・偏見コンテンツ等への一方的アクセス
- ⑫ バリューチェーン・部品統合 (サプライチェーン上の不十分な品質チェック)

上記のうち⑥には、AIとのインタラクション不備による誤用・想定された目的以外の利用(悪用)が含まれ、⑧のサイバーセキュリティリスクとも大きく関わる。悪用は生成AIに限定されるものではないが、後述のIPAの米国調査によれば、例えばフィッシングは、生成AIで明らかにリスクが増大していると思われる。また④のプライバシー漏えい、⑦の情報の一貫性、⑫のサプライチェーンのチェック不備等もセキュリティと重なるリスクと考えられる。このうち⑫のサプライチェーンリスクは生成AIに限られるものではないが、オープンソースソフトウェア(OSS: Open Source Software)調達等が生成AIに顕著な課題となる可能性はある。

### (c) 国内のAIガイドラインによるリスク統制

2024年4月19日、内閣府、総務省、経済産業省は省別にまとめられていたAI開発・利活用のガイドラインを統合した「AI事業者ガイドライン(第1.0版)」を公開した。同ガイドラインも「OECD AI Principles」に基づくリスク統制を志向しており、以下の統制項目を挙げている。

- ①人間中心(人権・人間の尊厳や虚偽情報の統制)
- ②安全性
- ③公平性
- ④プライバシー保護
- ⑤セキュリティ確保
- ⑥透明性
- ⑦アカウントビリティ
- ⑧教育・リテラシー
- ⑨公正競争確保
- ⑩イノベーション

同ガイドラインとAI RMFを比較すると、いずれも「OECD AI Principles」に準拠し、リスクベースのAIガバナンス及びマネジメントを目標とするが、AI RMFに比べ、同ガイドラインの項目①⑧⑨⑩のガバナンス及びマネジメントはやや広めであると思われる。また同ガイドラインは生成AIを想定した「高度なAIシステム」の統制のため、内部テストの実施、透明性の確保等が記載されている。ただし欧州AI法の「ハイリスクAI」(「2.2.3(2)(e) AI法の成立と実装準備」参照)と異なり、規制ではない。AI RMFとの詳細比較(クロスウォーク)は、2024年2月、日米それぞれに設置されたAIの安全性に取り組む「AIセーフティ・インスティテュート(AISI: AI Safety Institute)<sup>\*138</sup>」が共同で実施しており、最初の結果は2024年4月30日に公開された<sup>\*139</sup>。

なお「AI事業者ガイドライン」では、AIサービス提供者は安全な利用のための情報をステークホルダに提供する、AI利用者はAI提供者が想定した範囲内で利用する、としている。「4.2.3(2)(b) Generative AI Profileのリスク要因整理」の⑧で見たとおり、AIサービス提供者の想定を逸脱した誤用・悪用は大きなセキュリティ脅威でもあり、セキュリティの面からも情報の適切な提供とそれに基づいた利用が望まれる。

## 4.2.4 AIのサイバーセキュリティリスク認知状況

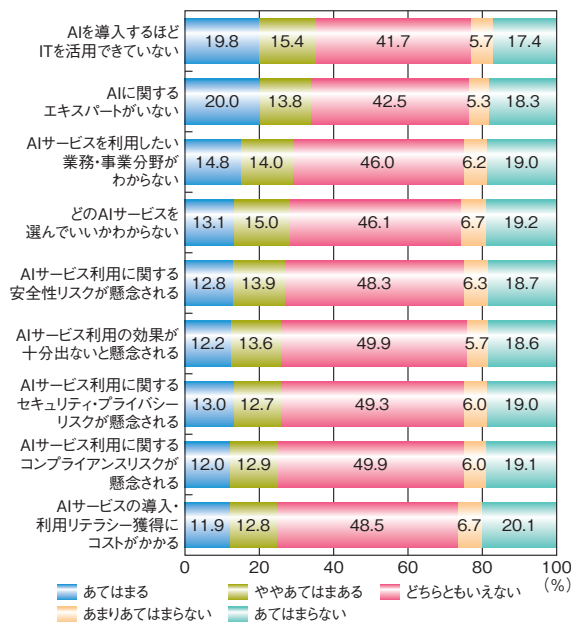
前項で、「Trustworthy AIを実現するために統制すべきリスク」を包括的に整理した。本項では、前項におけるサイバーセキュリティリスクの議論を踏まえ、AIシステムのサイバーセキュリティリスク要因とその脅威について分析を行う。

生成AIの急速な普及が言われているが、2024年4月時点でどんなAIサービスが利用され、利用者のセキュリティリスク認知はどの程度であろうか。IPAは2024年3月、AIサービスの業務利用、及び組織の「AI利用時のセキュリティ脅威・リスク調査」(以下、IPA国内調査)を実施した<sup>\*140</sup>。同調査は、国内企業、組織の役員4,941人からAIサービスの利用実態について回答を得た。更に回答者の中からAIサービスを業務で「利用している/許可している」「予定している」と回答した役員(以下、IT実務者)1,000人(就業301人以上の大企業466人、300人以下の中小企業534人)を抽出し、AIサービス選定時にセキュリティが意識されているか、企業、組織内のAI利用の規定整備や体制づくりができていないかを調査した。回答者全体の中で、AIサービスを業務で「既に利用している/許可している」あるいは「予定がある」とする回答者の割合は22.5%であった。

### (1) AI業務利用の状況

IPA国内調査で「いずれのAIサービスも利用しない/許可しない、予定もない」と回答した3,827人に、企業、組織がAIを利用しない/許可しない理由を尋ねたところ、「AIを導入するほどITを活用できていない」「AIに関するエキスパートがいない」という選択肢に「あてはまる」「ややあてはまる」と回答した合計の割合が上位であり、セキュリティ・プライバシー等のリスクは相対的には下位であった(次ページ図4-2-1)。

IT実務者にAI利用分野別に利用開始/許可時期を尋ねたところ、全体として、「今後利用/許可予定で

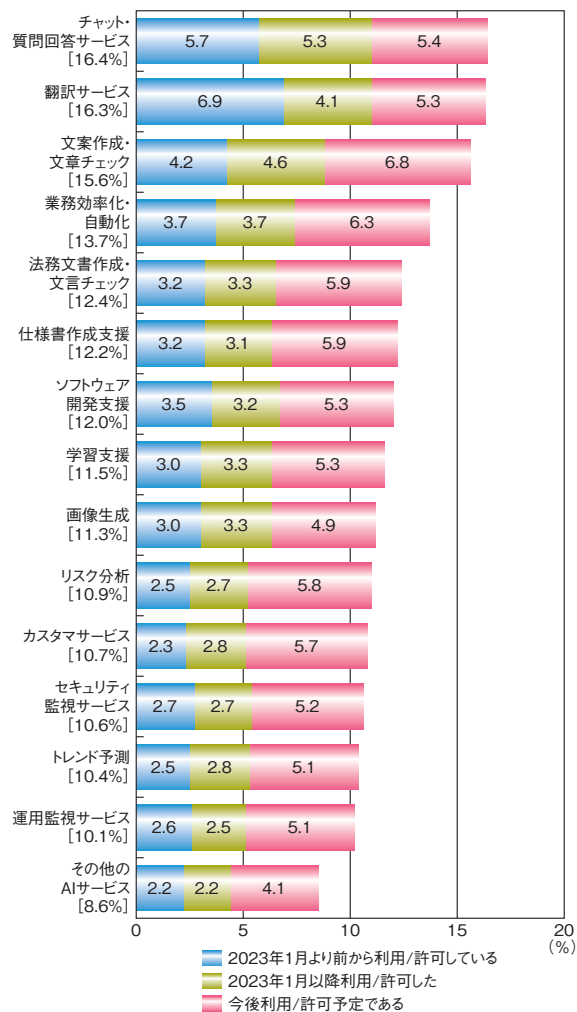


■ 図 4-2-1 企業、組織が AI を業務で利用しない／許可しない理由 (n=3,827)  
(出典)IPA 国内調査を基に編集

ある」との回答が最も多く、「2023年1月以降利用／許可した」と「2023年1月より前から利用／許可している」の回答はほぼ同等であった(図 4-2-2)。分野別では「チャット・質問回答サービス」「翻訳サービス」「文案作成・文章チェック」「業務効率化・自動化」が上位であった。生成 AI の普及に合わせ、急速に利用が広がっていることがうかがわれるが、従来の「業務効率化・自動化」等も今後の利用の拡大が見込まれることが分かる。

## (2) AI 導入で重要な評価尺度

AI の導入可否を決める場合に重要な評価尺度を尋ねた。AI の応用によって尺度が異なる可能性を考慮し、コンテンツ生成を主目的とする「生成 AI」、分類を主目的とする「分類 AI」それぞれについて質問したが、結果としては大きな差はなかった(次ページ図 4-2-3、図 4-2-4)。いずれも、「非常に重要だと思う」と「やや重要だと思う」を合計した割合を見ると、「出力の正確性」や「作業効率」に続いて「セキュリティ対策」(次ページ図 4-2-3、図 4-2-4 の赤枠)とプライバシー保護に関係した項目が重要な評価尺度となっていることが分かる。「学習の公平性」や「学習における法務・知財処理」「学習の網羅性」等がそれに続くが、透明性に関する「出力をどう生成したかの説明」「学習の追跡性」等は「生成 AI」「分類 AI」どちらも下位となっている。なお、セキュリティが重要な尺度とする回答には大企業と中小企業の差は見られなかった。



■ 図 4-2-2 企業が AI の利用／許可を開始した時期 (n=1,114)  
(出典)IPA 国内調査を基に編集

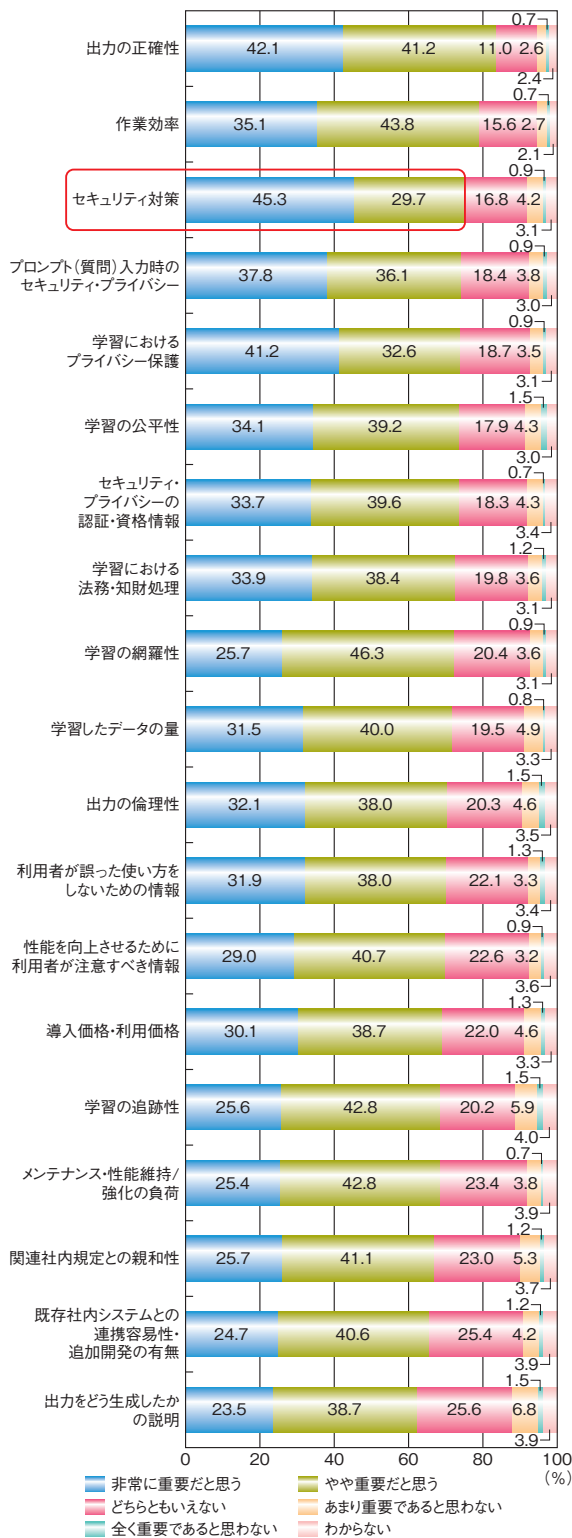
## (3) AI に関するセキュリティ施策の状況

企業、組織が AI について実施しているセキュリティ施策について尋ねた結果を図 4-2-5(次々ページ)に示す。

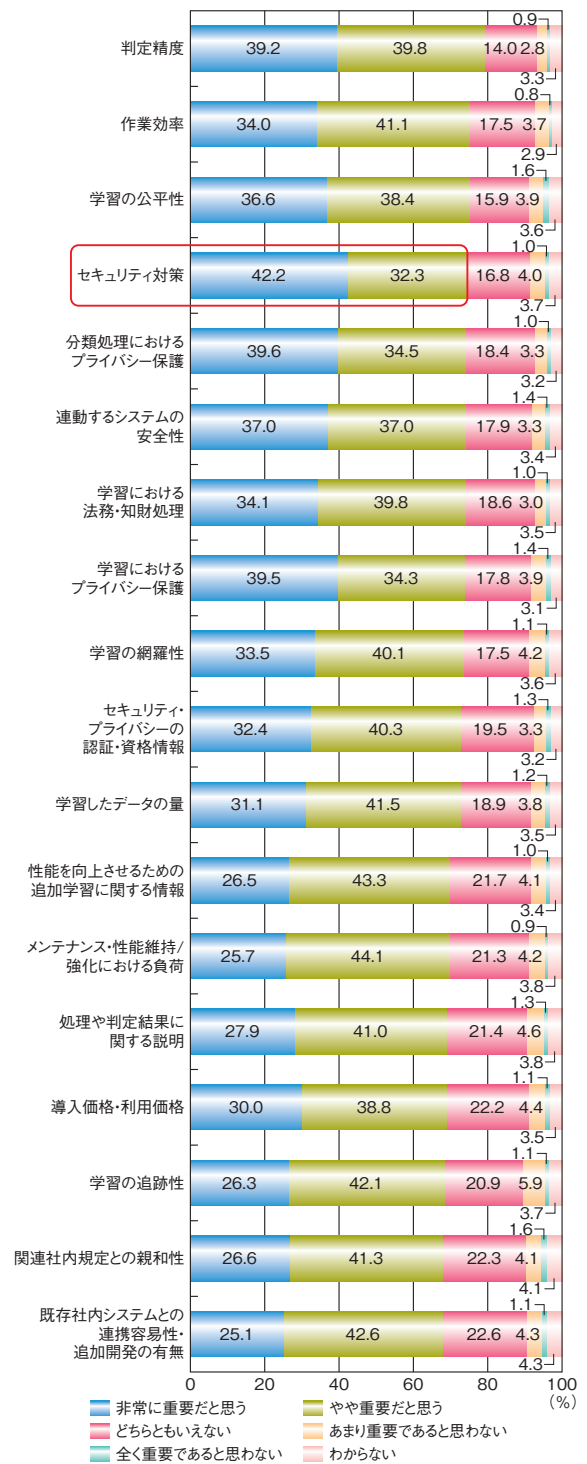
図 4-2-5(次々ページ)の選択肢は、セキュリティ施策に関わる要件で、赤枠が「分類 AI」、青枠が「生成 AI」に求められる要件である(中央の「社内システムの連携におけるセキュリティ要件」は双方に重なる)。若干ではあるが、「分類 AI」での対応が先行しているが「検討している」を含めても 40% 超にとどまっており、対応できているとは言い難い。特に、急速に広がっている生成 AI のプロンプト入力・管理等の規則化が 12% 程度にとどまっているのは課題と言える。

## (4) AI のセキュリティ脅威の認識

所属する企業にとってのセキュリティ脅威の大きさの認識について尋ねた結果を図 4-2-6(次々ページ)に示す。本設問でも「分類 AI」(図の赤枠)、「生成 AI」(図の



■ 図 4-2-3 企業が AI の導入可否において重視する尺度 (生成 AI) (n=1,000)  
(出典)IPA 国内調査を基に編集



■ 図 4-2-4 企業が AI の導入可否において重視する尺度 (分類 AI) (n=1,000)  
(出典)IPA 国内調査を基に編集

青粋) 双方に関わる脅威を挙げたが、個々の種別による大きな差は見られていない。これは、AIに関わるセキュリティ脅威が実際どのようなものか、経験が浅く実感されていない可能性を示している。

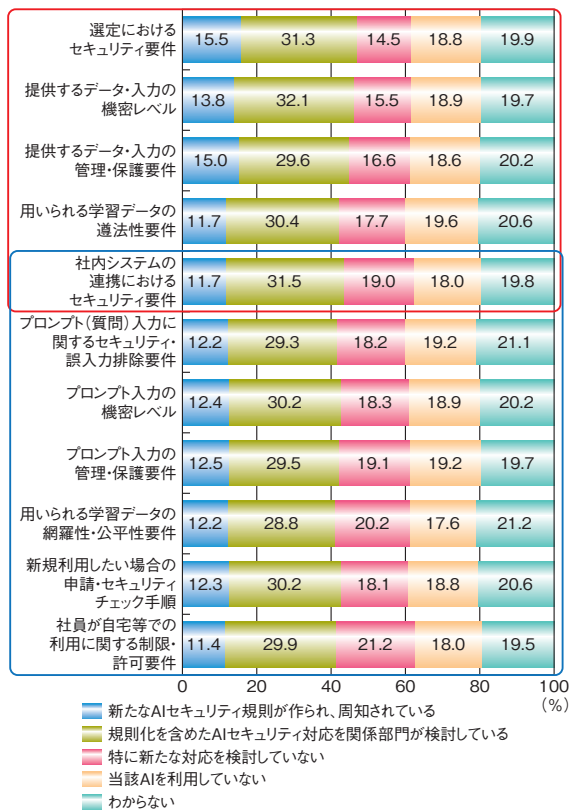


図 4-2-5 企業、組織が対応しているセキュリティ施策 (n=1,000)  
(出典)IPA 国内調査を基に編集

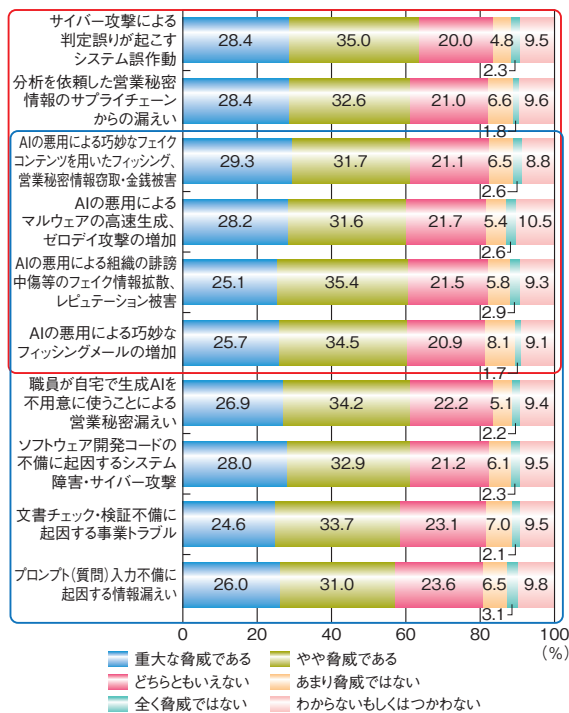


図 4-2-6 AIセキュリティ脅威の大きさの認知状況 (n=1,000)  
(出典)IPA 国内調査を基に編集

#### 4.2.5 AIのサイバーセキュリティリスクの分類

前項の調査では、国内の業務向け AI システムの利用者に関しては、AI のサイバーセキュリティリスクは重要という意識が強いものの脅威の認知やセキュリティ対応はこれから、という状況がうかがわれた。国内調査と並行して、IPA は 2024 年 2 ～ 3 月、米国における AI システムのサイバーセキュリティリスクの認知状況に関する調査(以下、IPA 米国調査)を行い、この中で有識者インタビューを実施した<sup>\*141</sup>。

本項では、IPA 米国調査を踏まえ、AI システムに関わるサイバーセキュリティリスクの詳細な分析を以下の三つの類型に分けて行う。

##### ① AI の悪用

代表例としては、無制限な軍事利用、暴力・犯罪・サイバー攻撃への悪用、フェイクコンテンツ拡散による権利侵害・対立扇動、悪意のボット・ウイルス生成等がある。

##### ② AI の機能特性による誤判断・誤用

代表例としては、誤判断による自動走行事故、生成 AI の不適切な学習やプロンプト処理による差別的回答・情報漏えい・誤情報拡散等がある。

##### ③ AI の機能特性を突いた攻撃脅威

AI モデルの特性により、特定のデータやパターンを含む入力に対して間違った予測・分類結果が出力される場合がある。こうした AI を間違えさせる入力は「敵対的サンプル (Adversarial example)」と呼ばれ、AI モデルの特性を調べる手段となっているが、これを逆手に取った多くの攻撃手法が存在する。一方、学習時点において学習データにノイズを混入させ、AI モデルの性能を劣化させる、あるいは意図的な誤判定を誘導する攻撃 (データポイズニング) もよく知られている。

以下では、前述の AI システムに関わるサイバーセキュリティリスクのそれぞれの実態と対応について検討する。

##### (1) AI の悪用

本節では、特にサイバー攻撃への悪用に注目する。AI のサイバー攻撃への悪用は、大別して①システムの脆弱性を突くもの、②人間の脆弱性を突くもの、に分けられる。このうち①は、AI を用いたセキュリティ対策と表裏一体の側面がある。すなわち、以下はサイバーセキュリティリスクの把握・対策に不可欠であり、対策強化の

ために AI の活用が望まれるが、こうした活用はシステムを狙う攻撃者にもメリットをもたらし、攻撃高度化・効率化のための悪用が懸念される。

- ソフトウェア記述不備等の脆弱性検出・リスク評価
- システム設定不備等の脆弱性検出・リスク評価
- 脆弱性を突いた攻撃予測 等

更に、新たな脆弱性に対するウイルス作成支援（または自動化）も警戒すべき AI 悪用の一つである。

一方②の人間に対する攻撃では、生成 AI により、虚偽を含むもっともらしい文章・動画・音声等の生成が飛躍的に容易化し、詐欺的攻撃への生成 AI の悪用が非常に懸念される。

以下では、代表的な悪用の類型ごとに説明する。

#### (a) 主流となるサイバー攻撃の準備支援(システムの脆弱性を突く悪用)

ランサムウェアや標的型攻撃等、現在の主流であるサイバー攻撃の準備、特に攻撃対象の情報収集・脆弱性発見を支援する AI の利用は、スキルの高くないサイバー攻撃者を助けるものとして懸念される。IPA 米国調査では、①公開情報収集・分析(OSINT)の効率化・精緻化、②脆弱性探索活動の自動化、③内部侵入のためのフィッシングメール作成の効率化・精緻化等が想定されるとしている。このうち③は次の「4.2.5(1)(c)フィッシング(人間の脆弱性を突く悪用)」で述べるとおり、現実には起こっていると思われる。①②がどこまで行われているかを知るのは難しいが、IPA 米国調査では、スキルのあまりない攻撃者が AI を利用して攻撃のコストを下げるのが先行している、という有識者の意見が見られた。一方で、重要インフラ等で高度な攻撃を仕掛ける攻撃者が、既存 LLM のチューニング、あるいはジェイルブレイク(遵法用フィルタの回避)により AI で新たな攻撃手法やウイルスを作成する、という意見もあるが、調査時点では既存攻撃のコスト削減が悪用の中心と見られる。英国サイバーセキュリティセンター(NCSC: National Cyber Security Centre)は、AI 悪用による攻撃の規模とインパクトの拡大が今後2年以内に起こる、と見ている<sup>\*142</sup>。

#### (b) ウイルス生成(システムの脆弱性を突く悪用)

2023年8月、ウイルス生成を目的とする悪意の AI モデルの存在が報告された<sup>\*143</sup>。このうち Wolf GPT は秘匿性の高い暗号化ウイルス生成や高度なフィッシングを目的とした Python ベースの ChatGPT 代替ツール、

XXXGPT はボットネットや遠隔操作ウイルス(RAT: Remote Access Trojan)、ATM用ウイルス作成キット等のコード生成を目的とする ChatGPT 代替ツールである。このようなツールがアンダーグラウンドコミュニティでどのように利用されているかは不明であり、実際のウイルスが生成 AI により作成されたか判定することは難しい。しかし、コーディングスキルに乏しい攻撃者を助け、攻撃コストを下けていることは想定できる。

一方 IPA 米国調査では、AI によるウイルス生成の脅威は、フィッシングやフェイクコンテンツの脅威に比べるとまだ大きくない(次に来る脅威である)という意見も聞かれた。生成 AI はコード生成の試行に適しているが、完成度の高いコードを作るにはまだ人手を必要とする。ただし、その工数は確実に削減されているとする意見もある。2024年4月時点では、AI の悪用は、システムより人間を標的とするもののほうが脅威だ、と認知されていると思われる。

#### (c) フィッシング(人間の脆弱性を突く悪用)

実際のフィッシングメールが生成 AI で作成されたかどうかの検証は難しい。しかし、IPA 国内調査とは別途行った国内有識者インタビュー、IPA 米国調査の有識者インタビューにおいては、フィッシングへの悪用は現実である、とする意見が大半である。2023年7月には、フィッシングを目的とした AI モデル FraudGPT が登場する<sup>\*144</sup>等、悪意の AI モデルも使われていると考えられる。

IPA 米国調査で公開された AI セキュリティ脅威とリスクのチャート<sup>\*145</sup>では、フィッシングへの悪用は現実には起こっており、かつ影響は大きい、として最大レベルの脅威とされた。IPA 国内調査とは別途行った国内有識者へのインタビューでも、つたない日本語が含まれていたフィッシングメールの文章が非常に流暢になっている、生成 AI は海外攻撃者に使われているという意見がある。

#### (d) フェイクコンテンツ生成・拡散(人間の脆弱性を突く悪用)

2018年のディープフェイク技術の悪用を始めて、国家が支援する活動家や協力者等が特定個人・組織を攻撃する生成 AI コンテンツを SNS に拡散させ続けている。例えば2023年10月、Barack Obama 元大統領の新たなフェイク動画が SNS サービス TikTok 上で拡散した。フェイク動画内の家族の料理人の突然死に関する本人に酷似した弁護の音声は、攻撃対象者の音声合成サービスを利用したものであることが判明してい



る<sup>\*146</sup>。「4.2.2(1)生成AIの急速な普及」で述べたディープフェイク発展技術とLLMが容易に利用できる等、AIの民主化の負の影響が背景となり、ロシア・ウクライナ戦争、イスラエル・ハマス間の武力衝突等でもフェイクコンテンツが拡散され続けている<sup>\*147</sup>(事例については「4.1.3虚偽を含んだ情報生成・拡散の事例」参照)。

フェイク文章生成については、前項でフィッシングへの悪用が大きな脅威であることを確認した。2023年7月、フランスのスタートアップ企業 Mithril Security はセキュリティ研究目的として、実験ツール PoisonGPT を公開した<sup>\*148</sup>。PoisonGPT は「悪意の AI モデルに気付かない利用者がフェイクニュースをどう生成・拡散させるか調べる」という調査意図が隠されたまま公開されたが、40セットがダウンロードされた後、PoisonGPT が公開されていた機械学習に関する共同作業プラットフォームである Web サイト Hugging Face<sup>\*149</sup> の運営者により利用規約違反として削除された。Mithril Security はこの物議をかもした調査について、AI モデルのサプライチェーン上の追跡性が重要であると述べている。

IPA 米国調査では、前述の政治的なフェイクコンテンツ拡散の脅威については「深刻である」とする意見と「影響はあるが、そこまで深刻ではない(一見してフェイクと分かるものが多い)」とする意見に分かれ、評価は定まっていない。ただし、2024年は世界的な「選挙イヤー」であることから、政党支持者・敵対的国家・分断で利益を得る組織等による選挙妨害・世論分断等の影響が懸念されており、SNS 事業者のフェイク対策が十分でない、との不安の声もある<sup>\*150</sup>。特に米国では、2024年11月の大統領選挙に向けた選挙妨害・世論分断工作等が懸念され、CISA は国家支援活動家の生成 AI 悪用についてまとめ、注意喚起している<sup>\*151</sup>。その手口は候補者のフェイク音声・動画、選挙スタッフの音声詐称、選挙事務所へのフィッシング、選挙インフラベンダーのなりすまし等多岐にわたる。2024年4月時点で、フェイクコンテンツ検知技術や生成 AI マーク埋め込み等の抑止技術、その実施体制は確立できていない。一時的にフェイクコンテンツを検知できてもすぐに新たな技術で生成・拡散されることも予想され、根本的な解決は難しいと考えられる。大量のフェイクコンテンツ拡散を前提として、大統領選挙で米国民がどのように対応するか、注目される。

なお、フェイクコンテンツ検出技術については国内でも研究と普及が進められている<sup>\*152</sup>。

## (2) AI の機能特性・学習不備による誤判定・誤用

「4.2.2(2) AI 品質の特性と新たなリスク」とで見たとおり、機械学習に基づく AI モデルの判定処理は分析が難しく、その統計的性質により誤判定はゼロにならない。またどのような入力に対し、どのような誤判定が起こるかの予測も難しい。こうした機能特性は AI モデルの「脆弱性」とも考えられ、誤判定に起因するインシデント、あるいは誤判定を狙った攻撃が脅威になり得る。

誤判定による直接的なインシデントは、サイバーフィジカルシステムへの AI 応用ではセーフティ(安全)の脅威(以下、セーフティ脅威)につながる場合もあるが、以下ではセーフティ脅威も、AI システムの一貫性に不具合が生じたセキュリティ脅威の発現類型と見て、検討に含める。

### (a) AI 誤判断によるセーフティ脅威

この脅威の中では、① AI で制御される製造ラインの誤判断による異常動作、② AI で制御される自律型ロボットの誤判断による異常動作、③自動運転車の誤判断による事故等の類型が代表例として考えられる。類型①②は場所(工場や利用エリア等)が特定され、AI システム利用環境のコントロール(入力のコントロール)がやりやすきことが多く、異常動作時にはフェールセーフ原則に基づき停止させれば大きな脅威にはなりにくいと思われる<sup>\*153</sup>。一方で、類型③は①②に比べ利用環境の自由度が高く、映像等の入力データもコントロールできないため、事前学習が行き届かず、誤判定が大きな脅威になり得る。

2023年10月2日、米国サンフランシスコで自動運転タクシーが隣を走行する車にはねられた女性をひく事故が発生した<sup>\*154</sup>。当該自動運転車は救急車の到着までひいた女性の上に停車したままであった。現実の大都市での走行を想定すれば、あり得ないことではない。様々な事故態様への準備が必要となる自動運転車の学習の難しさが顕在化することとなった。

日本においては2023年10月29日、福井県永平寺町にて特定条件のもとでの完全自動運転「レベル4<sup>\*155</sup>」で運行していた観光用自動運転車が自転車と接触事故を起こした(低速であり、けがはなかった)<sup>\*156</sup>。真後ろから見た特殊形状の自転車等、障害物検知学習が不足していたとされ、追加学習とスタッフの添乗によりサービスは2024年3月に再開された<sup>\*157</sup>。非常に限定された走行条件ながら、こちらも学習の課題が顕在化した。

以上の2例から、類型③のセーフティ脅威の削減に

は想定利用環境で起こり得るインシデント、及びインシデント対応で学習すべきデータの明確化が必要であると思われる。また、想定ケースに適合した学習データで、想定ケースについて網羅的に学習が行われたことの検証も求められる。

### (b) 生成 AI との不適切な質問応答

生成 AI への質問（プロンプト）は自然言語で行われ、生成 AI は質問の都度最も妥当と判定された回答を返す。質問によっては回答がセキュリティや、プライバシー、コンプライアンス上問題となることがある。以下ではその主要な類型を見ていく。

#### (ア) 不適切な入力による情報漏えい

質問として入力された文は、生成 AI 基盤モデルの学習に利用されることがある。履歴が AI サービス事業者側に残る入力データは事業者が保持し、以降のサービスで回答に使われる可能性がある。

2023年4月、Samsung Electronics Co., Ltd.（以下、Samsung社）の半導体部門の技術者が、ソースコード開発に ChatGPT を利用した際、製造技術に関わる営業秘密データを不用意に入力したと報じられた<sup>\*158</sup>。このとき ChatGPT への入力は OpenAI 社が保持する設定であり、ChatGPT がその情報を使って回答を生成できる状況となった。同年5月、Samsung社は ChatGPT の業務利用を禁じた。また米国では、API キーが埋め込まれたソースコードを開発者が ChatGPT に送った事案も報告されている<sup>\*159</sup>。

「4.2.4 AI のサイバーセキュリティリスク認知状況」の IPA 国内調査で見たように、生成 AI を業務利用する企業はまだ少数であり、導入においてはセキュリティが最大の懸念点の一つとなっている。Samsung 社の事案は生成 AI のセキュリティ対策を企業が考える契機となったと思われ、導入を検討する企業に影響を与えた可能性もある。

プロンプト入力における個人情報の扱いについては、AI ベンダー・サービス事業者への注意喚起が行われている。2023年6月2日、個人情報保護委員会は生成 AI サービス利用に関する注意喚起等を公開<sup>\*160</sup>し、個人情報保護を取り扱う事業者が生成 AI を利用する場合、プロンプトで入力する個人情報は必要最小限とすること、本人の同意を得ない個人データ入力が想定される場合は当該データを機械学習に利用しないよう確認することを注意した。また ChatGPT については OpenAI 社

に対し、本人の同意を得ないまま要配慮個人情報取得しないこと、取得する個人情報の利用目的を本人に日本語で通知または公表することを注意した。

図 4-2-6 (p.230) によれば、「プロンプト（質問）入力不備に起因する情報漏えい」が脅威であると回答している割合は、他のリスクを脅威と回答している割合と比べ、特段に多くなってはいない。これは、生成 AI の利用がまだ限定的で、営業秘密・個人情報を扱う社内業務への導入がためられる、という国内企業の状況を示している可能性がある。一方で、2023年後半以降、主要な生成 AI ベンダー・サービス事業者はビジネス向けの有償サービス等において、営業秘密・個人情報等の①機械学習への利用除外、②当該情報の管理と AI モデルの分離等を含む生成 AI のセキュリティ対策を強化しており<sup>\*161</sup>、今後利用が進むものと思われる。

#### (イ) 生成 AI 出力の脆弱性とその不適切な利用

大まかなソフトウェア仕様を自然言語で入力し、コードを試験的に生成させる等、ソフトウェア開発効率化のための生成 AI 利用が IT 事業者の間で広まっている<sup>\*161</sup>。複雑なコードの自動生成は難しいものの、試行錯誤的に行うコードを試す等のいわゆる仕様検討の「壁打ち」としての使い方が有効と見られるが、セキュリティ面では以下の課題がある。

- 学習データに含まれるコードに脆弱性がある場合、それが生成コードに反映されるリスクがある。
- スキルの低い開発者が生成 AI に頼りコードを作成した場合や、工数がひっ迫した状況で生成 AI に頼りコードを作成した場合等には、脆弱性を含んだコードが拡散するリスクがある。

上記リスクについて、実際に脆弱な生成 AI 由来のコードが拡散した事案は報告されていないが、GitHub 等の OSS 共有サイトで脆弱なコードが拡散してしまうリスクは IPA 米国調査で指摘され、現実に行き得る、とされている。また、生成 AI を利用した生成コードはセキュリティが弱い傾向にある、との実験報告もある<sup>\*162</sup>。生成 AI の利用範囲やコード共有時のルールの規定等、IT ベンダーや内製しているユーザー企業の対応が必要と思われる。

#### (3) AI の機能特性を突いた攻撃脅威

AI アルゴリズム自身の特性（あるいは不備）を悪用し、意図的に誤判定を起こさせることは、AI システムの性能

劣化、あるいは意図的な誤判定によるインシデント発生を目的とする攻撃の手段となり得る。また、AIシステムの運用・サービスにおいて、学習データの信頼性をサプライチェーンにわたってどう担保するか、データへの攻撃も重要な課題となる。以下では主要な脅威の類型について見ていく。より詳細な手法については人工知能学会の解説<sup>\*163</sup>、総務省とセキュリティベンダーによる手法のまとめ<sup>\*164</sup>等を参照されたい。

#### (a) プロンプトインジェクション

対話型生成 AI においては、不適切な応答が生成されることのないように様々な安全策が施されている。しかし、それらの裏をかくようなテキストを入力することによって安全策を回避し、詐欺や犯罪等を助長するメッセージ等を回答させる攻撃手法があり、これをプロンプトインジェクションと呼ぶ。自動車ディーラーに対する巧みなプロンプトインジェクションにより、わずか1ドルで新車を販売する約束を取り付けてしまうという事例も発生している。

#### (b) 敵対的サンプル

AI モデルに意図的に誤判定を起こさせる手段として、AI の判定動作を錯誤させる入力調べられている。これは敵対的サンプル(Adversarial sample)と呼ばれる。

人間が知覚できないノイズをデータに含め、人間には予想できない誤判定を起こさせる攻撃が、画像認識等の分野で研究されている<sup>\*165</sup>。AI モデル内で処理される入力データがノイズで変化して、分類されるべきカテゴリから遠くなり、誤判定に至るとされる。誤判定を起こさせる画像の特定やノイズ混入データの作成には攻撃対象の AI モデルに関する知識が要求され、容易でないと思われるが、自動走行を想定した研究事例では、交通標識にテープを貼ることで物理的に「ノイズ」を与え、標識の認識を失敗させるとことが可能であるという報告がある<sup>\*166</sup>。AI の利用場面によっては攻撃が容易にできる可能性があり、注意が必要である。

#### (c) モデルインバージョン

入力に対する判定の確からしさ(確信度)が出力として得られる場合、入力を調整して確信度を高め学習データを推定する攻撃手法はモデルインバージョン(Model inversion)と呼ばれる。顔認証システムで用いられる学習用顔画像の推定事例が、プライバシー侵害の可能性のあることからよく知られている<sup>\*167</sup>。ただし、入力を繰り返して調整を効率よく行う攻撃には周到な準備が必要

と考えられる。

#### (d) 推論攻撃

データベースに対して複数の検索を行い、検索された情報の対象者(個人)を特定する等の攻撃は推論攻撃と呼ばれるが、生成 AI への複数の質問によって同様な推論攻撃が可能となるリスクがある。例えば、特定のデータが学習データに含まれているかどうかを間接的な質問を繰り返すことで特定する攻撃が考えられる。2024年4月時点で、生成 AI の個々の質問応答は、不適切な回答がないようチューニングされつつあるが、複数質問により個人情報や犯罪関連情報(武器製造法等)を推定するリスクがどの程度かはまだ自明でない<sup>\*161</sup>。

#### (e) バックドア

AI モデル自体に細工(バックドア)を施し、特定のノイズやパターンデータ(トリガーと呼ぶ)等を含む入力に対して誤判定をさせる攻撃はバックドア攻撃と呼ばれる。バックドアは例えば、トリガーを入れたデータを事前に学習させて AI モデルを変化させておくことで用意される。大規模なデータを学習させる必要がある場合、学習データにバックドアがないことの検証は難しくなる。また AI モデルが公開・共有される場合、バックドアが仕掛けられるリスクが生じる。バックドアの仕掛けられた AI モデルは実際に共有サイトで確認されているという<sup>\*168</sup>。

#### (f) データポイズニング

ノイズを加えた大量のデータで学習を行わせ、AI モデルの判定を意図的に誤らせる、あるいは判定性能を劣化させる攻撃をデータポイズニングと呼ぶ。代表的な手法には、教師あり学習において、誤ったラベルを付けた学習データを混入させる「ラベルポイズニング」等がある。スパムメールフィルタの分類器の学習データやネットワークトラフィック分類システムの学習データを汚染する等のセキュリティ分野の AI システムの攻撃事例があげられており<sup>\*169</sup>、IPA 米国調査においては、現実に関わり得る脅威としてリスクが高いとされている。

データポイズニングを大量のデータに対して行うことは、攻撃対象の学習等に関わる内部不正があった場合、容易になる可能性があるが、AI モデルや学習データへの攻撃について、国内では「内部不正は脅威」とする意見と、「実施コストが高く、違う攻撃手法をとるのでは」という意見がある<sup>\*161</sup>。

上記 (a) ~ (f) の攻撃類型を含め、AI システムが誤判定や不適切な出力をした場合、影響を小さく抑えること（堅牢性の確保）が重要である。想定外の出力は悪意だけでなく、学習の不足、誤った利用等でも生じ、影響もセキュリティにとどまらずセキュリティ、プライバシー、人権侵害等に及ぶ。影響の低減施策では、こうしたリスクを包括的に検討する必要がある。

#### (g) AI モデルの窃取

AI モデルを不正コピーされ、同等の性能を持つコピー AI として利用されることは、セキュリティの視点から見ると窃取した AI モデルへの攻撃が容易になる、という脅威につながる。「4.2.5 (3) (c) モデルインバージョン」で記載した攻撃も AI モデル窃取を目的として用いられることがある。高価値のデータでコストをかけて学習した AI モデルの不正コピー AI が出回るというケースは今後あり得ると考えられる。保護対象として AI モデルのセキュリティを確保することが重要である<sup>\*170</sup>。

#### (h) サプライチェーン上の脅威

サービスサプライチェーン上の脆弱性を突く攻撃は AI システムに限らず、大きな脅威になり得る。AI については、学習データと AI システム自身のサプライチェーンセキュリティが求められる。学習データについては、真正性（データ改ざん・ノイズ追加等がないこと）、公平性（偏った学習でないこと）、プライバシー保護（学習に個人データが含まれる場合の匿名性確保）等が担保されていることを検証する必要がある。

AI システムについては、企業が生成 AI を利用する場合、社外にある AI モデル、及び AI モデルと連動する社内システム（営業秘密等が管理される）の情報管理とセキュリティに関する責任分担が重要となる<sup>\*161</sup>。更に、基盤モデル（AI モデル）が OSS 由来である場合、開発サプライチェーンのセキュリティ確保が非常に重要である。

### 4.2.6 AIセキュリティ対策の動向

IPA 国内調査によれば、国内企業、組織の AI セキュリティに対して重要性の認知はできているが、体制やルール策定等の本格的な取り組みはこれから、という状況である。本項では、研究機関・政府機関等から公開されている AI セキュリティガイドライン等について紹介する。

#### (1) 国内ガイドライン

AI のセキュリティガイドラインとしては、AI ベンダーを対象とする開発ガイドラインが先行している。2023 年 9 月、機械学習工学研究会は「機械学習システムセキュリティガイドライン Version 2.00」を公開した<sup>\*171</sup>。同ガイドラインは企業・民間団体が主導しており、「4.2.5 (3) AI の機能特性を突いた攻撃脅威」に記載されたような攻撃脅威への対策を事例として示し、脅威分析等になじみのない AI 開発者に具体的な対応の仕方を示している。

2023 年 11 月、内閣府は「セキュア AI システム開発ガイドライン」を公開した<sup>\*172</sup>。同ガイドラインは広島 AI プロセス<sup>\*119</sup> の補完文書として英国 NCSC、米国 CISA との共同執筆によるもので、セキュア・バイ・デザイン原則による AI システム開発、SBOM（Software Bill of Materials）の利用によるサプライチェーンセキュリティ確保等が記載されている。

更に 2024 年 4 月、総務省・経済産業省は「AI 事業者ガイドライン（第 1.0 版）」を公開した<sup>\*173</sup>。前述の AI RMF に類する国内向けの AI ガバナンス指針であり、AI 開発者・AI 提供者・AI 利用者に対してセキュリティとプライバシーを含む共通指針を示している。また広島 AI プロセスに基づき、「高度な AI システム（多目的に用いられる基盤モデル等）」の開発者に対しては共通指針に加え、「安全性のテスト」「セキュリティ確保」「モニタリング結果公開」等の要件を示している。

#### (2) 米国のガイドライン

「4.2.3 (2) ガイドライン等によるリスク対応の枠組み」に示したとおり、2023 年 1 月、NIST は Trustworthy AI の包括的なガバナンス指針となる AI RMF を公開した。更に Biden 大統領の大統領令である EO 14110 により、NIST は 2024 年 4 月に生成 AI 対応の新たなガイドラインを公開した。一つは生成 AI に関する AI RMF の拡張版 NIST AI 600-1<sup>\*135</sup>、もう一つはソフトウェアサプライチェーンセキュリティ強化を目的とした「セキュアソフトウェア開発フレームワーク (SSDF<sup>\*174</sup>)」を拡張した新たな開発ガイドライン NIST AI 800-218A<sup>\*175</sup> で、生成 AI、及びデュアルユース基盤モデル<sup>\*176</sup> のデータ保護を対象にしたプロファイルが記載されている。

これらのガイドラインは日本国内のガイドラインとの整合、セキュリティ対策策定・実践にも影響があると思われる。

## 4.2.7 まとめ

本節では、AIのセキュリティを「AIを悪用したセキュリティ脅威」「AIに対するセキュリティ脅威」に絞り込んで解説した。IPA 国内調査及び IPA 米国調査の結果から、AIに関わるセキュリティインシデントはそう顕著ではないことが確認されたが、それは脅威が小さいことを意味しない。AIの民主化により、AIリスクの統制はより難しく、応用はより広範に、技術進歩はより高速になっている。このため、新たな応用が日々生まれ、新たなリスクが日々増えている。その全体像は我々にまだ見えていないと考えるべきであり、全体を把握する作業を続けなければならない。

AIを使わないから大丈夫、ということはない。AIの

悪用による従来システムへの攻撃の増加・進化は止めることができない。情報漏えいやシステム障害等の従来型の被害に加え、フェイクコンテンツや偏った情報による詐欺、人権侵害、世論分断等はこれまでセキュリティとはやや遠い話であったが、AIリスクマネジメントではセーフティも含め、全部を地続きなものとしてとらえる必要があると思われる。

2024年4月時点で、AIセーフティ、セキュリティ、プライバシーの対策は、開発者側のガバナンス強化・規則策定等から始まっている。しかし、AIのセキュリティにおいて、誰がどう使うのか、それをどうコントロールするのかの問題は非常に大きい。使う側が何をすべきか、何ができるかの議論も進めることが重要と思われる。

※ 1 <https://rm.coe.int/information-disorder-toward-an-interdisciplinary-framework-for-research/168076277c> [2024/5/2 確認]  
※ 2 一般社団法人セーフ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](https://www.saferinternet.or.jp/wordpress/wp-content/uploads/Disinformation_report.pdf) [2024/5/2 確認]  
※ 3 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](https://www.eeas.europa.eu/eeas/1st-eeas-report-foreign-information-manipulation-and-interference-threats_en) [2024/5/2 確認]  
※ 4 Marc Laity, 2015, NATO AND THE POWER OF NARRATIVE, Peter Pomerantsev ed., Information at War: From China's Three Warfares to NATO's Narratives, London: LEGATUM INSTITUTE, 22-27, p.24  
※ 5 消費者庁：ステルスマーケティングに関する検討会報告書 [https://www.caa.go.jp/policies/policy/representation/fair\\_labeling/stealth\\_marketing](https://www.caa.go.jp/policies/policy/representation/fair_labeling/stealth_marketing) [2024/5/2 確認]  
※ 6 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> [2024/5/2 確認]  
※ 7 大澤淳「主戦場となるサイバー空間“専守防衛”では日本を守れない」月刊 Wedge、2021年12月号、pp.24-27  
※ 8 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  
※ 9 大澤淳：台湾有事とハイブリッド戦争 [https://www.spf.org/iina/articles/osawa\\_02.html](https://www.spf.org/iina/articles/osawa_02.html) [2024/5/2 確認]  
※ 10 Office of the Director of National Intelligence: Background to “Assessing Russian Activities and Intentions in Recent US Elections” : The Analytic Process and Cyber Incident Attribution [https://www.dni.gov/files/documents/ICA\\_2017\\_01.pdf](https://www.dni.gov/files/documents/ICA_2017_01.pdf) [2024/5/2 確認]  
※ 11 CYBERSCOOP : White House attributes Ukraine DDoS incidents to Russia's GRU <https://cyberscoop.com/ukraine-ddos-russia-attribution-white-house-neuberger/> [2024/5/2 確認]  
※ 12 A. B. Манойло et al, “Операции информационно-п

сихологической войны,” Горячая линия-Телеком, 2018. p108-110.

※ 13 Information at War: From China's Three Warfares to NATO's Narratives <https://li.com/wp-content/uploads/2024/05/information-at-war-from-china-s-three-warfares-to-nato-s-narratives-pdf.pdf> [2024/6/13 確認]

高木耕一郎「新領域から「バトル・オブ・ナラティブ」へ - 新領域 (宇宙、サイバー、電磁波)、心理・認知領域含む多次元環境下における将来戦」、戦略研究学会編「戦略研究 27 多次元環境下の戦略」芙蓉書房出版、2020 年、pp.49-71

※ 14 公益財団法人笹川平和財団安全保障研究グループ：“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](https://www.spf.org/global-data/user172/cyber_security_2021_web1.pdf) [2024/5/2 確認]

※ 15 A. B. Маноило et al, “Операции информационно-психологической войны,” Горячая линия-Телеком, 2018.

Daniel Bagge, “Unmasking Maskirovka: Russia's Cyber Influence Operations,” 2019.

※ 16 <https://www.airitilibrary.com/Publication/alDetailedMesh?docid=P20220613001-202112-202206130009-202206130009-19-34> [2024/5/2 確認]

※ 17 Soroush Vosoughi, Deb Roy, and Sinan Aral: The spread of true and false news online <https://www.science.org/doi/10.1126/science.aap9559> [2024/5/2 確認]

※ 18 Haruka Nakajima Suzuki, Midori Inaba: Psychological Study on Judgment and Sharing of Online Disinformation <https://ieeexplore.ieee.org/document/10196864> [2024/5/2 確認]

※ 19 Bradley D Menz, Natansh D Modi, Michael J Sorich, Ashley M Hopkins: Health Disinformation Use Case Highlighting the Urgent Need for Artificial Intelligence Vigilance: Weapons of Mass Disinformation <https://pubmed.ncbi.nlm.nih.gov/37955873/> [2024/5/2 確認]

※ 20 BBC: Gaza hospital: What video, pictures and other evidence tell us about Al-Ahli hospital blast <https://www.bbc.com/news/world-middle-east-67144061> [2024/5/2 確認]

※ 21 IDF (The Israel Defense Forces) Announcement: Briefing by IDF Spokesperson, Rear Admiral Daniel Hagari <https://idfanc.activetrail.biz/ANC1810202362> [2024/5/2 確認]

※ 22 WIRED: Who's Responsible for the Gaza Hospital Explosion? Here's Why It's Hard to Know What's Real <https://www.wired.com/story/al-ahli-baptist-hospital-explosion-disinformation-osint/> [2024/5/2 確認]

※ 23 Reuters: Biden vows aid for Gaza, Israel as protests rock Middle East <https://www.reuters.com/world/biden-heads-middle-east-inflamed-by-gaza-hospital-blast-2023-10-18/> [2024/5/2 確認]

※ 24 Institute for Strategic Dialogue: Capitalising on crisis: Russia, China and Iran use X to exploit Israel-Hamas information chaos [https://www.isdglobal.org/digital\\_dispatches/capitalising-on-crisis-russia-china-and-iran-use-x-to-exploit-israel-hamas-information-chaos/](https://www.isdglobal.org/digital_dispatches/capitalising-on-crisis-russia-china-and-iran-use-x-to-exploit-israel-hamas-information-chaos/) [2024/5/2 確認]

※ 25 JFC: イスラエル・パレスチナ情勢をめぐる大量の誤情報 / 偽情報 検証方法を解説【ファクトチェックまとめ】 <https://www.factcheckcenter.jp/fact-check/international/israel-palestine-conflict-fact-check-summary/> [2024/5/2 確認]

※ 26 日本経済新聞: イスラエルでハイブリッド戦 ハマス側サイバー攻撃周到 <https://www.nikkei.com/article/DGXZQOUC109PP0Q3A011C200000/> [2024/5/2 確認]

※ 27 Foreign Affairs: Gaza and the Future of Information Warfare <https://www.foreignaffairs.com/middle-east/gaza-and-future-information-warfare> [2024/5/2 確認]

※ 28 Reuters: Disinformation surge threatens to fuel Israel-Hamas conflict <https://jp.reuters.com/article/idUSKBN311118/> [2024/5/2 確認]

※ 29 The New York Times: In a Worldwide War of Words, Russia, China and Iran Back Hamas <https://www.nytimes.com/2023/11/03/technology/israel-hamas-information-war.html> [2024/5/2 確認]

※ 30 JFC: 「(画像) 男性が子どもたちを瓦礫から救出する画像」は AI で作成【ファクトチェック】 <https://www.factcheckcenter.jp/fact-check/international/ai-generated-image-of-man-rescuing-children-from-rubble/> [2024/5/2 確認]

Euronews: Israel-Hamas War: This viral image of a baby trapped under rubble turned out to be fake [https://www.euronews.com/my-europe/2023/10/24/israel-hamas-war-this-viral-image-of-a-](https://www.euronews.com/my-europe/2023/10/24/israel-hamas-war-this-viral-image-of-a-baby-trapped-under-rubble-turned-out-to-be-fake)

baby-trapped-under-rubble-turned-out-to-be-fake [2024/5/2 確認] Deutsche Welle: Fact check: AI fakes in Israel's war against Hamas <https://www.dw.com/en/fact-check-ai-fakes-in-israels-war-against-hamas/a-67367744> [2024/5/2 確認]

Radio Free Asia: Israel-Hamas war: How tech, social media spur misinformation <https://www.rfa.org/english/news/afcl/fact-check-israel-hamas-misinformation-11082023172217.html> [2024/5/2 確認]

JFC: 「(動画) アメリカの人気モデルがイスラエル支持を表明」は誤り AI で改変【ファクトチェック】 <https://www.factcheckcenter.jp/fact-check/international/american-model-supports-israel/#> 拡散した動画に ai による改変の形跡 [2024/5/2 確認]

※ 31 JFC: 「(画像) 男性が子どもたちを瓦礫から救出する画像」は AI で作成【ファクトチェック】 <https://www.factcheckcenter.jp/fact-check/international/ai-generated-image-of-man-rescuing-children-from-rubble/> [2024/5/2 確認]

※ 32 この画像は、在フランス中国大使館の X の投稿 (<https://twitter.com/AmbassadeChine/status/1718262759249326313?s=20&ef=factcheckcenter.jp> [2024/5/2 確認]) の画像を IPA がダウンロードしたものを掲載したものである。IPA が顔部分をぼかす加工を行った。また、JFC のファクトチェック結果 (JFC: 「(画像) 男性が子どもたちを瓦礫から救出する画像」は AI で作成【ファクトチェック】 <https://www.factcheckcenter.jp/fact-check/international/ai-generated-image-of-man-rescuing-children-from-rubble/> [2024/5/2 確認]) を基に丸の囲みや矢印を追加する加工を行った。

※ 33 JFC: 「(画像) 男性が子どもたちを瓦礫から救出する画像」は AI で作成【ファクトチェック】 <https://www.factcheckcenter.jp/fact-check/international/ai-generated-image-of-man-rescuing-children-from-rubble/> [2024/5/2 確認]

Radio Free Asia: Israel-Hamas war: How tech, social media spur misinformation <https://www.rfa.org/english/news/afcl/fact-check-israel-hamas-misinformation-11082023172217.html> [2024/5/2 確認]

※ 34 NHK: 世界を分断する SNS 発「赤ちゃん」の物語 (ナラティブ) <https://www3.nhk.or.jp/news/html/20231027/k10014237551000.html> [2024/5/2 確認]

※ 35 Deutsche Welle: Fact check: AI fakes in Israel's war against Hamas <https://www.dw.com/en/fact-check-ai-fakes-in-israels-war-against-hamas/a-67367744> [2024/5/2 確認]

※ 36 JFC: 「(動画) アメリカの人気モデルがイスラエル支持を表明」は誤り AI で改変【ファクトチェック】 <https://www.factcheckcenter.jp/fact-check/international/american-model-supports-israel/#> 拡散した動画に ai による改変の形跡 [2024/5/2 確認]

※ 37 TIME: Inside the Israel-Hamas Information War <https://time.com/6549544/israel-and-hamas-the-media-war/> [2024/5/2 確認]

※ 38 University of Maryland: American Public Attitudes on Israel/Palestine During the Israel-Gaza War [https://criticalissues.umd.edu/sites/criticalissues.umd.edu/files/UMCIP\\_October2023\\_Israel-Gaza\\_Results.pdf](https://criticalissues.umd.edu/sites/criticalissues.umd.edu/files/UMCIP_October2023_Israel-Gaza_Results.pdf) [2024/5/2 確認]

University of Maryland: American Public Attitudes on Israel/Palestine During the Israel-Gaza War: Part 2 [https://criticalissues.umd.edu/sites/criticalissues.umd.edu/files/UMCIP\\_11.3-5.2023\\_Israel-Gaza\\_Results.pdf](https://criticalissues.umd.edu/sites/criticalissues.umd.edu/files/UMCIP_11.3-5.2023_Israel-Gaza_Results.pdf) [2024/5/2 確認]

※ 39 JFC: 福島第一原発の処理水と汚染水の違いは何? 海洋放出は危険? 【ファクトチェックまとめ】 <https://www.factcheckcenter.jp/fact-check/nuclear/fukushima-daiichi-nuclear-plant-treated-water-ocean-release-fact-check-summary/> [2024/5/2 確認]

Logically Ltd.: Logically Bulletin: Coordinated Chinese campaign targets Japan's release of treated nuclear wastewater <https://www.logically.ai/resources/fukushima-daiichi-wastewater-release> [2024/5/2 確認]

※ 40 外務省: 外務省幹部とされる人物との ALPS 処理水の取扱いについての面談に関する報道について [https://www.mofa.go.jp/mofaj/press/release/press5\\_000052.html](https://www.mofa.go.jp/mofaj/press/release/press5_000052.html) [2024/5/2 確認]

外務省: 外務省のものとしてされる偽文書に関する報道について [https://www.mofa.go.jp/mofaj/press/release/press1\\_001532.html](https://www.mofa.go.jp/mofaj/press/release/press1_001532.html) [2024/5/2 確認]

※ 41 台湾事實査核中心 (台湾ファクトチェックセンター): 【易生誤解】 網傳「日本媒體報導日本排放核廢水, 原銷往中國、香港、澳門 2 萬條魚, 今晚已經改銷, 運往台灣」? <https://tfc-taiwan.org.tw/articles/9521> [2024/5/2 確認]

※ 42 経済産業省: ALPS 処理水の海洋放出に関する偽情報について <https://www.meti.go.jp/press/2023/09/20230902002/20230902002.html> [2024/5/2 確認]

※ 43 Logically Ltd.: Logically Bulletin: Coordinated Chinese campaign targets Japan's release of treated nuclear wastewater <https://www.logically.ai/resources/fukushima-daiichi-wastewater-release> [2024/5/2 確認]

※ 44 JFC: 「処理水放出で海の色が変化」は誤り【ファクトチェック】 <https://www.factcheckcenter.jp/fact-check/nuclear/discharge-treated-water-sea-color-change-false/> [2024/5/2 確認]

※ 45 この画像は、X の投稿 (<https://twitter.com/raystube/status/1694578936540451191>) を IPA がキャプチャした画面を掲載したものである。IPA がアカウント名をぼかす加工を行った。

※ 46 本段落の各事例については以下の資料 p.12 を参照。公益財団法人笹川平和財団安全保障研究グループ: 「外国からのディスインフォメーションに備えよ!~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](https://www.spf.org/global-data/user172/cyber_security_2021_web1.pdf) [2024/5/2 確認]

※ 47 インド太平洋防衛フォーラム: 2020 年の台湾選挙への介入を目的として、政治的影響力という多くの武器を駆使する中国共産党 <https://ipdefenseforum.com/ja/2019/12/2020年の台湾選挙への介入を目的として、政治的影響力> [2024/5/2 確認]

※ 48 大紀元: Venus Upadhyaya 「台湾への直接的な選挙妨害か 総統選 8 日前、前例のないサイバー攻撃」 <https://www.epochtimes.jp/2024/01/196253.html> [2024/5/2 確認]

TBS NEWS DIG: 「中国から毎日数百万回のサイバー攻撃」台湾外交部長 <https://newsdig.tbs.co.jp/articles/-/775036?display=1> [2024/5/2 確認]

※ 49 インド太平洋防衛フォーラム: 選挙が近づく台湾、中国共産党の威圧行動と情報操作に警戒 <https://ipdefenseforum.com/ja/2023/12/選挙が近づく台湾、中国共産党の威圧行動と情報> [2024/5/2 確認]

※ 50 Reuters: 台湾総統が「戦時の逃亡準備」、1 月選挙まで中国報道続々調査 <https://jp.reuters.com/world/taiwan/NSBN4DOUJHJHLHUEJQ3TCENUKE-2024-04-01/> [2024/5/17 確認]

中国での報道例としては例えば以下がある。  
网易: 美台岛撤侨计划曝光, 蔡英文随时准备逃亡! 布林肯急忙“踩刹车” <https://www.163.com/dy/article/I7RPC5F905534DFV.html> [2024/5/17 確認]

※ 51 中央通訊社: 郭無患「共軍宣布海空聯合演訓 國安人士: 介入選舉動機明確」 <https://www.cna.com.tw/news/aip/202308190067.aspx> [2024/5/2 確認]

※ 52 中央通訊社: 翟思嘉「國防部: 將整合軍煤力量反制中共認知作戰」 <https://www.cna.com.tw/news/aip/202308220124.aspx> [2024/5/2 確認]

※ 53 台湾事實查核中心: 2024 總統大選不實訊息 <https://tfc-taiwan.org.tw/topic/9640> [2024/5/2 確認]

Liberty Times: 合成總統、副總統假影片搞詐 選前首見參選人深偽片查 介 選 意 圖 <https://news.ltn.com.tw/news/society/paper/1615999> [2024/5/2 確認]

※ 54 台湾事實查核中心: 2024 選舉查證筆記第一集: 台灣首見選前 AI 造假音檔 教你判別偽造影音小撇步 <https://tfc-taiwan.org.tw/articles/9781> [2024/5/2 確認]

※ 55 U.S. Department of State: How the People's Republic of China Seeks to Reshape the Global Information Environment <https://www.state.gov/gec-special-report-how-the-peoples-republic-of-china-seeks-to-reshape-the-global-information-environment/> [2024/5/2 確認]

※ 56 Yahoo! ニュースにおいて「CGTN」で検索 (<https://news.yahoo.co.jp/search?p=CGTN&ei=utf-8>) すると、多数の日本語版の記事が表示される。

※ 57 Shen, Puma. "How China Initiates Information Operations Against Taiwan," p. 29

※ 59 フォークス台湾: 総統選 / 中国が政治的威圧と偽情報拡散で総統選に介入 台湾、情報の即時訂正体制で防衛図る <https://japan.focustaiwan.tw/politics/202401050005> [2024/5/2 確認]

※ 60 日本経済新聞: 台湾総統選後の東アジア 中国、国際的に「認知戦」展開 <https://www.nikkei.com/article/DGXZQOCD173HS0X10C24A1000000/> [2024/5/2 確認]

※ 61 JFC: (能登半島地震) 災害時に広がる偽情報 5 つの類型 地震や津波に関するデマはどう拡散するのか <https://www.factcheckcenter.jp/explainer/others/5-types-of-disinformation-about-disaster/> [2024/5/2 確認]

※ 62 NHK: 能登半島地震の偽情報 海外から多く「インプレゾンビ」が <https://www3.nhk.or.jp/news/html/20240202/k10014341931000.html> [2024/5/2 確認]

※ 63 日本経済新聞: 能登半島地震、岸田首相「虚偽情報の流布許されず」 <https://www.nikkei.com/article/DGXZQOUA020TY0S4A100C2000000/> [2024/5/2 確認]

首相官邸: 令和6年能登半島地震についての会見 [https://www.kantei.go.jp/jp/101\\_kishida/statement/2024/0102kaiken.html](https://www.kantei.go.jp/jp/101_kishida/statement/2024/0102kaiken.html) [2024/5/2 確認]

※ 64 総務省: 令和6年能登半島地震におけるインターネット上の偽・誤情報への対応 [https://www.soumu.go.jp/main\\_content/000923727.pdf](https://www.soumu.go.jp/main_content/000923727.pdf) [2024/5/2 確認]

※ 65 日本経済新聞: 災害時の偽情報対策探る 現在は要請どまり、EU は法規制 <https://www.nikkei.com/article/DGXZQOUA061020W4A100C2000000/> [2024/5/2 確認]

※ 66 デジタル空間における情報流通の健全性確保の在り方に関する検討会事務局: プラットフォーム事業者ヒアリングの総括 (暫定版※) [https://www.soumu.go.jp/main\\_content/000946374.pdf](https://www.soumu.go.jp/main_content/000946374.pdf) [2024/5/27 確認]

※ 67 THE SOUFAN CENTER: IntelBrief: AI-Powered Disinformation in the Israel-Hamas War and Beyond <https://thesoufancenter.org/intelbrief-2023-october-26/> [2024/5/2 確認]

※ 68 VOA: 'Deepfake' of Biden's Voice Called Early Example of US Election Disinformation <https://learningenglish.voanews.com/a/deepfake-of-biden-s-voice-called-early-example-of-us-election-disinformation/7455392.html> [2024/5/2 確認]

※ 69 Reuters: Deepfaking it: America's 2024 election collides with AI boom <https://jp.reuters.com/article/idUSKBN2XL0IR/> [2024/5/2 確認]

※ 70 NHK: 「選挙イヤー」の2024年 世界で高まる「フェイクへの懸念」 <https://www3.nhk.or.jp/news/html/20231216/k10014289161000.html> [2024/5/2 確認]

※ 71 WIRED: Fake Taylor Swift Quotes Are Being Used to Spread Anti-Ukraine Propaganda <https://www.wired.com/story/russia-ukraine-taylor-swift-disinformation/> [2024/5/2 確認]

※ 72 Recorded Future: Obfuscation and AI Content in the Russian Influence Network "Doppelgänger" Signals Evolving Tactics <https://go.recordedfuture.com/hubfs/reports/ta-2023-1205.pdf> [2024/5/2 確認]

※ 73 ABC News: Taylor Swift and No AI Fraud Act: How Congress plans to fight back against AI deepfakes <https://abcnews.go.com/US/taylor-swift-ai-fraud-act-congress-plans-fight/story?id=106765709> [2024/5/2 確認]

※ 74 The Wall Street Journal: Lab Leak Most Likely Origin of Covid-19 Pandemic, Energy Department Now Says <https://www.wsj.com/articles/covid-origin-china-lab-leak-807b7b0a> [2024/5/2 確認]

※ 75 独立行政法人国民生活センター: これまでに寄せられた新型コロナウイルス関連の消費者トラブル [https://www.kokusen.go.jp/soudan\\_now/data/coronavirus\\_jirei.html](https://www.kokusen.go.jp/soudan_now/data/coronavirus_jirei.html) [2024/5/2 確認]

※ 76 JFC: ワクチン <https://www.factcheckcenter.jp/tag/vaccine/> [2024/5/2 確認]

※ 77 Springer Nature: Fujio Toriumi, Takeshi Sakaki, Tetsuro Kobayashi & Mitsuo Yoshida: Anti-vaccine rabbit hole leads to political representation: the case of Twitter in Japan <https://link.springer.com/article/10.1007/s42001-023-00241-8> [2024/5/2 確認]

※ 78 笹原和俊、デジタル影響工作に対する計算社会科学のアプローチ、一田和樹他、ネット世論操作とデジタル影響工作、原書房、2023年3月

※ 79 WHO: Infodemic [https://www.who.int/health-topics/infodemic#tab=tab\\_1](https://www.who.int/health-topics/infodemic#tab=tab_1) [2024/5/2 確認]

※ 80 POLITICO: State report: Russian, Chinese and Iranian disinformation narratives echo one another <https://www.politico.com/news/2020/04/21/russia-china-iran-disinformation-coronavirus-state-department-193107> [2024/5/2 確認]

※ 81 佐々木孝博、ロシアによるデジタル影響工作、一田和樹他、ネット世論操作とデジタル影響工作、原書房、2023年3月

※ 82 BBC NEWS JAPAN: コンサートホール襲撃 ロシアはなぜウクライナのせいにするのか <https://www.bbc.com/japanese/articles/c51npj7v1lxo> [2024/5/2 確認]

※ 83 U.S. Department of State: GEC Special Report: August 2020 Pillars of Russia's Disinformation and Propaganda Ecosystem [https://www.state.gov/wp-content/uploads/2020/08/Pillars-of-Russia%E2%80%99s-Disinformation-and-Propaganda-Ecosystem\\_08-04-20.pdf](https://www.state.gov/wp-content/uploads/2020/08/Pillars-of-Russia%E2%80%99s-Disinformation-and-Propaganda-Ecosystem_08-04-20.pdf) [2024/5/2 確認]

※ 84 Microsoft Corporation: ウクライナの防衛: サイバー戦争の初期の教訓 <https://news.microsoft.com/ja-jp/2022/07/04/220704-defending-ukraine-early-lessons-from-the-cyber-war/> [2024/5/2

確認]

※ 85 JFC : ウクライナ <https://www.factcheckcenter.jp/tag/ukraine/> [2024/5/2 確認]

※ 86 朝日新聞 : ロシアの偽情報作戦、ソ連時代から「お家芸」ウクライナ危機の深層 <https://digital.asahi.com/articles/ASQ2S7H84Q2SUHBI03X.html> [2024/5/2 確認]

※ 87 The New York Times : Russia has been laying groundwork online for a 'false flag' operation, misinformation researchers say. <https://www.nytimes.com/2022/02/19/business/russia-has-been-laying-groundwork-online-for-a-false-flag-operation-misinformation-researchers-say.html> [2024/5/2 確認]

※ 88 藤村厚夫、世界のメディアの変容、一田和樹他、ネット世論操作とデジタル影響工作、原書房、2023年3月

※ 89 The Wall Street Journal : ロシアで SNS 「テレグラム」急成長の理由 <https://jp.wsj.com/articles/telegram-thrives-amid-russias-media-crackdown-11647826301> [2024/5/2 確認]

※ 90 Internews : Ukrainians increasingly rely on Telegram channels for news and information during wartime <https://internews.in.ua/news/ukrainians-increasingly-rely-on-telegram-channels-for-news-and-information-during-wartime/> [2024/5/2 確認]

※ 91 EL PAIS : Ukraine considers banning Telegram if app is confirmed as threat to national security <https://english.elpais.com/international/2024-04-02/ukraine-considers-banning-telegram-if-app-is-confirmed-as-threat-to-national-security.html> [2024/5/2 確認]

※ 92 University of Cambridge : The failure of Russian propaganda <https://www.cam.ac.uk/stories/donbaspropaganda> [2024/5/2 確認]

※ 93 内閣官房 : 国家安全保障戦略について <https://www.cas.go.jp/jp/siryou/221216anzenhoshou.html> [2024/5/2 確認]

※ 94 産経新聞 : <独自>陸自に「認知戦」対処専門部隊新設 安保3文書に明記 <https://www.sankei.com/article/20221208-MLNG77HAEZOQPIYMA2IPB7SMU/> [2024/5/2 確認]

※ 95 読売新聞オンライン : 海自に電子戦や偽情報対策担う部隊新設、25年までに2000人規模…3文書案 <https://www.yomiuri.co.jp/politics/20221209-0YT1T50304/> [2024/5/2 確認]

※ 96 防衛省、自衛隊 : 認知領域を含む情報戦への対応 <https://www.mod.go.jp/j/approach/defense/infowarfare/index.html> [2024/5/27 確認]

国家安全保障会議決定、閣議決定 : 防衛力整備計画 <https://www.mod.go.jp/j/policy/agenda/guideline/plan/pdf/plan.pdf> [2024/5/27 確認]

※ 97 首相官邸 : 令和5年4月14日(金)午前・内閣官房長官記者会見 [https://www.kantei.go.jp/jp/tyoukanpress/202304/14\\_a.html](https://www.kantei.go.jp/jp/tyoukanpress/202304/14_a.html) [2024/5/2 確認]

※ 98 日本経済新聞 : 防衛省、偽情報対策で新ポスト 諸外国の意図や影響分析 <https://www.nikkei.com/article/DGXZQOUA013J90R00C22A4000000/> [2024/5/2 確認]

※ 99 外務省 : 令和6年度概算要求の概要 <https://www.mofa.go.jp/mofaj/files/100546567.pdf> [2024/5/2 確認]

※ 100 総務省 : 「デジタル空間における情報流通の健全性確保の在り方に関する検討会」の開催 [https://www.soumu.go.jp/menu\\_news/s-news/01ryutsu02\\_02000374.html](https://www.soumu.go.jp/menu_news/s-news/01ryutsu02_02000374.html) [2024/5/2 確認]

※ 101 G7 2023 HIROSHIMA SUMMIT : Ministerial Declaration The G7 Digital and Tech Ministers' Meeting 30 April 2023 [https://www.soumu.go.jp/joho\\_kokusai/g7digital-tech-2023/topics/pdf/pdf\\_20230430/ministerial\\_declaration\\_dtmm.pdf](https://www.soumu.go.jp/joho_kokusai/g7digital-tech-2023/topics/pdf/pdf_20230430/ministerial_declaration_dtmm.pdf) [2024/5/2 確認]

※ 102 G7 2023 HIROSHIMA SUMMIT : Existing Practices against Disinformation (EPaD) [https://www.soumu.go.jp/main\\_content/000905620.pdf](https://www.soumu.go.jp/main_content/000905620.pdf) [2024/5/2 確認]

※ 103 NHK : SNS などの偽情報対策 日米で連携して対処へ 協力文書に署名 <https://www.3nhk.or.jp/news/html/20231206/k10014279951000.html> [2024/5/2 確認]

※ 104 読売新聞オンライン : ネット上の誹謗中傷は迅速削除、SNS 大手に義務付けへ…法改正で削除基準の透明化も <https://www.yomiuri.co.jp/national/20240111-0YT1T50187/> [2024/5/2 確認]

※ 105 読売新聞オンライン : 能登半島地震巡る偽情報対策、被災自治体とOP技術を使い実証実験へ…岸田首相「虚偽情報の流布許さない」 <https://www.yomiuri.co.jp/politics/20240124-0YT1T50000/> [2024/5/2 確認]

※ 106 JFC : 日本ファクトチェックセンターが AI 活用 LINE でユーザーからの質問に答えます <https://www.factcheckcenter.jp/info/others/ai-answers-user-questions-on-line/> [2024/5/2 確認]

※ 107 The White House : Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence <https://www.whitehouse.gov/briefing-room/presidential-actions/2023/10/30/executive-order-on-the-safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence/> [2024/5/2 確認]

※ 108 GOV.UK : The Bletchley Declaration by Countries Attending the AI Safety Summit, 1-2 November 2023 <https://www.gov.uk/government/publications/ai-safety-summit-2023-the-bletchley-declaration/the-bletchley-declaration-by-countries-attending-the-ai-safety-summit-1-2-november-2023> [2024/5/2 確認]

※ 109 NIST : U.S. Commerce Secretary Gina Raimondo Announces Key Executive Leadership at U.S. AI Safety Institute <https://www.nist.gov/news-events/news/2024/02/us-commerce-secretary-gina-raimondo-announces-key-executive-leadership-us> [2024/5/2 確認]

GOV.UK : AI Safety Institute <https://www.gov.uk/government/organisations/ai-safety-institute> [2024/5/2 確認]

※ 110 <https://aisi.go.jp/> [2024/5/2 確認]

※ 111 読売新聞オンライン : 選挙での「ディープフェイク」に歯止め、IT20社が生成AI偽情報対策で合意…OP技術研究組合も支持 <https://www.yomiuri.co.jp/economy/20240217-0YT1T50140/> [2024/5/2 確認]

※ 112 読売新聞社大阪本社社会部、情報バンデミック あなたを惑わすものの正体 第2章 発信者を追う なぜ広めるのか、中央公論社、2022年11月

※ 113 RISTEX : SDGsの達成に向けた共創的研究開発プログラム(情報社会における社会的側面からのトラスト形成)について <https://www.jst.go.jp/ristex/funding/solve-digist/> [2024/5/2 確認]

※ 114 公益財団法人笹川平和財団安全保障研究グループ : 「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](https://www.spf.org/global-data/user172/cyber_security_2021_web1.pdf) [2024/5/2 確認]

※ 115 JFC : JFC ファクトチェック講座 3 : 検証の4ステップ「横読み」で効率的に <https://www.factcheckcenter.jp/course/fact-check-course/4-step-verification-efficient-skimming/> [2024/5/2 確認]

※ 116 RAND : The Rise of Generative AI and the Coming Era of Social Media Manipulation 3.0 <https://www.rand.org/pubs/perspectives/PEA2679-1.html> [2024/5/2 確認]

※ 117 総務省 : 平成30年版 情報通信白書 第2章 ICTによる新たなエコノミーの形成 <https://www.soumu.go.jp/johotsusintokei/whitepaper/ja/h30/pdf/n2100000.pdf> [2024/6/19 確認]

IPA : AI 白書 2019 <https://www.ipa.go.jp/publish/wp-ai/qv6pgp000000w5z-att/000088602.pdf> [2024/6/13 確認]

※ 118 European Commission : AI Act <https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai> [2024/5/27 確認]

※ 119 総務省 : 広島 AI プロセスについて [https://www8.cao.go.jp/cstp/ai/ai\\_senryaku/7kai/11hiroshimaaipurosesu.pdf](https://www8.cao.go.jp/cstp/ai/ai_senryaku/7kai/11hiroshimaaipurosesu.pdf) [2024/5/27 確認]

※ 120 <https://www.ipa.go.jp/archive/publish/wp-security/qv6pgp000000v5l-att/000079041.pdf> [2024/5/27 確認]

※ 121 大規模言語モデル : 大量のデータセットとディープラーニング技術により、文章や単語の出現確率を推定するモデル。自然言語処理の精度を大幅に向上させた。

※ 122 ZDNET : What is ChatGPT and why does it matter? Here's what you need to know <https://www.zdnet.com/article/what-is-chatgpt-and-why-does-it-matter-heres-everything-you-need-to-know/> [2024/5/27 確認]

※ 123 基盤モデル : 大量のデータ(画像・動画・音声等を含む)を学習し、様々な用途(アプリケーション)にチューニングできるモデル。大規模言語モデルもその一つ。

※ 124 GAN (Generative Adversarial Network) : GANは敵対的生成ネットワークと呼ばれ、ラベルのない入力データ(教師なし学習)から実際には存在しない人の顔画像等を高品質で生成できる。

※ 125 Meta社 : Build the future of AI with Meta Llama 3 <https://llama.meta.com/llama3/> [2024/5/27 確認]

※ 126 総務省 : 令和元年版 情報通信白書 第3節 (2) 進むAIの民主化 <https://www.soumu.go.jp/johotsusintokei/whitepaper/ja/r01/html/nd113220.html> [2024/5/27 確認]

※ 127 深層学習(ディープラーニング) : パラメータ化されたモジュールを多層的に組み合わせたニューラルネットワークモデルを扱う学習手法。

※ 128 総務省 : 令和5年版 情報通信白書 第3節 インターネット上の偽・誤情報の拡散等 <https://www.soumu.go.jp/johotsusintokei/whitepaper/ja/r05/pdf/n2300000.pdf> [2024/5/27 確認]

※ 129 IPA の「AI 白書 2019」(<https://www.ipa.go.jp/publish/wp->



ai/qv6pgp000000w5z-att/000088602.pdf[2024/5/27 確認])の「第5章 AIの社会実装課題と対策」参照。

- ※ 130 いわゆる汎用人工知能が人間の知性を凌駕するシンギュラリティがリスクとして議論されることがあるが、本稿ではこの課題は扱わない。
- ※ 131 内閣府：AIガバナンスに関する議論の方向性について（ディスカッションペーパー） [https://www8.cao.go.jp/cstp/ai/ningen/r5\\_1kai/siry03.pdf](https://www8.cao.go.jp/cstp/ai/ningen/r5_1kai/siry03.pdf)[2024/5/27 確認]
- ※ 132 ISO：ISO/IEC 42001:2023 Information technology <https://www.iso.org/standard/81230.html>[2024/5/27 確認]
- ※ 133 経済産業省：AIマネジメントシステムの国際規格が発行されました <https://www.meti.go.jp/press/2023/01/20240115001/20240115001.html>[2024/5/27 確認]
- ※ 134 NIST:Artificial Intelligence Risk Management Framework (AI RMF 1.0) <https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf>[2024/5/27 確認]
- ※ 135 NIST: Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile <https://airc.nist.gov/docs/NIST.AI.600-1.GenAI-Profile.ipd.pdf>[2024/5/27 確認]
- ※ 136 OECD：OECD AI Principles overview <https://oecd.ai/en/ai-principles>[2024/5/27 確認]
- ※ 137 本白書では文献引用上の正確性を期す必要がない場合、表記の統一のため、悪意のあるプログラム、マルウェア等を総称して「ウイルス」と表記する。
- ※ 138 米国のAISI (U.S. AISI) は2024年2月7日にNISTに設置された。  
NIST: U.S. ARTIFICIAL INTELLIGENCE SAFETY INSTITUTE <https://www.nist.gov/aisi>[2024/5/27 確認]  
日本のAISI (AISI Japan) は2024年2月14日にIPAに設置された。  
AISI:Japan AI Safety Institute <https://aisi.go.jp>[2024/5/27 確認]
- ※ 139 AISI:AI事業者ガイドラインと米国NIST AIリスクマネジメントフレームワーク (RMF) とのクロスウォーク [https://aisi.go.jp/2024/04/30/ai\\_rmf\\_crosswalk1\\_news/](https://aisi.go.jp/2024/04/30/ai_rmf_crosswalk1_news/)[2024/5/27 確認]
- ※ 140 IPA：IPA テクニカルウォッチ「AI利用時のセキュリティ脅威・リスク調査報告書」 <https://www.ipa.go.jp/security/reports/technicalwatch/20240704.html>[2024/7/5 確認]
- ※ 141 IPA：AI RISK AND THREATS [https://www.ipa.go.jp/security/reports/technicalwatch/m42obm000000hzkm-att/2024\\_IPA\\_Report1\\_FINAL\\_forPublic.pdf](https://www.ipa.go.jp/security/reports/technicalwatch/m42obm000000hzkm-att/2024_IPA_Report1_FINAL_forPublic.pdf)[2024/6/18 確認]
- ※ 142 NCSC：The near-term impact of AI on the cyber threat <https://www.ncsc.gov.uk/report/impact-of-ai-on-cyber-threat>[2024/5/27 確認]
- ※ 143 Cyber security news: Hackers Released New Black Hat AI Tools XXXGPT and Wolf GPT <https://cybersecuritynews.com/black-hat-ai-tools-xxxgpt-and-wolf-gpt/>[2024/5/27 確認]
- ※ 144 SlashNext: The State of Phishing 2023 <https://slashnext.com/wp-content/uploads/2023/10/SlashNext-The-State-of-Phishing-Report-2023.pdf>[2024/5/27 確認]
- ※ 145 IPA 米国調査「AI RISK AND THREATS」([https://www.ipa.go.jp/security/reports/technicalwatch/m42obm000000hzkm-att/2024\\_IPA\\_Report1\\_FINAL\\_forPublic.pdf](https://www.ipa.go.jp/security/reports/technicalwatch/m42obm000000hzkm-att/2024_IPA_Report1_FINAL_forPublic.pdf)[2024/6/18 確認])の「Table 7: Summary of AI Threats and Risk Chart」参照。
- ※ 146 The New York Times: 'A.I. Obama' and Fake Newscasters: How A.I. Audio Is Swarming TikTok <https://www.nytimes.com/2023/10/12/technology/tiktok-ai-generated-voices-disinformation.html>[2024/5/27 確認]
- ※ 147 例えばロシア・ウクライナ戦争については、以下を参照。  
The Japan News: Examining Generative AI / Russian Side Seeks to Undermine Ukraine Via Disinformation; Fake Video Shows Military Leader Criticizing Zelenskyy <https://japannews.yomiuri.co.jp/society/social-series/20240221-169948/>[2024/5/27 確認]  
イスラエル・ハマス間の武力衝突については以下を参照。  
WIRED: Generative AI Is Playing a Surprising Role in Israel-Hamas Disinformation <https://www.wired.com/story/israel-hamas-war-generative-artificial-intelligence-disinformation/>[2024/5/27 確認]
- ※ 148 MITHRIL SECURITY: PoisonGPT: How We Hid a Lobotomized LLM on Hugging Face to Spread Fake News <https://blog.mithrilsecurity.io/poisongpt-how-we-hid-a-lobotomized-llm-on-hugging-face-to-spread-fake-news/>[2024/5/27 確認]
- ※ 149 <https://huggingface.co>[2024/5/27 確認]
- ※ 150 Reuters: 焦点：ソーシャルメディア大手、「選挙イヤー」のフェイク対策が後手に <https://jp.reuters.com/business/technology/6UH5ANX25JLXPN5226HWQES2YM-2024-01-12/>[2024/5/27 確認]
- ※ 151 CISA: Risk in Focus: Generative A.I. and the 2024 Election

Cycle <https://www.cisa.gov/resources-tools/resources/risk-focus-generative-ai-and-2024-election-cycle>[2024/5/27 確認]

- ※ 152 人工知能学会、越前功、馬場口登、笹原和俊、インフォデミック時代におけるフェイクメディア克服の最前線 人工知能学会誌 Vol.38 No.2 (2023/3), pp.189-196
- ※ 153 統計的な誤判定による誤動作の対応を「統計的に起こる故障対応の一貫」と考えるセーフティ専門家もいる。
- ※ 154 朝日新聞：完全無人タクシーが女性をひいて体の上に停車 ミサンフランシスコ <https://digital.asahi.com/articles/ASRB43JCWRB4UHB100S.html>[2024/5/27 確認]
- ※ 155 自動運転 レベル4：自動車専用道路や敷地内・送迎ルート等の限定エリアで、人間が介在しない完全な自動運転が行われるレベル。
- ※ 156 NHK：「レベル4」自動運転事故 カメラが自転車を認識できず 福井 <https://www3.nhk.or.jp/news/html/20231110/k10014254121000.html>[2024/5/27 確認]
- ※ 157 日本経済新聞:自動運転レベル4 運行再開 昨年に事故、福井・永平寺 <https://www.nikkei.com/article/DGXZQOUE1610E0W4A310C2000000/>[2024/5/27 確認]
- ※ 158 TechTimes: Samsung Employees Use ChatGPT at Work, Unknowingly Leak Critical Source Codes <https://www.techtimes.com/articles/289996/20230404/samsung-employees-used-chatgpt-work-unknowingly-leaked-critical-source-codes.htm>[2024/5/27 確認]
- ※ 159 CIO: CIOs are worried about the informal rise of generative AI in the enterprise <https://www.cio.com/article/650764/cios-are-worried-about-the-informal-rise-of-generative-ai-in-the-enterprise.html>[2024/5/27 確認]
- ※ 160 個人情報保護委員会:生成 AI サービスの利用に関する注意喚起等について [https://www.ppc.go.jp/files/pdf/230602\\_kouhou\\_houdou.pdf](https://www.ppc.go.jp/files/pdf/230602_kouhou_houdou.pdf)[2024/5/27 確認]
- ※ 161 IPA による国内有識者インタビューを基にした。
- ※ 162 ACM Digital Library: Do Users Write More Insecure Code with AI Assistants? <https://dl.acm.org/doi/abs/10.1145/3576915.3623157>[2024/5/27 確認]
- ※ 163 人口知能学会、大塚 玲、AIセキュリティの研究動向、人工知能学会誌 Vol.38 No.2 (2023/3), pp.181-188
- ※ 164 総務省、三井物産セキュアディレクション株式会社：総務省×MBSD：詳細解説 [https://www.mbsd.jp/aisec\\_portal/detail\\_category.html](https://www.mbsd.jp/aisec_portal/detail_category.html)[2024/5/27 確認]
- ※ 165 Ian J. Goodfellow, Jonathon Shlens and Christian Szegedy: EXPLAINING AND HARNESSING ADVERSARIAL EXAMPLES <https://arxiv.org/pdf/1412.6572.pdf>[2024/5/27 確認]
- ※ 166 David Silver: Adversarial Traffic Signs <https://medium.com/self-driving-cars/adversarial-traffic-signs-fd16b7171906>[2024/6/24 確認]
- ※ 167 Matt Fredrikson, Somesh Jha and Thomas Ristenpart: Model Inversion Attacks that Exploit Confidence Information and Basic Countermeasures <https://www.cs.cmu.edu/~mfredrik/papers/fjr2015ccs.pdf>[2024/5/27 確認]
- ※ 168 JFrog Ltd.: Data Scientists Targeted by Malicious Hugging Face ML Models with Silent Backdoor <https://jfrog.com/blog/data-scientists-targeted-by-malicious-hugging-face-ml-models-with-silent-backdoor/>[2024/5/27 確認]
- ※ 169 OWASP Japan: ML02:2023 データポイズニング攻撃 (Data Poisoning Attack) [https://coky-t.github.io/owasp-machine-learning-security-top-10-ja/ml02\\_2023-data\\_poisoning\\_attack](https://coky-t.github.io/owasp-machine-learning-security-top-10-ja/ml02_2023-data_poisoning_attack)[2024/5/27 確認]
- ※ 170 学習モデルの知的財産としての扱い、それに対するオーナーシップは誰が主張できるか、等も別途検討すべき重要課題である。
- ※ 171 日本ソフトウェア科学会 機械学習工学研究会: : 機械学習システムセキュリティガイドライン <https://github.com/mlse-jssst/security-guideline>[2024/5/27 確認]
- ※ 172 内閣府:セキュア AI システム開発ガイドラインについて <https://www8.cao.go.jp/cstp/stmain/20231128ai.html>[2024/5/27 確認]
- ※ 173 [https://www.meti.go.jp/shingikai/mono\\_info\\_service/ai\\_shakai\\_jisso/pdf/20240419\\_1.pdf](https://www.meti.go.jp/shingikai/mono_info_service/ai_shakai_jisso/pdf/20240419_1.pdf)[2024/5/27 確認]
- ※ 174 NIST: Secure Software Development Framework <https://csrc.nist.gov/projects/ssdf>[2024/5/27 確認]
- ※ 175 NIST: Secure Software Development Practices for Generative AI and Dual-Use Foundation Models <https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-218A.ipd.pdf>[2024/6/20 確認]
- ※ 176 デュアルユース基盤モデル: 悪用されると、国家安全保障、経済安全保障等に深刻な問題をもたらすと考えられる基盤モデル。

# 付録

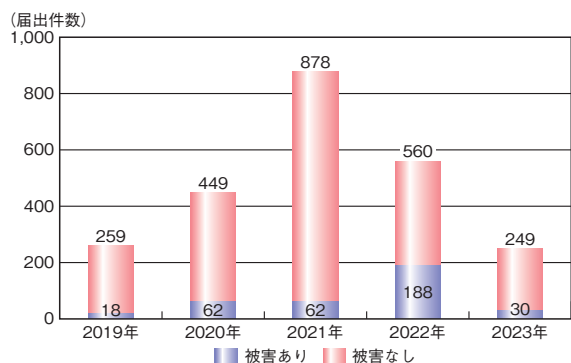
## 資料

## 資料A 2023年のコンピュータウイルス届出状況

IPA が 2023 年 1 月から 12 月の期間に受け付けたコンピュータウイルス（以下、ウイルス）届出の集計結果について述べる。

### A.1 届出件数

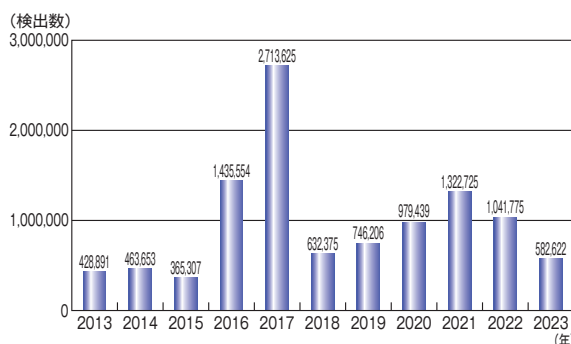
2023 年の年間届出件数は、前年の 560 件より 311 件（55.5%）少ない 249 件であった（図 A-1）。そのうち、ウイルス感染の実被害があった届出は 30 件であった。



■図 A-1 ウイルス届出件数推移（2019～2023 年）

### A.2 届出のあったウイルス等検出数

2023 年に寄せられたウイルス等の検出数は、前年の 104 万 1,775 個より 45 万 9,153 個（44.1%）少ない 58 万 2,622 個であった（図 A-2）。



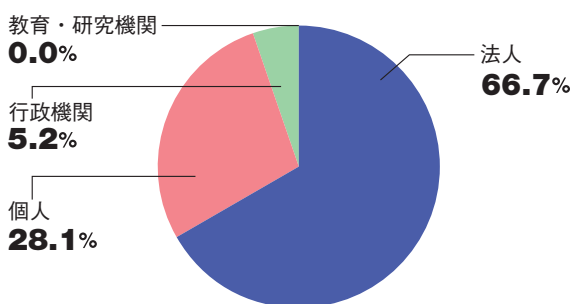
■図 A-2 ウイルス等検出数推移（2013～2023 年）

### A.3 届出者の主体別届出件数

2023 年の主体別届出件数は前年と比較すると、全体的に減少した。主体別の比率では「法人」からの届出が 66.7%（166 件）と最も多かった（表 A-1、図 A-3）。

届出者の主体	2021 年	2022 年	2023 年
法人	284	388	166
個人	578	145	70
行政機関	15	18	13
教育・研究機関	1	9	0
合計（件）	878	560	249

■表 A-1 ウイルス届出者の主体別届出件数（2021～2023 年）



■図 A-3 ウイルス届出者の主体別届出件数の比率（2023 年）

### A.4 傾向

2023 年でウイルス感染の実被害に遭った届出 30 件のうち、ランサムウェアの感染被害が 11 件あった。また、Emotet の感染被害も同じく 11 件あり、2022 年で実被害に遭った届出 188 件のうち、Emotet の感染被害が 145 件であったことに比べると大幅に減少したものの届出はされている。なお、Emotet に関しては不定期に休止・再開を繰り返しており、今後、再び大規模な攻撃活動が開始される可能性もあるため、引き続き警戒をしていただきたい。

これらの届出件数の詳細は、下記の資料から参照可能であり、ランサムウェアの攻撃手口や対策に関しては、本白書の「1.2.1 ランサムウェア攻撃」にて詳しく述べているので、ぜひそちらを一読いただきたい。

#### 参照

■コンピュータウイルス・不正アクセスの届出状況[2023年(1月～12月)]

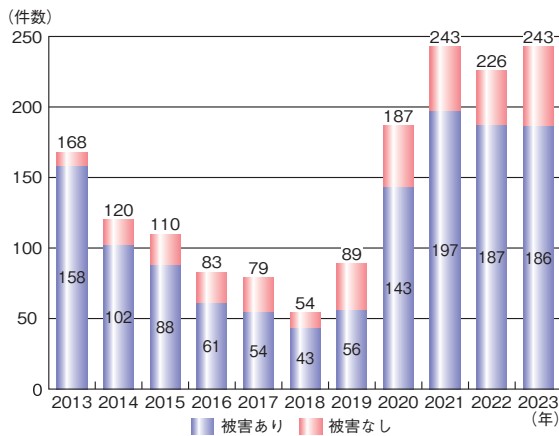
<https://www.ipa.go.jp/security/todokede/crack-virus/ug65p9000000nnpa-att/2023-report.pdf>

## 資料B 2023年のコンピュータ不正アクセス届出状況

IPA が2023年1月から12月の期間に受け付けたコンピュータ不正アクセス（以下、不正アクセス）届出の集計結果について述べる。

### B.1 届出件数

2023年の年間届出件数は、前年の226件より17件(7.5%)多い243件であった(図B-1)。そのうち、実被害があった届出は186件であった。



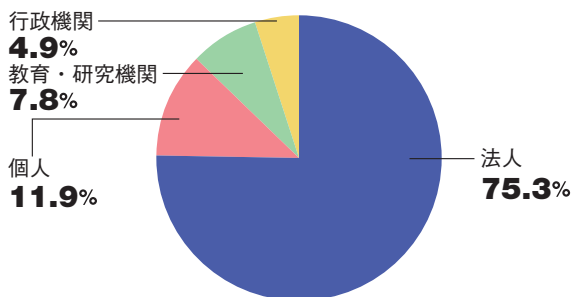
■ 図 B-1 不正アクセス届出件数推移 (2013年～2023年)

### B.2 届出者の主体別届出件数

2023年は前年と比較すると、「法人」からの届出件数が増加した一方で、その他の届出件数は減少している。届出者の主体別の比率で見ると「法人」からの届出が75.3%(183件)と最も多かった(表B-1、図B-2)。

届出者の主体	2021年	2022年	2023年
法人	156	137	183
個人	46	50	29
教育・研究機関	22	21	19
行政機関	19	18	12
合計(件)	243	226	243

■ 表 B-1 不正アクセス届出者の主体別届出件数 (2021～2023年)

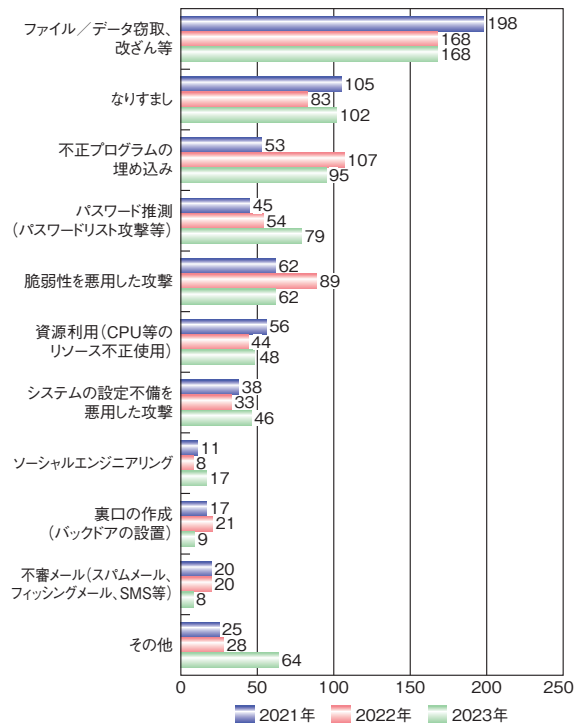


■ 図 B-2 不正アクセス届出者の主体別届出件数の比率 (2023年)

### B.3 手口別件数

届出を攻撃行為(手口)により分類した件数を図B-3に示す。なお、以降の分類も含め、届出1件につき、複数の分類項目が該当する場合がある。その場合は該当する項目のそれぞれにカウントした。

2023年の届出において最も多く見られた手口は、前年と同様に「ファイル/データ窃取、改ざん等」の168件であり、次いで「なりすまし」が102件、「不正プログラムの埋め込み」が95件であった。



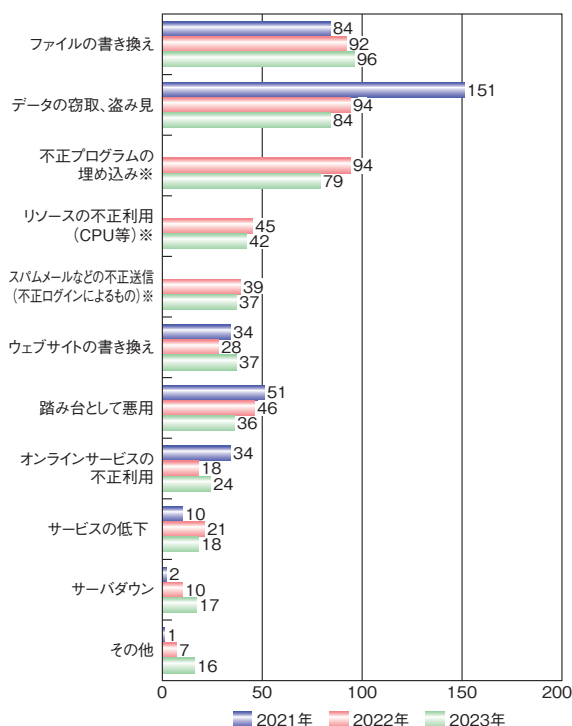
■ 図 B-3 不正アクセス手口別件数の推移 (2021～2023年)

### B.4 被害内容別件数

届出のうち、実際に被害に遭った届出について、被害内容により分類した件数を図B-4に示す。2023年の届出において最も多く見られた被害は、「ファイルの書き換え」の96件であった。次いで「データの窃取、盗み見」が84件、「不正プログラムの埋め込み」が79件であった。

なお、具体的な被害事例については、「コンピュータウイルス・不正アクセスに関する届出について」(<https://www.ipa.go.jp/security/todokede/crack-virus/about.html>)において「コンピュータウイルス・不正アクセスの届出事例[2023年上半期(1月～6月)]」及び「コン

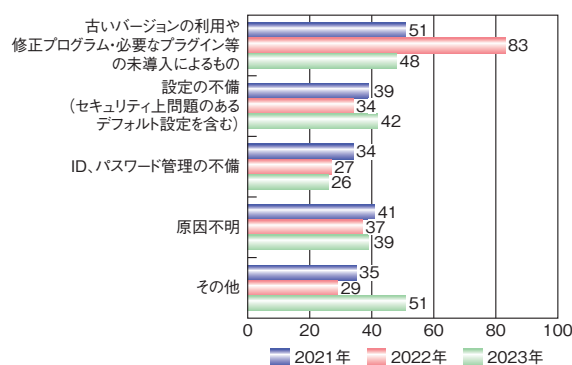
ピュータウイルス・不正アクセスの届出事例 [2023 年下半期 (7 月～12 月)]」を紹介している。こちらも、ぜひ参考にしていただきたい。



■図 B-4 不正アクセス被害内容別件数の推移 (2021～2023 年)  
※被害内容が多様化したため、2022 年から項目を細分化した。

## B.5 原因別件数

実際に被害に遭った届出について、不正アクセスの原因となった問題点／弱点で分類した件数を図 B-5 に示す。2023 年の届出において最も多く見られた原因は、前年と同様に「古いバージョンの利用や修正プログラム・必要なプラグイン等の未導入によるもの」であり 48 件であった。次いで「設定の不備(セキュリティ上問題のあるデフォルト設定を含む)」が 42 件、「ID、パスワード管理の不備」が 26 件であった。



■図 B-5 不正アクセス原因別件数の推移 (2021～2023 年)

## B.6 傾向と対策

不正アクセスの傾向と対策について述べる。

### (1) 傾向

図 B-1 に示した 2023 年に届出された 243 件について、不正アクセス (被害なしも含む) の傾向を分析したところ、「Web サイトの脆弱性や設定不備の悪用に関する不正アクセス」が 65 件、「VPN 装置の脆弱性やリモートデスクトップサービスの設定不備を悪用したランサムウェア攻撃に関する不正アクセス」が 52 件確認された。また、「パスワードリスト攻撃や総当たり攻撃で、認証を突破されたことによる、メールアカウント等の不正アクセス」が 44 件あった。

### (2) 対策

(1) で示した脆弱性や設定不備の対策としては、利用している機器やソフトウェアに関する脆弱性情報の収集や修正プログラムの適用、設定の定期的な見直しといった、基本的なセキュリティ対策を実施することが重要である。企業・組織においては、脆弱性診断やペネトレーションテスト等を行い、確実に脆弱性や設定不備を解消することが望まれる。なお、ソフトウェア等の脆弱性対策に関しては、本白書の「1.2.5 ソフトウェアの脆弱性を悪用した攻撃」も参照していただきたい。

メールアカウント等の不正アクセスに関する対策としては、企業・組織やシステム利用者に限らず、他者に推測されにくい複雑なパスワードを設定する、パスワードの使い回しをしない等の基本的な対策を実施することに加え、利用しているシステムで多要素認証等のセキュリティオプションが用意されている場合には積極的に採用する等、今一度、アカウントが適切に管理できているか見直すことを勧める。

## 参照

■コンピュータウイルス・不正アクセスの届出状況 [2023 年 (1 月～12 月)]

<https://www.ipa.go.jp/security/todokede/crack-virus/ug65p9000000nnpa-att/2023-report.pdf>

## 資料C ソフトウェア等の脆弱性関連情報に関する届出状況

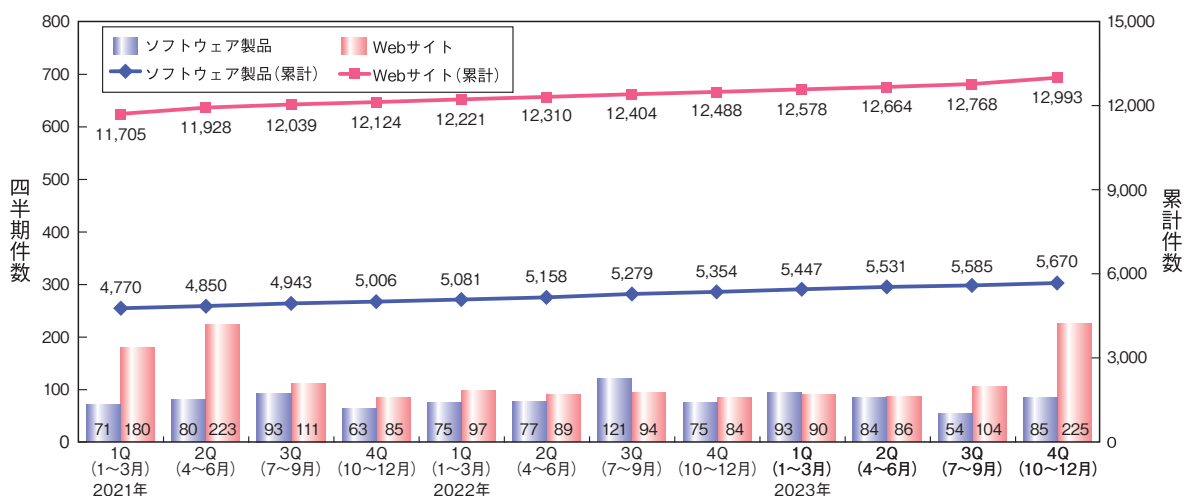
IPA が受け付けたソフトウェア製品や Web サイトの脆弱性の情報について、届出件数や処理の状況を述べる。

Web サイトに関するもの 1 万 2,993 件、合計 1 万 8,663 件で、Web サイトに関する届出が全体の 69.6% を占めている(図 C-1)。

### C.1 脆弱性の届出概況

2023 年末時点で、届出受付開始(2004 年 7 月 8 日)からの累計は、ソフトウェア製品に関するもの 5,670 件、

表 C-1 に示すように、届出受付開始から各四半期末時点までの就業日 1 日あたりの届出件数は、2023 年第 4 四半期末時点で 3.93 件となっている。



■ 図 C-1 脆弱性関連情報の届出件数の四半期別推移

2021年1Q (1~3月)	2021年2Q (4~6月)	2021年3Q (7~9月)	2021年4Q (10~12月)	2022年1Q (1~3月)	2022年2Q (4~6月)	2022年3Q (7~9月)	2022年4Q (10~12月)	2023年1Q (1~3月)	2023年2Q (4~6月)	2023年3Q (7~9月)	2023年4Q (10~12月)
4.04	4.06	4.05	4.02	4.01	3.99	3.98	3.97	3.95	3.94	3.92	3.93

■ 表 C-1 就業日 1 日あたりの届出件数 (届出受付開始から各四半期末時点)

### C.2 ソフトウェア製品の脆弱性届出の処理状況

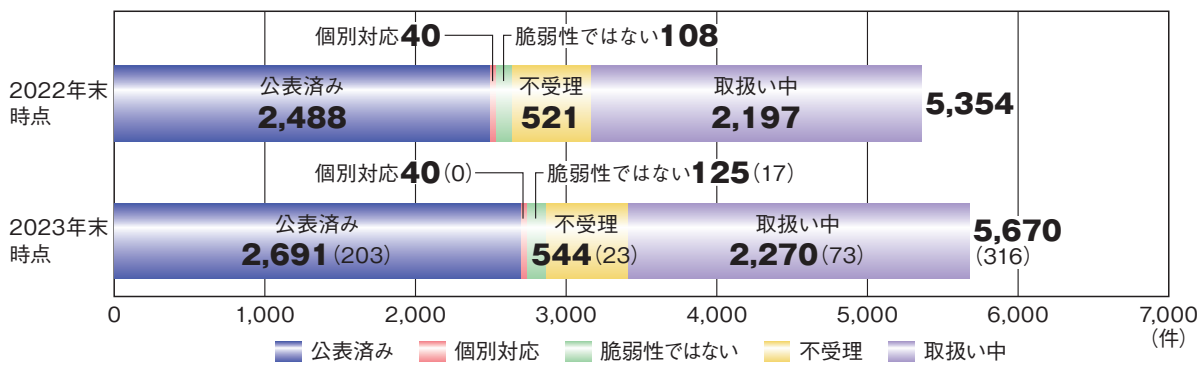
ソフトウェア製品に関する脆弱性届出の 2023 年における処理件数及び 2023 年末時点での処理状況別の累計件数について図 C-2 に示す。

2023 年の届出のうち、JPCERT/CC が調整を行い、製品開発者が脆弱性の修正を完了し、JVN で対策情報を公表した「公表済み」のものは 203 件で累計 2,691 件、JVN で公表せず製品開発者が「個別対応」を行ったものは 0 件で累計 40 件、製品開発者が「脆弱性ではない」と判断したものは 17 件で累計 125 件、告示で定める届出の対象に該当せず「不受理」としたものは 23 件で累計 544 件となり、これらをまとめた「処理の終了」

件数は 243 件で累計 3,400 件に達した。また、「取扱い中」の届出は 73 件増加して 2,270 件となり、ソフトウェア製品に関する届出は累計 5,670 件となった。

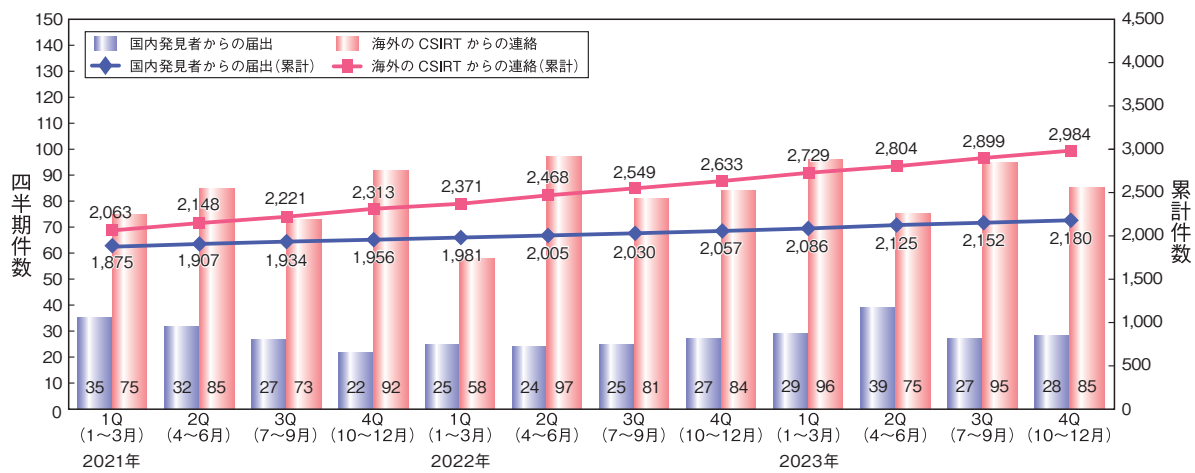
ソフトウェア製品の脆弱性対策情報の公表件数の累計は、国内発見者からの届出を公表したものが 2,180 件、海外の CSIRT から JPCERT/CC が連絡を受けたものを JVN で公表したものが 2,984 件となった。これらソフトウェア製品の脆弱性対策情報の公表件数の期別推移を図 C-3 に示す。

なお、複数の届出についてまとめて 1 件の脆弱性対策情報として公表する場合があるため、図 C-2 の「公表済み」の件数と図 C-3 の公表件数は異なっている。



※ ( )内の数値は2022年末時点と2023年末時点の差分

■ 図 C-2 ソフトウェア製品の脆弱性関連情報の届出の処理状況の推移



■ 図 C-3 ソフトウェア製品の脆弱性対策情報の公表件数

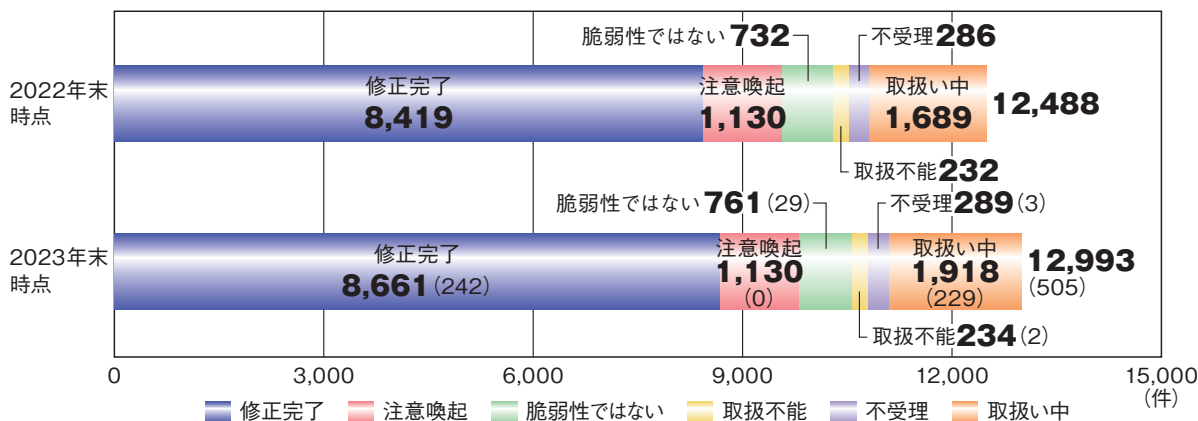
### C.3 Webサイトの脆弱性届出の処理状況

Webサイトに関する脆弱性届出の2023年における処理件数及び2023年末時点での処理状況別の累計件数について図C-4に示す。

2023年の届出のうち、IPAが通知を行いWebサイト運営者が「修正完了」としたものは242件で累計8,661件、IPAが「注意喚起」等を行った後に処理を終了したものは0件で累計1,130件、IPA及びWebサイト運営者が「脆弱性ではない」と判断したものは29件で累計761件、Webサイト運営者と連絡が不可能なもの、また

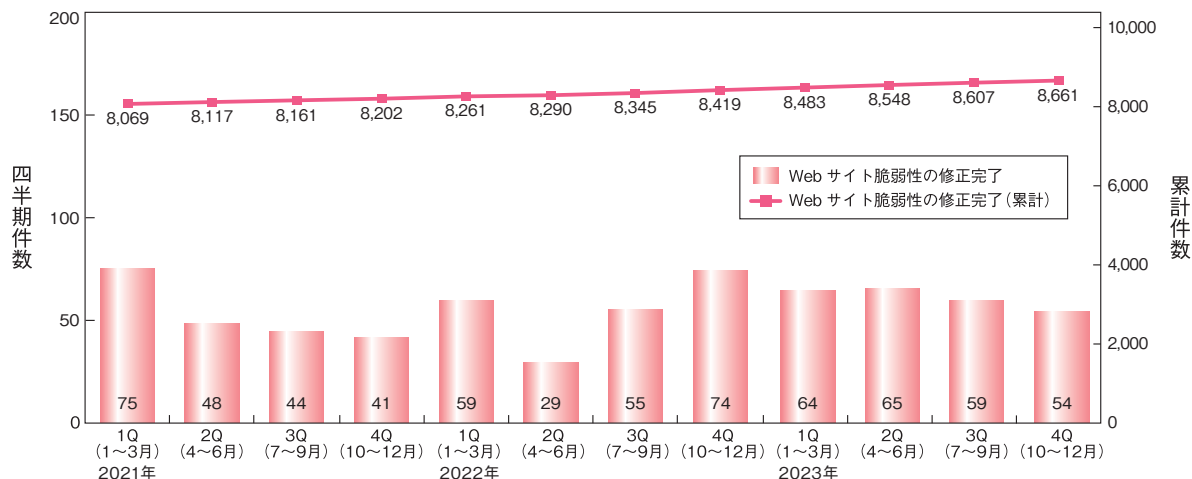
はIPAが対応を促しても修正完了した旨の報告をしない、修正を拒否する等、Webサイト運営者の対応により「取扱不能」なものは2件で累計234件、告示で定める届出の対象に該当せず「不受理」としたものは3件で累計289件となり、これらをまとめた「処理の終了」件数は276件で累計1万1,075件に達した。また、「取扱い中」の届出は229件増加して1,918件となり、Webサイトに関する届出は累計1万2,993件となった。

これらのうち、「修正完了」件数の期別推移を図C-5に示す。



※( )内の数値は2022年末時点と2023年末時点の差分

■ 図 C-4 Web サイトの脆弱性関連情報の届出の処理状況の推移



■ 図 C-5 Web サイトの脆弱性の修正完了件数

参照

■ ソフトウェア等の脆弱性関連情報に関する届出状況 [2023年第4四半期(10月~12月)]  
<https://www.ipa.go.jp/security/reports/vuln/software/2023q4.html>

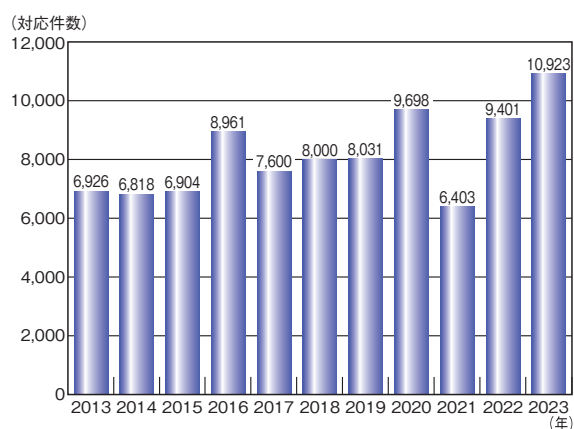


## 資料D 2023年の情報セキュリティ安心相談窓口の相談状況

IPA が 2023 年 1 月から 12 月の期間に対応した、相談状況の集計結果について述べる。

### D.1 相談対応件数

2023 年の年間相談対応件数は 10,923 件となり、2022 年の相談対応件数 9,401 件より 1,522 件（16.2%）の増加となった（図 D-1）。



■図 D-1 相談対応件数の推移（2013～2023 年）

### D.2 相談者の主体別相談件数

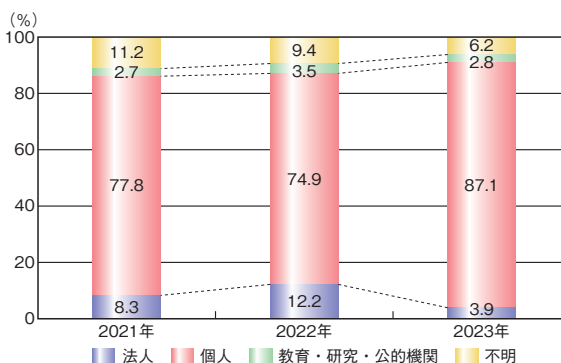
相談者の主体別では、2023 年も個人からの相談が 9,514 件（87.1%）と最も多かった。

主体別相談比率の推移では、法人からの相談比率は 2022 年と比較して 8.3% 減少した一方、個人からの相談比率は 12.2% 増加した（表 D-1、図 D-2）。

法人については、2022 年に多かった「Emotet 関連」の相談の減少が、要因の一つと考えられる。また個人については、「ウイルス警告の偽警告」についての相談の増加が要因の一つと考えられる（「D.4 手口別相談件数」参照）。

相談者の主体	2021 年	2022 年	2023 年
法人	530	1,145	427
個人	4,984	7,043	9,514
教育・研究・公的機関	170	330	308
不明	719	883	674
合計（件）	6,403	9,401	10,923

■表 D-1 情報セキュリティ安心相談窓口の主体別相談件数（2021～2023 年）



■図 D-2 情報セキュリティ安心相談窓口の主体別相談件数の比率推移（2021～2023 年）

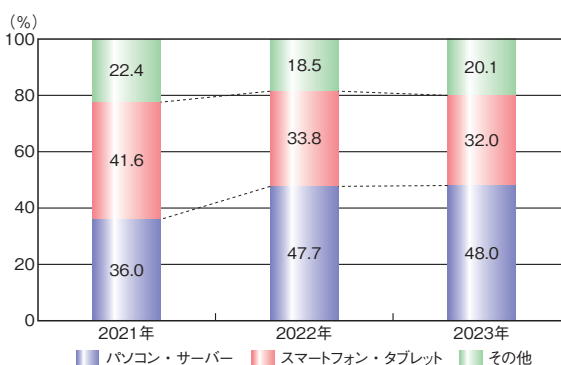
### D.3 相談者の機器種別相談件数

相談機器種別では、2023 年は「パソコン・サーバー」に関する相談が 5,240 件（48.0%）と最も多かった。

相談者の機器種別相談比率は、2022 年と比較して同じ水準で推移しており、大きな変化はなかった（表 D-2、図 D-3）。

相談機器種別の主体	2021 年	2022 年	2023 年
パソコン・サーバー	2,304	4,487	5,240
スマートフォン・タブレット	2,666	3,173	3,492
その他	1,433	1,741	2,191
合計（件）	6,403	9,401	10,923

■表 D-2 情報セキュリティ安心相談窓口の機器種別相談件数（2021～2023 年）



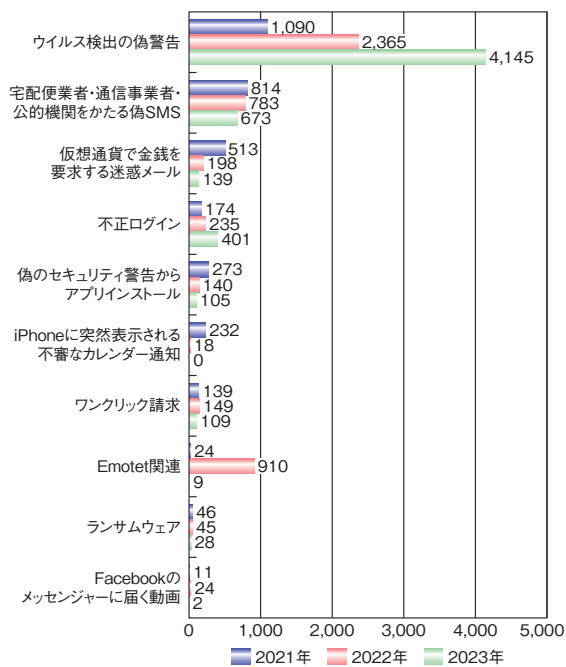
■図 D-3 情報セキュリティ安心相談窓口の機器種別相談件数の比率推移（2021～2023 年）

### D.4

#### 手口別相談件数

主要手口ごとの相談件数を図 D-4 に示す。2023 年の相談で最も多く寄せられたのは、「ウイルス検出の偽警告」に関する相談で4,145件(37.9%)であった。次いで、「宅配便業者・通信事業者・公的機関をかたる偽SMS」に関する相談が673件(6.2%)、「不正ログイン」に関する相談が401件(3.7%)であった。上位三つの手口による相談件数の合計は5,219件で、全相談件数(10,923件)の47.8%であった。

問い合わせの多い手口については、情報セキュリティ安心相談窓口の発行する「安心相談窓口だより」や、「手口検証動画」で注意喚起を行っている。ぜひ参考にしてほしい。



■ 図 D-4 主要手口別相談件数の推移 (2021~2023年)

**参照**

- 安心相談窓口だより  
<https://www.ipa.go.jp/security/anshin/attention/index.html>
- 手口検証動画シリーズ  
<https://www.ipa.go.jp/security/anshin/measures/verificationmov.html>



第19回 IPA

# 「ひろげよう情報セキュリティ コンクール」2023 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53,312点の応募作品の中から、受賞した作品の一部をご紹介します。

## 最優秀賞

〈独立行政法人情報処理推進機構〉

### 〈標語部門〉

それでいい？  
使いまわしの  
パスワード

大阪府 大阪市立大淀小学校 5年 今岡 陽菜歌さん

### 〈ポスター部門〉

扱いに注意！君の味方は敵にもなる



神奈川県 神奈川県立神奈川工業高等学校 3年 村石 琉音さん

### 〈4コマ漫画部門〉

フィッシング



兵庫県 西宮市立鳴尾中学校 3年  
奥埜 和花さん

# 優秀賞

〈独立行政法人情報処理推進機構〉

## 〈標語部門〉

信じるの 知らない人の その言葉

大阪府 堺市立南八下小学校 4年  
市ノ瀬 瑚珀さん

セキュリティ 「面倒くさい」が 命とり

大阪府 大阪教育大学附属平野中学校 1年  
稲垣 敢太さん

詐欺メール 「緊急」「至急」疑おう

大阪府 東大谷高等学校 1年  
小倉 結子さん

## 〈ポスター部門〉

フィッシングに注意!



兵庫県 雲雀丘学園小学校 6年  
オストハイダ 真紋さん

“Fake” Wi-Fi



愛媛県 松山市立勝山中学校 3年  
渡辺 梨緒さん



兵庫県 兵庫県立姫路工業高等学校 2年  
川上 心優さん

## 〈4コマ漫画部門〉

なぞかけ



岡山県 岡山市立御津南小学校 5年  
北山 穂風さん

「まあ、いっか。」の結末は…。



広島県 呉市立東畑中学校 3年  
長尾 妃芽さん

こわ〜い話



広島県 広島県立呉商業高等学校 3年  
井上 心彩さん

## IPAの便利なツールとコンテンツ

情報セキュリティ対策ベンチマーク		診断
用途・目的	自組織のセキュリティレベルを診断	
利用対象者	情報セキュリティ担当者	
特長	<ul style="list-style-type: none"><li>他組織と比較した自組織のセキュリティレベルが判る</li><li>自組織に不足しているセキュリティ対策が判る</li></ul>	
<b>概要</b>		
「セキュリティ対策の取り組み状況に関する評価項目」27問と「企業プロフィールに関する評価項目」19問、計46問に回答すると以下の診断結果を表示します。		
<b>■提供される診断結果</b>		
<ul style="list-style-type: none"><li>セキュリティレベルを示したスコア(最高点135点、最低点27点)</li><li>情報セキュリティリスクの指標と企業規模、業種が自組織と近い他組織について診断項目別に比較</li><li>結果に応じた推奨される取り組み</li></ul>		



脆弱性体験学習ツール「AppGoat」		学習
用途・目的	脆弱性に関する基礎的な知識の学習	
利用対象者	<ul style="list-style-type: none"><li>アプリケーション開発者</li><li>Webサイト管理者</li></ul>	
特長	脆弱性の概要や対策方法等、脆弱性に関する基礎的な知識を実習形式で体系的に学べるツール	
<b>概要</b>		
SQLインジェクション、クロスサイト・スクリプティング等の12種類のWebアプリケーションに関連する脆弱性について学習できるツールです。利用者は学習テーマ毎の演習問題に対して、埋め込まれた脆弱性の発見、プログラミング上の問題点の把握、対策手法を学べます。		
<b>■活用方法例</b>		
<ul style="list-style-type: none"><li>Webアプリケーション用学習ツール(個人学習モード)を利用した、自宅等での個人学習</li><li>Webアプリケーション用学習ツール(集合学習モード)を利用した、学校の講義や組織内のセミナー等における複数人での学習</li></ul>		
<b>■動作環境・必須ソフトウェア</b>		
Windows 10、11		

脆弱性対策情報データベース「JVN iPedia」		対策
用途・目的	自組織で使用しているソフトウェア製品の脆弱性の確認と対策	
利用対象者	<ul style="list-style-type: none"><li>システム管理者</li><li>製品・サービスの保守を担う担当者</li></ul>	
特長	国内外で公開されたソフトウェア製品の脆弱性対策情報が掲載された、キーワード検索可能なデータベース	
<b>概要</b>		
<b>■掲載情報例</b>		
<ul style="list-style-type: none"><li>脆弱性の概要</li><li>脆弱性の深刻度 CVSS 基本値</li><li>脆弱性がある製品名とそのベンダー名</li><li>本脆弱性に関わる製品ベンダー等のリンク</li><li>共通脆弱性識別子 CVE</li></ul>		
<b>■活用方法例</b>		
<ul style="list-style-type: none"><li>ネット記事等に記載された CVE 番号を JVN iPedia で検索し、脆弱性の詳細を確認</li><li>自組織で使用している製品名で検索し、脆弱性の詳細を確認</li></ul>		

## MyJVN バージョンチェッカ for .NET

<https://jvndb.jvn.jp/apis/myjvn/vccheckdotnet.html>



用途・目的	パソコンにインストールされたソフトウェア製品のバージョンが最新かどうかの確認
利用対象者	パソコン利用者全般
特長	インストールされている対象製品が最新バージョンかどうかをまとめて確認できる
<b>概要</b>	
<b>■判定対象ソフトウェア製品</b>	
<ul style="list-style-type: none"><li>• Adobe Reader</li><li>• JRE</li><li>• Lhaplus</li><li>• Mozilla Firefox</li><li>• Mozilla Thunderbird</li><li>• iTunes</li><li>• Lunascape</li><li>• Becky! Internet Mail</li><li>• OpenOffice.org</li><li>• VMware Player</li><li>• Google Chrome</li><li>• LibreOffice</li></ul>	
<b>■活用方法例</b>	
毎朝、MyJVN バージョンチェッカを実行して、使用しているソフトウェアが最新かどうかをチェックし、最新でなければそのソフトウェアを更新する	
<b>■動作環境・必須ソフトウェア</b>	
Windows 10、11	

## 注意警戒情報サービス

<https://jvndb.jvn.jp/alert/>



用途・目的	脆弱性対策に必要な最新情報の収集
利用対象者	<ul style="list-style-type: none"><li>• システム管理者</li><li>• 製品・サービスの保守を担う担当者</li></ul>
特長	国内で広く利用され、脆弱性が悪用されると影響の大きいサーバー用オープンソースソフトウェアのリリース情報と IPA が発信する「重要なセキュリティ情報」を提供
<b>概要</b>	
<b>■掲載情報例</b>	
<ul style="list-style-type: none"><li>• Apache HTTP Server</li><li>• Apache Struts</li><li>• Apache Tomcat</li><li>• BIND</li><li>• Joomla!</li><li>• OpenSSL</li><li>• WordPress</li><li>• 重要なセキュリティ情報</li></ul>	
<b>■活用方法例</b>	
定期的に自組織で使用しているオープンソースソフトウェアのリリース情報や IPA が発信する「重要なセキュリティ情報」が公表されているかどうかを確認し、公表されていれば内容の確認、必要に応じ対応を行う	

## サイバーセキュリティ注意喚起サービス「icat for JSON」

<https://www.ipa.go.jp/security/vuln/icat.html>



用途・目的	IPA が発信する「重要なセキュリティ情報」のリアルタイム取得
利用対象者	<ul style="list-style-type: none"><li>• システム管理者</li><li>• サービスの保守を担う担当者</li><li>• 個人利用者</li></ul>
特長	Web ページに HTML タグを埋め込むと、Web ページから IPA が発信する「重要なセキュリティ情報」を配信
<b>概要</b>	
<b>■「重要なセキュリティ情報」発信例</b>	
<ul style="list-style-type: none"><li>• 利用者への影響が大きい製品の脆弱性情報</li><li>• 広く使われる製品のサポート終了情報</li><li>• サイバー攻撃への注意喚起</li></ul>	
<b>■活用方法例</b>	
icat を自組織の従業員がよくアクセスする Web ページ（イントラページ等）に表示させ、ソフトウェア更新等の対策を促す	

## MyJVN 脆弱性対策情報フィルタリング収集ツール(mjcheck4)

<https://jvndb.jvn.jp/apis/myjvn/mjcheck4.html>



用途・目的	自組織で使用しているソフトウェア製品の脆弱性の確認と対策
利用対象者	<ul style="list-style-type: none"><li>・システム管理者</li><li>・製品・サービスの保守を担う担当者</li></ul>
特長	JVN iPedia に登録されている脆弱性対策情報をフィルタリングして自社システムに関連する脆弱性情報を効率よく収集

### 概要

#### ■フィルタリング例

- ・製品名
- ・CVSSv3
- ・公開日 等

#### ■活用方法例

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

#### ■動作環境・必須ソフトウェア

Windows 10、11

## Web サイトの攻撃兆候検出ツール「iLogScanner」

<https://www.ipa.go.jp/security/vuln/ilogscanner/>



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

### 概要

#### ■アクセスログ、エラーログから検出可能な項目例

- ・SQL インジェクション
- ・OS コマンド・インジェクション
- ・ディレクトリ・トラバーサル
- ・クロスサイト・スクリプティング

#### ■認証ログ(Secure Shell、FTP)から検出可能な項目例

- ・大量のログイン失敗
- ・短時間の集中ログイン
- ・同一ファイルへの大量アクセス
- ・認証試行回数

#### ■活用方法例

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

## 5分できる！情報セキュリティ自社診断

<https://www.ipa.go.jp/security/guide/sme/5minutes.html>







用途・目的	自社の情報セキュリティ対策状況を診断
利用対象者	中小企業・小規模事業者の経営者、管理者、従業員
特長	<ul style="list-style-type: none"><li>・設問に答えるだけで自社のセキュリティ対策状況を把握することができる</li><li>・診断後は、診断結果に即した対策が確認できる</li></ul>


### 概要



「5分できる！情報セキュリティ自社診断」は、情報セキュリティ対策のレベルを数値化し、問題点を見つけるためのツールです。

25の質問に答えるだけで診断することができ、解説編を参照することで、診断編にある設問の内容を自社で対応していない場合に生じる情報セキュリティへのリスクと、今後どのような対策を設けるべきかを把握することができます。



<b>情報セキュリティ・ポータルサイト「ここからセキュリティ！」</b> <a href="https://www.ipa.go.jp/security/kokokara/">https://www.ipa.go.jp/security/kokokara/</a>				
用途・目的	<ul style="list-style-type: none"> <li>情報セキュリティや情報リテラシーに関する情報収集</li> <li>国内の主なレポート、ガイドライン、学習・診断等のツール等の利用</li> </ul>			
利用対象者	<ul style="list-style-type: none"> <li>インターネットの一般利用者(小学生~大人)</li> <li>企業の管理者/一般利用者</li> </ul>			
特長	情報セキュリティ関連の民間及び公的な団体が公開する無償の資料、情報、ツールを網羅的に掲載。目的別、用途別、役割別に情報を選択し利用が可能			
<b>概要</b>				
<ul style="list-style-type: none"> <li>セキュリティベンダー、公的機関、政府等から発信される注意喚起や、資料・動画・ツール等のコンテンツを網羅的に掲載したポータルサイト</li> <li>コンテンツを「被害に遭ったら」「対策する」「教育・学習」「セキュリティチェック」「データ &amp; レポート」に分類。必要な情報が見つかりやすい</li> <li>教育学習は対象者を細分化し、それぞれに適した教育学習コンテンツを紹介</li> </ul>				

<b>サイバーセキュリティ経営可視化ツール</b> <a href="https://www.ipa.go.jp/security/economics/checktool.html">https://www.ipa.go.jp/security/economics/checktool.html</a>		
用途・目的	セキュリティ対策の実施状況のセルフチェック	
利用対象者	原則として、従業員 300 名以上の企業の CISO 等、サイバーセキュリティ対策の実施責任者	
特長	サイバーセキュリティ経営ガイドライン Ver3.0 に準拠したセキュリティ対策の実施状況を成熟度モデルで自己診断し、レーダーチャートで可視化	
<b>概要</b>		
<p>経営者がサイバーセキュリティ対策を実施する上で責任者となる担当幹部（CISO 等）に指示すべき“重要 10 項目”が、適切に実施されているかどうかを 5 段階の成熟度モデルで自己診断し、その結果をレーダーチャートで可視化するツールです。</p> <p>診断結果は、経営者への自社のセキュリティ対策の実施状況の説明資料として利用できます。経営者が対策状況を定量的に把握することで、サイバーセキュリティに関する方針の策定や適切なセキュリティ投資の検討、投資家等ステークホルダとのコミュニケーション等に役立てることができます。</p> <p>■提供される主な機能</p> <ul style="list-style-type: none"> <li>重要 10 項目の実施状況の可視化</li> <li>診断結果と業種平均との比較</li> <li>対策を実施する際の参考事例</li> <li>グループ企業同士の診断結果の比較</li> </ul>		

<b>5分でできる！情報セキュリティポイント学習</b> <a href="https://www.ipa.go.jp/security/sec-tools/5mins_point.html">https://www.ipa.go.jp/security/sec-tools/5mins_point.html</a>		
用途・目的	自社の情報セキュリティ教育の実施	
利用対象者	中小企業の経営者、管理者、従業員等	
特長	<ul style="list-style-type: none"> <li>自社診断の質問を 1 テーマ 5 分で学べる</li> <li>インストール不要、無料の学習ツール</li> </ul>	
<b>概要</b>		
情報セキュリティについて学習できるツールです。身近にある職場の日常の 1 コマを取り入れた親しみやすい学習テーマで、セキュリティに関する様々な事例を疑似体験しながら適切な対処法を学ぶことができます。		

付録



## 安心相談窓口だより

<https://www.ipa.go.jp/security/anshin/attention/index.html>



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

### 概要

IPA 情報セキュリティ安心相談窓口では、寄せられる相談に関して手口を実際に検証し、そこで得られた知見をその後の相談対応にフィードバックするとともに、注意喚起等、情報発信にも活かしています。

「安心相談窓口だより」では中でも多く相談が寄せられる相談内容の「手口」「対処」「対策」について、パソコンやスマートフォンの操作等にあまり詳しくない人でも理解できるように分かりやすく説明を行っています。

記事は不定期に公開されますので、「安心相談窓口だより」を定期的に確認することで、最新のネット詐欺等の手口や対策を知り、被害の未然防止に役立てることができます。

手口に関する内容以外にも、被害にあわないための日ごろから気を付けるポイントについての記事も公開しています。



## 映像で知る情報セキュリティ 各種映像コンテンツ

<https://www.ipa.go.jp/security/videos/list.html>



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

### 概要

「サイバー攻撃」「内部不正」「ワンクリック請求」「偽警告」等の脅威をテーマにした動画のほか、「中小企業向け情報セキュリティ対策」「新入社員向け」「保護者／小学生／中高生向け」といった訴求対象者別の動画を公開しています。動画の視聴により、スマートフォン・パソコンを使用する際に利用者に求められる振舞いや対策を身に付けることができます。

情報セキュリティの自己研さんを目的とした個人の視聴のほか、組織内の研修用としての利用が可能です。

#### ■動画のタイトル例

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



# 索引

## A

- AI(Artificial Intelligence : 人工知能)  
.....9, 97, 101, 132, 224
- AiTM(adversary-in-the-middle) ..... 33
- AI 安全性サミット(AI Safety Summit) ..... 98
- AI 事業者ガイドライン .....73, 80, 227, 235
- AI セーフティ・インスティテュート  
..... 73, 102, 111, 221, 227
- AI 戦略 ..... 73
- AI の民主化 ..... 225
- AI リスクマネジメントフレームワーク(AI RMF : AI  
Risk Management Framework) ... 102, 225, 235
- 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) ..... 114
- APT12 ..... 216
- APT(Advanced Persistent Threat) 攻撃  
.....24, 172, 188, 209
- Artificial Intelligence Act(AI 法) .... 110, 224, 227
- ASEAN 地域フォーラム(ARF : ASEAN Regional  
Forum) ..... 72
- ASM(Attack Surface Management) 導入ガイド  
ンス .....27, 82
- Attack Surface Management(ASM) ... 27, 75, 82

## B

- BlackTech ..... 25, 94, 189

## C

- C&C(Command and Control) サーバー  
.....24, 35, 88, 94, 185
- Camaro Dragon ..... 179
- CCRA(Common Criteria Recognition  
Arrangement) ..... 129, 159
- CEO 詐欺 ..... 29, 32
- CI / CD パイプラインにおけるセキュリティの留意点  
に関する技術レポート ..... 75
- Citrix Bleed ..... 36, 57
- Clop(CI0p) ..... 10, 38
- CMVP(Cryptographic Module Validation  
Program) ..... 163

- CNA(CVE Numbering Authority) ..... 54
- CosmicEnergy ..... 175
- CRYPTREC ..... 73, 167
- CSIRT(Computer Security Incident Response  
Team) ..... 26, 33, 112, 114, 155, 172
- CVE(Common Vulnerabilities and Exposures :  
共通脆弱性識別子) ..... 54, 174, 179
- Cyber Av3ngers ..... 171
- CYROP(CYber Range Open Platform) ..... 121
- CYXROSS ..... 70

## D

- DDoS 攻撃 ..... 33, 35, 95, 179, 188
- DNS(Domain Name System) ..... 34, 188
- DSA(Digital Signature Algorithm) ..... 169
- DX 推進スキル標準(DSS-P) ..... 116
- DX リテラシー標準(DSS-L) ..... 116

## E

- Earth Kasha ..... 24
- ECDSA ..... 169
- EC サイト構築・運用セキュリティガイドライン ..... 62
- EDR(Endpoint Detection and Response)  
..... 21, 27, 150
- Emotet ..... 156
- EO 14028 ..... 105
- EO 14110 ..... 101, 104, 235
- ESXiArgs ..... 10
- EUCC(European cybersecurity certification  
scheme) ..... 129
- EU サイバーレジリエンス法案(CRA : EU Cyber  
Resilience Act) ..... 105, 108, 177, 189
- e- ネットキャラバン ..... 69

## G

- G7 広島サミット ..... 35, 71, 95, 98
- GDPR(General Data Protection Regulation :  
EU 一般データ保護規則) ..... 106, 111

## I

- ICT サイバーセキュリティ総合対策 ..... 86
- IEC(International Electrotechnical  
Commission : 国際電気標準会議) ..... 126

IEEE (The Institute of Electrical and Electronics Engineers, Inc.)	127
IETF (Internet Engineering Task Force)	127
IoC (Indicator of Compromise : 侵害指標)	21, 106
IoT	35, 69, 86, 130, 136, 179
IoT-domotics	131
IoT 製品に対するセキュリティ適合性評価制度	79, 162, 189
IoT セキュリティガイドライン	130
IoT ボットネット対策	86
ISA/IEC 62443 シリーズ	137
ISMAP-LIU (イスマップ・エルアイユー : ISMAP for Low-Impact Use)	70, 164
ISMAP 管理基準	164, 165
ISMAP クラウドサービスリスト	164
ISO (International Organization for Standardization : 国際標準化機構)	126
ISO/IEC 15408	129, 159, 161
ISO/IEC 27000 ファミリー	128, 198
ISO/IEC JTC 1/SC 27	127
ITSS+	118
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)	126, 135
IT スキル標準 (ITSS)	118
IT 製品の調達におけるセキュリティ要件リスト	159
IT セキュリティ評価及び認証制度 (JISEC : Japan Information Technology Security Evaluation and Certification Scheme)	79, 159, 163
<b>J</b>	
J-CRAT (Cyber Rescue and Advice Team against targeted attack of Japan : サイバーレスキュー隊)	23, 85
JTC 1 (Joint Technical Committee 1 : 第一合同技術委員会)	126
JVN iPedia	54, 57
<b>L</b>	
Lattice Attack	169
LockBit	11, 19, 69, 94, 109, 173

## M

Microsoft Office	37
Mirai	92, 179, 183, 185, 187
MOVEit Transfer	10, 38, 56
Mustang Panda	25

## N

NICTER (Network Incident analysis Center for Tactical Emergency Response)	87, 187
NIS 指令 (Network and Information Systems Directive) ・ NIS2 指令	107, 177
NOTICE (National Operation Towards IoT Clean Environment)	69, 87, 187
NVD (National Vulnerability Database)	54

## O

OSINT (Open Source Intelligence)	213, 231
----------------------------------	----------

## P

PIMS (Privacy Information Management System : プライバシー情報マネジメントシステム)	135
Play	173
Proself	24, 38

## R

RomCom	38
--------	----

## S

SaaS	70, 164, 192, 193, 198
Sandworm	172
SBD (Security By Design) マニュアル	70
SC3 セキュリティ人材育成フレームワーク	118
SECCON	122
SecHack365	122
SECURITY ACTION	148, 153
Shields Ready	175
SIM スワップ	94
SMS (ショートメッセージ)	12, 39, 42, 158
Software Bill of Materials (SBOM : ソフトウェア部品表)	69, 78, 105, 176, 235
SQL インジェクション	38, 55, 61

Storm-0558	25
Storm-0978	38

## T

TCG(Trusted Computing Group)	127
Telegram	213, 220
Tropic Trooper	24
Trustworthy AI	111, 227, 235

## U

U.S. Cyber Trust Mark プログラム	105
UNC4841	25

## V

Volt Typhoon	8, 106, 188
VPN	18, 23, 36, 84, 93, 159

## W

Web サイト改ざん	15, 58
Windows	44, 45, 126
WispRider	25

## あ

アイデンティティ管理	134
暗号鍵管理システム設計指針(基本編)	167
暗号資産	72, 90, 93, 183, 188
暗号モジュール試験及び認証制度(JCMVP : Japan Cryptographic Module Validation Program)	163
安全なウェブサイトの作り方	62
安全保障等の機微な情報等に係る政府情報システムの取扱い	76
安保 3 文書	116
イスラエル・ハマスの武力衝突	107, 212, 232
イスラエル・パレスチナ情勢	97
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)	47, 94
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)	12, 22, 84, 100, 115, 185
インターネットトラブル事例集 2023 年版	158

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	100
インフォデミック	219
ウェブ健康診断仕様	62
営業秘密	51, 80, 82, 150, 226, 233
エコチェンバー	212, 222
遠隔操作アプリ(ソフトウェア)	43, 44, 47, 48
遠隔操作ウイルス(RAT : Remote Access Trojan)	20, 231
欧州刑事警察機構(Europol : European Union Agency for Law Enforcement Cooperation)	69, 94, 98, 100, 109
オープンソースソフトウェア(OSS : Open Source Software)	69, 105, 108, 177, 227
オープンリダイレクト(Open Redirect)	61
お助け隊サービス 2 類	153

## か

環太平洋パートナーシップ協定(TPP 協定 : Trans-Pacific Partnership Agreement)	107
機械学習システムセキュリティガイドライン Version 2.00	235
機器検証サービス	69, 79, 83
偽・誤情報	157, 209
技術情報管理認証制度	82, 151
業界別サイバーレジリエンス強化演習(CyberREX : Cyber Resilience Enhancement eXercise by industry)	124
共通鍵暗号	168
共通脆弱性タイプ一覧(CWE : Common Weakness Enumeration)	54
共通脆弱性評価システム(CVSS : Common Vulnerability Scoring System)	38, 55, 75
虚偽情報	109, 156, 208
クラウドサービス	19, 33, 51, 159, 164, 192
クラウドサービスの安全性評価に関する検討会	164
クレジットカード	12, 41, 82, 92, 156
クロスサイト・スクリプティング	55, 61
経営者向けインシデント対応机上演習	153
経済安全保障重要技術育成プログラム(K Program)	72
経済安全保障推進法	73
軽量暗号	167, 169, 190

公開鍵暗号	169, 197		
攻撃対象領域(アタックサーフェス)	21, 27, 132, 149		
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	78, 178		
国立研究開発法人情報通信研究機構(NICT: National Institute of Information and Communications Technology)	69, 87, 89, 121, 167, 187		
国立情報学研究所(NII: National Institute of Informatics)ストラテジックサイバーレジリエンス研究開発センター	71		
個人情報保護委員会	19, 44, 71, 156, 195, 233		
コネクテッドカー	182		
コモンクライテリア(共通基準)	159, 160		
コラボレーション・プラットフォーム	79, 155		
<b>さ</b>			
最高 AI 責任者(CAIO: Chief AI Officer)	101		
最高情報セキュリティ責任者(CISO: Chief Information Security Officer)	91, 113, 124, 148, 154		
サイドチャネル攻撃	130, 169, 170		
サイバーインテリジェンス情報共有ネットワーク	94		
サイバー危機対応机上演習(CyberCREST: Cyber Crisis REsponse Table top exercise)	124		
サイバー警察局	69, 90, 92, 117		
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)	13, 29, 83		
サイバーセキュリティ 2023	68, 177		
サイバーセキュリティお助け隊サービス	69, 79, 153		
サイバーセキュリティ経営ガイドライン Ver3.0 実践のためのプラクティス集	68, 78, 154		
サイバーセキュリティ企画演習(CyberSPEX: Cyber Security Planning Exercise)	125		
サイバーセキュリティ協議会	71		
サイバーセキュリティ経営ガイドライン	26, 68, 78, 149, 154		
サイバーセキュリティ経営可視化ツール	68, 78, 154		
サイバーセキュリティ経営戦略コース	123		
サイバーセキュリティ戦略	68, 100, 103, 112, 176		
サイバーセキュリティ体制構築・人材確保の手引き	149		
サイバーセキュリティネクサス(CYNEX: Cyber Security NEXUS)	69, 121		
サイバーセキュリティフレームワーク(CSF: Cyber Security Framework)	104, 175, 176		
サイバー特別捜査隊	69, 90, 94, 98		
サイバーフィジカルシステム(CPS: Cyber Physical System)	134, 226, 232		
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: the Cyber/Physical Security Framework)	77, 134		
サイバーレジリエンス	26, 74, 106		
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)	69, 78, 151		
サプライチェーンリスク	69, 104, 149		
サポート詐欺	43, 48, 158		
産学情報セキュリティ人材育成交流会	123		
産業競争力強化法等の一部を改正する法律	82		
産業サイバーセキュリティ研究会	76, 117, 189		
産業サイバーセキュリティセンター(ICSCoE: Industrial Cyber Security Center of Excellence)	86, 123, 177, 178		
産業用制御システム向け侵入検知製品等の導入手引書	178		
事業継続計画(BCP: Business Continuity Plan)	22, 26, 197		
実践的サイバー防御演習(CYDER: CYber Defense Exercise with Recurrence)	100, 121		
自由で開かれたインド太平洋	100		
重要インフラにおける情報セキュリティ確保に係る安全基準等策定指針	68		
重要インフラのサイバーセキュリティに係る行動計画	70, 73, 177		
重要インフラのサイバーセキュリティに係る安全基準等策定指針	69, 70, 165, 177		
常時リスク診断・対処(CRSA)のエンタープライズアーキテクチャ(EA)	74		
情報処理安全確保支援士(登録セキスベ)	119		
情報セキュリティ安心相談窓口	39, 92		
情報セキュリティサービス基準	69, 83		
情報セキュリティサービス基準適合サービスリスト	79, 83		

情報セキュリティサービス審査登録制度	69, 79, 83	セキュアソフトウェア開発フレームワーク(SSDF)	235
情報セキュリティサービスに関する審査登録機関基準	83	セキュリティ・キャンプ	120
情報セキュリティ早期警戒パートナーシップ	58	セキュリティ・クリアランス制度	73
情報セキュリティマネジメント試験	119	セキュリティ・バイ・デザイン(セキュア・バイ・デザイン)	70, 74, 104, 235
情報セキュリティマネジメントシステム(ISMS : Information Security Management System)	127, 151, 198, 225	ゼロデイ脆弱性	25, 37, 56, 85, 172, 180
情報戦	209	ゼロトラストアーキテクチャ	70, 74
情報操作型サイバー攻撃	208, 209, 222	組織における内部不正防止ガイドライン	51, 150
情報漏えい	11, 48, 58, 150, 193, 233	ソフトウェア管理に向けた SBOM(Software Bill of Materials)の導入に関する手引	69
新型コロナウイルス	37, 97, 115, 208, 218	<b>た</b>	
人工知能システムのセキュリティ脅威に対処するためのガイダンス	132	ダークウェブ	11, 21, 94, 188
侵入型ランサムウェア攻撃	17, 20, 21	耐量子計算機暗号	167, 169
推論攻撃	234	地域 SECURITY	69, 79, 152
スマートカード	159, 161	中核人材育成プログラム	123
スマート工場化でのシステムセキュリティ対策事例調査報告書	178	中小企業の情報セキュリティ対策ガイドライン	153, 154, 197
制御システム(ICS : Industrial Control System)	171	ディープフェイク	28, 101, 212, 216, 225, 231
制御システムのセキュリティリスク分析ガイド	154, 178	ディスインフォメーション(Disinformation)	208, 210, 215, 221
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)	125	データガバナンス法(Data Governance Act)	109
脆弱性	21, 26, 54, 173, 186, 231	データポイズニング	234
生成 AI(Generative AI)	58, 97, 101, 156, 208, 224	敵対的サンプル(Adversarial sample)	234
政府機関等における情報システム運用継続計画ガイドライン	70	デジタル空間における情報流通の健全性確保の在り方に関する検討会	217, 220
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	74, 159, 163	デジタルサービス法(DSA : Digital Services Act)	97, 109
政府機関等の対策基準策定のためのガイドライン	83, 163	デジタル市場法(DMA : Digital Markets Act)	109
政府情報システムにおける脆弱性診断導入ガイドライン	74	デジタル社会推進標準ガイドライン	74, 75
政府情報システムにおけるセキュリティ・バイ・デザインガイドライン	74	デジタル人材育成プラットフォーム	116
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))	70, 83, 164	デジタルスキル標準	116
責任共有モデル	196	テレワーク	14, 37, 50, 82
セキュア AI システム開発ガイドライン	235	電子署名	162, 163
		トラストサービス規準	198
		<b>な</b>	
		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)	25, 68, 100, 158, 165, 177
		内部不正	13, 51, 150, 234
		ナラティブ(Narrative)	209, 210, 223

なりすまし	29, 32, 39, 84, 173, 182
二重の脅迫(二重恐喝)	14, 17, 21, 93, 173
偽 EC サイト	43, 47
偽のセキュリティ警告	42, 43, 45
日 ASEAN サイバーセキュリティ政策会議	72, 99
日 ASEAN サイバーセキュリティ能力構築センター (Asean Japan Cybersecurity Capacity Building Centre : AJCCBC)	123
日 ASEAN 能力向上プログラム強化プロジェクト	99, 123
日米豪印サイバーセキュリティ・パートナーシップ：共 同原則	99
日本 ASEAN 友好協力 50 周年	99, 115
日本産業標準調査会 (JISC : Japanese Industrial Standards Committee)	126
認知戦	208, 210
ネット詐欺	42, 48
ネットワーク貫通型攻撃	23, 84
ノーウェアランサム攻撃	11, 14, 17, 21, 93

## は

バイオメトリクス	135
パスキー認証	196, 197
バックドア	234
ばらまき型メール	84
ハルシネーション	212, 226
万博向けサイバー防御講習 (CIDLE)	122
ビジネスメール詐欺 (BEC : Business Email Compromise)	9, 28, 32, 84
ビッグデータ	80, 135
標的型攻撃	23, 84, 85, 94, 172, 231
標的型サイバー攻撃特別相談窓口	85
広島 AI プロセス	73, 99, 224, 235
ファクトチェック	213, 221, 222
フィッシング	9, 12, 33, 39, 93, 231
フィルターバブル	212, 222
フェイクニュース	101, 157, 209
副業詐欺	43, 46, 48
不正アクセス	19, 23, 33, 49, 95, 196
不正競争防止法の改正	80
不正送金	43, 44, 94
プラス・セキュリティ人材	116, 117
プロテクションプロファイル (PP : Protection Profile)	160, 162

プロンプトインジェクション	234
米国国立標準技術研究所 (NIST : National Institute of Standards and Technology)	54, 70, 103, 163, 176, 225
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)	10, 74, 104, 171, 175
防衛産業サイバーセキュリティ基準	72, 77
ボットネット	35, 86, 179, 183, 185, 188

## ま

マイクロターゲティング	210, 222
マイナポータル	41, 70
マナビ DX (マナビ・デラックス)	116
マルインフォメーション (Malinformation)	208
ミスインフォメーション (Misinformation)	208
民間宇宙システムにおけるサイバーセキュリティ対策 ガイドライン	78
モデルインバージョン (Model inversion)	234

## ら

ランサムウェア	10, 13, 17, 93, 109, 171
ランダムサブドメイン攻撃	34
リークサイト	21, 93
リフレクション攻撃	34
リモートデスクトップ	14, 18, 20, 150
量子鍵配送 (QKD : Quantum Key Distribution)	129, 136
ロシア・ウクライナ戦争	34, 105, 107, 219, 232

著作・製作 独立行政法人情報処理推進機構（IPA）

編集責任 高柳 大輔 小山 明美 涌田 明夫 白石 歩 井上 佳春  
小川 隆一

執筆者 IPA  
浅見 侑太 板垣 寛二 伊藤 彰朗 伊東 麻子 伊藤 吉史  
井上 佳春 内海 百葉 大久保 直人 大友 更紗 小川 賢一  
小川 隆一 小幡 宗宏 甲斐 成樹 金山 栄一 金子 成徳  
神谷 健司 唐亀 侑久 河合 真吾 神田 雅透 黒岩 俊二  
小杉 聡志 小山 明美 小山 祐平 佐川 陽一 佐藤 栄城  
篠塚 耕一 白石 歩 白鳥 悦正 新保 淳 銭谷 謙吾  
高塚 光幸 竹内 智子 武智 洋 田島 威史 田島 凜  
丹野 菜美 近澤 武 辻 宏郷 長迫 智子 中島 健児  
楢原 龍史 西尾 秀一 西村 奏一 野村 春佳 橋本 徹  
長谷川 智香 平尾 謙次 福岡 尊 福原 聡 富士 愛恵里  
藤井 明宏 古居 敬大 松島 伸彰 宮本 冬美 森 淳子  
安田 進 山下 恵一 吉野 和博 吉原 正人 吉本 賢樹  
渡邊 祥樹

株式会社日立製作所 相羽 律子  
三菱電機株式会社 神余 浩夫  
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 セキュリティ危機管理チーム 満塩 尚史  
株式会社 KDDI 総合研究所 三宅 優  
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  
情報規格調査会 JTC 1 / SC 27 / WG 5 小委員会

協力者 IPA  
和泉 隆平 板橋 博之 伊藤 真一 江島 将和 大澤 淳  
釜谷 誠 亀山 友彦 岸野 照明 北村 弘 栗原 史泰  
桑名 利幸 古明地 正俊 塩田 英二 清水 碩人 瀬光 孝之  
高見 穰 高柳 大輔 田口 聡 田村 智和 土屋 正  
遠山 真 中島 尚樹 中野 美夏 西原 栄太郎 日向 英俊  
松田 修平 真鍋 史明 宮崎 卓行

一般社団法人 JPCERT コーディネーションセンター 石寺 桂子  
Trend Micro Incorporated 木村 仁美  
長崎県立大学 島 成佳  
国立研究開発法人情報通信研究機構 サイバーセキュリティ研究所  
経済産業省 商務情報政策局 サイバーセキュリティ課



## おわりに

ロシア・ウクライナ戦争の収束の兆しが見えないところに、イスラエル・ハマス間の武力衝突が勃発した2023年。戦場での戦闘とサイバー戦に加え、生成AIの進化や台頭によって精巧に加工された虚偽情報を用いた情報戦が繰り返されているといいます。一方、私達の身の回りにも本物の画像を細工したフェイクニュースや詐欺目的と思われる虚偽情報がSNS等で数多く飛び交っています。本白書では新たに設けた「第4章 注目のトピック」に、前年に引き続き、虚偽情報拡散に関する節を設け、多くの事例について解説しています。これに加え、AIのセキュリティについても第4章に節を設けました。IPAには2024年2月、AIを安全に利用し、利便性を享受できるよう、AIの安全性に関する評価手法や基準の検討等を行うAIセーフティ・インスティテュート(AISI)が設置されました。今後、本白書においてもAIに関する記述は欠かせないものになりそうです。

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。なお、出典元がIPA 以外の場合、当該出典元の許諾が必要となる場合があります。
- ・本白書は2023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、<sup>TM</sup>または<sup>®</sup>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 2024

変革の波にひそむ脅威：リスクを見直し対策を

2024年7月30日 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](mailto: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