S3 セキュリティー・トラスト

S3.04 人・社会とセキュリティー

キーワード：ユーザブルセキュリティー、コグニティブセキュリティー、ダークパターン、デセプティブパターン、誤情報、偽情報、ファクトチェック、ユーザー調査、サプライチェーンのセキュリティー、ソフトウェアの透明性、ソフトウェア部品表、サイバーセキュリティー研究倫理

本領域の論文・特許のデータについては以下をご参照ください
・ダッシュボード版：https://www.jst.go.jp/crds/report/CRDS-FY2025-FR-02.html
・PDF 版：https://www.jst.go.jp/crds/pdf/2025/FR/CRDS-FY2025-FR-03/CRDS-FY2025-FR-03_41S304.pdf

1．研究開発領域の定義

情報サービスを利用するユーザーの観点からセキュリティーの問題を解決し、社会に受容されて人々に活用され、社会を守るセキュリティー技術の研究開発を行う領域である。セキュリティーの観点において、人の脆弱性を狙った攻撃の解決や、情報サービスが実現できることと情報サービスを利用する人が期待することのギャップ（Socio-technical gap）の解決、多数の人・組織の関わりにおいてセキュリティー技術の普及を阻害する要因の解決、法制度などの社会的要請に応えられるセキュリティー技術の設計、情報が社会に拡散することによる影響の分析・対策などが盛んに研究開発されている。

2．本領域の意義

インターネットの発展に伴い、多種多様な活動がデジタル化され、情報システムの複雑な相互接続が急速に進む中で、セキュリティー技術の対象は個人、組織、産業などに拡大している。これまで、サイバー攻撃の検知・対策や脆弱性のないセキュアなシステム・サービスを実現するセキュリティー技術が研究されてきた。また、プライバシーでは、EU一般データ保護規則（General Data Protection Regulation：GDPR）などのプライバシー保護に関する法規制が世界的に進み、情報サービスを利用するユーザーに法制度との関係を提示する必要性や、法規制に適合したシステムを開発することが求められている。近年では、情報サービスの進展により、フィッシング詐欺や偽情報（Disinformation）・誤情報（Misinformation）[1]などが社会的な問題となっている。

情報サービスを利用するのは人であり、情報システムを開発・運用するのも人である。上述した問題を解決するためには、情報サービスを利用するユーザーや開発者・運用者などの人の思考や行動、ユーザーや開発者、運用者などが情報サービスを利用したり、開発・運用したりする際の社会的・組織的なプロセスやルールにも着目したセキュリティー技術の確立が必要である。

人・社会とセキュリティーは、個人、組織、産業、社会、国家にまで関わり、人々が安心・安全に利用できる情報サービスや社会の実現に貢献するために不可欠である。人・社会に着目するこの領域の研究開発は、これまでのセキュリティー分野の情報技術に加え、心理学、認知科学、経済学などの人文・社会科学を含めた学際的アプローチにより総合的に取り組む必要がある。

1　偽情報、誤情報とは、悪意（Intent to harm）と虚偽（Falseness）により分類すると、偽情報は悪意がありかつ虚偽の情報、誤情報は悪意がないかつ虚偽の情報である。それ以外に、悪意があるかつ虚偽の情報をマルインフォメーション（Malinformation）と呼ぶ［Claire Wardle, “Understanding Information disorder,” FIRST DRAFT, https://firstdraftnews.org/long-form-article/understanding-information-disorder/（2025年9月1日アクセス）］。
例えば、偽情報は「意図的／意識的に作られたウソ、虚偽の情報」、誤情報は「勘違い／誤解により拡散した間違い情報」などである［総務省「インターネットとの向き合い方～ニセ・誤情報にだまされないために～第2版」https://www.soumu.go.jp/use_the_internet_wisely/special/nisegojouhou/（2025年9月1日アクセス）］。

3．研究開発の動向

3.1　概要

人・社会とセキュリティーに関連する「ユーザブルセキュリティー」、「コグニティブセキュリティー」、「ソフトウェアサプライチェーンのセキュリティー」、「法制度とセキュリティー・プライバシー」、「サイバーセキュリティー研究倫理」について概要を述べる。

- **ユーザブルセキュリティー**

ユーザブルセキュリティーは、情報サービスを利用するユーザーや開発者・運用者などの人を中心にしてセキュリティーやプライバシーの問題解決に取り組む分野であり、1990年代後半頃に提唱され、現在も継続的に研究が進められている。この分野では、ユーザーが情報サービスを利用したり、開発・運用したりするさまざまなシーンにおいて、認識（perception）・行動（behavior）を観測してユーザーを理解することで、ユーザーがより適切な認識や意思決定（decision-making）を行いやすいセキュリティー・プライバシー技術の確立を目指している。

ユーザブルセキュリティーの研究では、ユーザーのより深い理解を試みる社会科学・心理学のアプローチを取り入れた観測が行われる。ユーザーがセキュリティー技術を十分に活用できていない状況や技術の誤解（misconception）が発生する理由や、フィッシングなどによってだまされる要因などを解明することで、専門性が必ずしも高くないユーザーにも十分な恩恵が得られるセキュリティー技術が研究されている。

開発者や運用者の観点では、開発者が脆弱なシステムを作ってしまう原因の理解と、その原因を解消できるように開発者をサポートする技術の創出、セキュリティー技術者が解析作業を行う際の明文化されていないノウハウ（暗黙知）を形式知化する試み、組織のセキュリティー運用でシステム管理者が抱える困難さと解決策の模索なども研究対象となっている。

また、法規制とセキュリティー・プライバシーについても研究対象となっている[2]。プライバシーに関する法規制と実際のシステム・サービスが提供する機能とのギャップや、ユーザーが抱いているプライバシーに関する認識と法規制として定められている内容とのギャップ、法規制と開発者の認識とのギャップを解消することで、法規制への適合とユーザーや開発者・運用者の利便性とを両立する研究が行われている。

- **コグニティブセキュリティー**

コグニティブセキュリティー[1), 2)]は、「人間の認知（Cognition）」に注目することで、人を中心とするセキュリティー・プライバシーに関わる問題の本質を理解し、解決を図ろうとする研究分野であり、2019年頃から提唱されている。ここでの認知とは、人が情報をどのように知覚・解釈・判断し、行動に移すかという一連の認知的プロセスを指す。コグニティブセキュリティーは、ユーザーがセキュリティー上のリスクをどのように捉えて意思決定するのか、また攻撃者が認知の脆弱性（例えば、ユーザーの注意の偏り、直感的な判断の偏り）をどのように悪用するのかを明らかにし、認知的プロセスに基づいて安全な行動や自律的な意思決定を促す技術や制度を確立することを目指している。

フィッシングなどのサイバー詐欺やソーシャルエンジニアリング、フェイクニュースなどの偽・誤情報、ダークパターン[3]などの操作的デザインはユーザーの認知に対する典型的な脅威であり、コグニティブセキュリティーの研究対象となっている。さらに、コグニティブセキュリティーの対象は個人にとどまらず、組織や国家にまで影響を及ぼす問題も対象としている。近年では、偽・誤情報の拡散による組織・国家の意思形成・合

2　法規制とセキュリティー・プライバシーについては、国立研究開発法人科学技術振興機構研究開発戦略センター『研究開発の俯瞰報告書　システム・情報科学技術分野（2024年）』[3)]の2.4.4 人・社会とセキュリティー (3) 研究開発領域の概要　［研究開発の動向］❷セキュリティー・プライバシーに関する法規制とシステムデザインも参照いただきたい。

3　「ダークパターン（dark patterns）」は、「デセプティブパターン（deceptive patterns）」と呼ばれる場合もある。

意形成のかく乱、世論操作などが深刻な問題として認識されている。これらの問題に対処するためには、認知科学、心理学、経済学などによる科学的なアプローチが求められている。コグニティブセキュリティーは人間の認知とその限界を科学的に理解し、それを踏まえた対策技術の実現や、教育プログラム、制度策定の基盤となるものである。

- **サイバーセキュリティー研究倫理**

サイバーセキュリティーの研究は、その研究行為や研究成果が社会やそれを利用する人々に対して直接的な影響を与える可能性がある。内容によっては悪い影響も含むこともあり、倫理的な配慮が必要である。特に、十分に前例のない研究対象や研究手法を取り扱う際には倫理的な問題にも直面しやすい。研究者の行動規範として研究者の間で認識され順守されるものが研究倫理であり、これにより研究が社会から信頼を獲得し、社会的要請（社会に対する貢献）に応えることができる。特にサイバーセキュリティーに関する研究倫理では、社会との関わりの中での研究行為・成果の説明責任や法令・関連規則などの順守が重要視されている。この考え方は、2012年に米国の国土安全保障省からMenlo Report[4)] と呼ばれるセキュリティーを含むICT研究全般の研究倫理原則が発行されて以降、欧米の国際学会を中心に認識が広まった。

日本国内では、2016年頃から各研究コミュニティーにおいてサイバーセキュリティーの研究倫理に関する認識状況の確認や課題の整理、アクションプランの考案などを行う動きが広まった。2018年には、日本学術振興会のサイバーセキュリティ第192委員会にワーキンググループが設置され、以後研究分野全体を横断するかたちで啓発活動が推進されている。これにより、研究倫理上配慮すべき具体的項目の整理や、個別の研究事例についての妥当性を相談する場の開設が行われるようになった。例えば、サイバーセキュリティー分野で国内最大級の学術シンポジウムである「コンピュータセキュリティシンポジウム（Computer Security Symposium：CSS）」では、2018年に研究倫理相談窓口が開設され、2019年には「サイバーセキュリティ研究における倫理的配慮のためのチェックリスト」が整備された。研究論文の投稿者は同リストの活用により倫理的配慮のセルフチェックが可能になっている。サイバーセキュリティーの研究倫理としては、ヒューマンファクターの問題、および脆弱性発見したときの責任ある情報開示（Responsible disclosure）が頻繁に議論の対象になっている。前者では、研究行為においてユーザーのプライバシーや心身が適切に保護されることが推奨され、後者では脆弱性を発見した際、適切に関係各所に情報が開示され、論文発表までに適切な対策が採られることが推奨される。

- **ソフトウェアサプライチェーンのセキュリティー**

サプライチェーンは複雑化・グローバル化しており、今日ではプロダクト開発が単一の組織で完結することはまれである。プロダクト開発では、さまざまな人々や組織によって作られた構成要素（ソフトウェアやハードウェアのコンポーネント）が利用されている。サプライチェーンのセキュリティーリスクは、この構成要素（ソフトウェアとハードウェアのコンポーネント）の脆弱性に起因している場合が多い。

プロダクト開発では、オープンソースソフトウェア（Open Source Software：OSS）が利用されており、効率的かつ迅速な開発のために大きな役割を果たしている。一方で、OSSに脆弱性が発見された場合、この脆弱性は、サプライチェーンを介して容易かつ広範囲に波及し、そのOSSを利用するすべてのプロダクトに影響を及ぼす。 OSSの脆弱性に迅速に対応するためには、プロダクトで利用されているソフトウェアコンポーネントを管理したソフトウェア部品表（Software Bill of Materials：SBOM）を準備しておくことが有効であり各国で注目されている。

米国では、2021年5月に「サイバーセキュリティー強化のための大統領令」が発され、ソフトウェアサプライチェーンに関するセキュリティーの向上に資するガイドラインの策定やSBOMの必須化などが米国連邦政府の公式な取り組みとして定められた。日本でもSBOMに関する検討が進められている。経済産業省では「サイバー・フィジカル・セキュリティー確保に向けたソフトウェア管理手法等検討タスクフォース」でSBOMの

作成・共有・活用などに関する議論、およびそれらの実証に向けた検討が進められている。2023年7月には「ソフトウェア管理に向けたSBOM（Software Bill of Materials）の導入に関する手引」が策定されている。情報処理推進機構（Information-technology Promotion Agency：IPA）は、企業向けに具体的な導入手順や運用モデルが示した「SBOM導入・運用の手引き」（2024年12月）を公開している。

SBOMはソフトウェアの透明性（Software Component Transparency）を実現する有効な手法の一つであり、ソフトウェアの開発者と利用者の両方にとってメリットがある。ソフトウェア開発者は、開発しているソフトウェアの脆弱性を管理でき、ソフトウェアの使用者はソフトウェアの脆弱性対策の情報として利用できる。SBOMの導入・活用は、サプライチェーンのリスクを根元から低減する効果が想定されており、その活用が期待されている。

- **重要インフラやサプライチェーンのレジリエンス**

重要インフラやサプライチェーンに関わる攻撃により被害の拡大を受けて、重要インフラやサプライチェーンのレジリエンスが重要視されている。各国政府から各種指針が公開されており、2014年に米国の国立標準技術研究所（National Institute of Standards and Technology：NIST）から「重要インフラのサイバーセキュリティーを改善するためのCyber Security Framework（CSF）Version1.0が発行された。2024年2月には約10年ぶりにCSF Version 2.0に改訂され、業種や企業規模などに依存しない汎用的・体系的なガイドラインとなっている。日本では、内閣サイバーセキュリティセンター（National center of Incident readiness and Strategy for Cybersecurity：NISC）から各種の行動計画や安全基準などの策定指針が公表されている[4]。また、経済産業省からは、「サイバー・フィジカル・セキュリティ対策フレームワークVer1.0」（2019年4月）が公表され、その展開としてビルや工場、スマートホームなどを対象としたガイドラインが公表されている。

3.2 トピックス

人・社会とセキュリティーに関するトピックスとして、「偽・誤情報の対策のためのツールボックス」、「集合知を活用した偽・誤情報対策」、「ダークパターンの高度化・複雑化」、「サイバーセキュリティーの研究倫理の強化」、「研究プロジェクトと法制度の動向」について述べる。

- **偽・誤情報の対策のためのツールボックス**

偽・誤情報の対策では、「nature human behaviour」に掲載された論文“Toolbox of individual-level interventions against online misinformation”（2024年5月）[5)] で、個人レベルでの介入研究が包括的にレビューされている。この論文は、オンライン上の偽・誤情報に対抗する行動的・認知的介入手法を、最新の研究を含む81本の科学論文をエビデンスとして包括的に整理したものである。この論文の特徴は、偽・誤情報に関わる問題のさまざまな側面に対処するために、オンラインツールボックスがウェブサイトで公開されている点であり、研究者だけでなく、政策立案者、教育関係者、一般の人々もアクセスできる。

オンラインツールボックスは、概念的ツールボックス（Conceptual toolbox）とエビデンス的ツールボックス（Evidence toolbox）の2種類に分かれており、概念的ツールボックスでは介入手法が整理されている。介入手法は、3つのカテゴリーとして、行動を対象とする「ナッジ（Nudge）」、能力（Competence）を対象とする「ブーストと教育的介入（Boost and educational interventions）」、人の信じる内容といった

4 「重要インフラの情報セキュリティ対策に係る第4次行動計画」（2020年1月改訂）、「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（第5版）」（2019年5月改訂）、「重要インフラのサイバーセキュリティに係る行動計画」（2024年3月改定）、「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（2023年7月）」

信念（Beliefs）を対象とする「反駁方略（Refutation strategies）」に分けられている。エビデンスのツールボックスは、世界中で行われた81本の科学的介入研究を参照できるデータベースとなっており、各研究の介入手法や対象国、対象者の数、主要な結果などを参照することができる。対象国ごとの研究数が世界地図に表示されており、北米、欧州を対象とした研究が多く行われていることが見て取れる。近年は南米やアフリカなどの国を対象とした研究も行われているが、日本を対象とした研究は少ない。

- **集合知を活用した偽・誤情報対策**

ファクトチェック（Fact Checking）とは、情報の真偽を検証する行為であり、偽・誤情報の対策手段の一つとしてプラットフォーム事業者やメディア関連団体、ファクトチェック推進団体、行政機関などによって実施されている。その重要性が注目されている一方、ファクトチェックにはチェックを行う主体の中立性、判断材料の不完全性、チェック体制のリソース不足などの問題がある。このような中で、「コミュニティーノート（community notes）」と呼ばれる手法が導入されはじめ、偽・誤情報への新たな対応策として注目されている。コミュニティーノートとは、SNSの利用者（一般ユーザー）が投稿に対して補足情報や異なる視点による情報を自発的に記述し、他のユーザーがその投稿の信頼性や妥当性を評価することで、集合知的にチェックする仕組みである。コミュニティーノートの課題も指摘されている。特に、偽・誤情報が拡散する初期段階においては、コミュニティーノートの対応が後手に回る傾向があり、結果として拡散の抑制には十分な効果を発揮できていないという問題が報告されている[6)]。そのため、誤解を招きやすい投稿をアルゴリズムにより優先的に抽出し、優先度の高い投稿をコミュニティーノートの貢献者に速やか提示するなど、コミュニティーノートのタイミングギャップを埋める新たな手法の研究開発が必要である。

- **ダークパターンの高度化・複雑化**

ダークパターン（Dark Pattern）とは、ユーザーが無意識のうちに自身に不利な行動を取るように誘導するデザインである。米国連邦取引委員会（Federal Trade Commission：FTC）、および国際消費者保護ネットワーク（International Consumer Protection and Enforcement Network：ICPEN）が2024年に公表した報告書では、サブスクリプションサービスを提供するウェブサイトやアプリのうち、76%が少なくとも1種類の“疑わしい”ダークパターンを含んでいたことが示されている[7)]。このような報告もあり、規制当局や研究者の間で、ダークパターンに対するプラットフォーム設計の透明性と説明責任、そしてユーザー保護の必要性が改めて強く認識されることとなった。

米国や欧州では、ダークパターンに対して法的な規制が始まっている。研究では、一部のダークパターンについての実態調査と問題指摘が行われている。例えば、ウェブサイトにおけるオプトアウトの提供や、サブスクリプションを解除する設定の難しさが調査されている。また、ユーザーが設定しやすいウェブサイト設計の推奨例が提案されている。ダークパターンの手法は高度化・複雑化しており、ユーザーの認知バイアスを意図的に操作するような設計が数多く見られる。近年、「ダーケストパターン（Darkest Pattern）」と呼ばれる新たな概念が提唱されている[8)]。ダークパターンとダーケストパターンの違いは、操作がどこに埋め込まれているかという点にある。ダークパターンでは、操作は主にユーザーインターフェース（User Interface：UI）に埋め込まれているのに対し、ダーケストパターンでは、その操作はシステムのアーキテクチャーに埋め込まれている。つまり、利用者である一般ユーザーがダーケストパターンを識別することは難しく、コグニティブセキュリティーの観点から高度な情報技術と心理学的理解の両方に基づく対策が必要となる。

- **サイバーセキュリティーの研究倫理の強化**

研究倫理への関心の高まりを受けて、サイバーセキュリティーの国際学会においても倫理的配慮の明文化（Call for Paperへの記載など）が進んでいる。特に、セキュリティー・プラバシーに関する4大トップ国際学会の一つであるUSENIX Security Symposiumでは、2025年に倫理的配慮に関するガイドラインが制定

され、研究者が倫理的問題に向き合い、論文の中でどのように説明責任を果たすべきかが明確に示された。さらに、USENIX Security Symposiumでは倫理的問題の有無にかかわらず、すべての投稿論文で研究倫理に関する記述が必須となっている。論文著者は、研究に倫理的懸念が存在したかどうか、存在した場合にどのように対応したのか、存在しないと判断した理由も含めて明示することが求められている。

このように、研究倫理は、論文投稿時の形式的な要件にとどまらず、研究の信頼性、透明性、社会的な説明責任を支える重要な要素となっている。

- **研究プロジェクトと法制度の動向**

①米国・IARPA（Intelligence Advanced Research Projects Activity）ReSCINDプロジェクト[5]

ReSCIND（Reimagining Security with Cyberpsychology-Informed Network Defenses）プロジェクトは、攻撃者の手口や認知に着目した研究プロジェクトである。攻撃者が持つ認知バイアスや意思決定の癖に注目し、それらを逆手に取った防御手法の開発を目的としている。このプロジェクトでは、攻撃者の判断に影響を与える心理的要因を特定し、それに基づいて「サイバー心理学的防御」を構築する。将来的には、こうした防御を自動的に適応させるシステムによって、攻撃者の行動を誘導・無力化することを目指している。

②米国・MITRE Engage[6]

MITRE社が提供するフレームワーク「Engage」は、攻撃者との認知的なかけひきに基づく新しい防御戦略を体系化したものである。攻撃者に対して偽情報を提示することで誤判断を促し、組織への実害を防ぐ「アドバーサリー・エンゲージメント（Adversary Engagement）」の実践を支援するものである。 MITRE Engageは、MITRE ATT&CKとの連携により、攻撃者の行動パターンに対して各フェーズ（Prepare、Expose、Affect、Elicit、Understand）に基づいた防御計画の立案と実行を可能にするものである。このように、コグニティブセキュリティーの分野では、人間の心理や行動に関する知見を取り入れた攻防の研究が進んでおり、従来の技術的対策を認知面から補完・強化するアプローチとして期待されている。

③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」（2023年11月～）

AIの誤用・悪用によってプラットフォーム上で偽・誤情報がまん延しつつある現状や、こうした問題に対する国際的な関心の高まりを背景に、デジタル空間における情報流通の健全性を確保するための総合的な検討が進められている。検討会では、プラットフォーム事業者に対するヒアリングも実施されている。議論では、「表現の自由」や「知る権利」の実質的な保障、民主主義の実現、安全で信頼できる情報流通空間としてのデジタル空間の構築といった基本理念のもと、国内外のマルチステークホルダーが連携・協力しながら、継続的かつ安定的に対応を進めることの重要性が強調されている。特に、「インターネット上の偽・誤情報対策に係るマルチステークホルダーによる取組集」には関係者が相互に参照しやすい形で取り組み事例が整理・公開されており、今後の国内外における連携や協働の一助となることが期待される。

3.3 課題

人・社会とセキュリティーに関する課題として、「コグニティブセキュリティーのための認知特性の解明と対策」、「コグニティブセキュリティー特有の研究倫理」、「多様なユーザーを対象とした調査」、「ユーザー募集のための環境整備」、「ユーザー調査手法の確立」、「大規模言語モデルを用いたユーザー調査の可能性」、「OSS開発のセキュリティー研究」について述べる。

5 IARPA, “RESCIND: REIMAGINING SECURITY WITH CYBERPSYCHOLOGY-INFORMED NETWORK DEFENSES,” https://www.iarpa.gov/research-programs/rescind（2025年9月1日アクセス）

6 “MITRE | Engage”, https://engage.mitre.org/（2025年9月1日アクセス）

- **コグニティブセキュリティーのための認知特性の解明と対策[2)]**

コグニティブセキュリティーは、人間の認知に注目して、人を中心とするセキュリティー・プライバシーに関わる問題の本質を理解し解決を図ろうとする研究分野である。

人間の認知研究は、過去から多くの研究が行われ、既に、認知バイアスや二重過程理論、記憶メカニズムなど、さまざまな知見が蓄積されている。しかし、AIやSNSなどの情報技術が普及した現在のデジタル社会で発生しているフィッシングや偽・誤情報、ダークパターン、世論操作などの問題に対する人間の認知についての知見は十分に検証されておらず対策への活用も進んでいない。また、生成AIは画像や文書の生成だけでなく音声や映像による自然な会話も可能となりつつあり、SNS上ではアテンションエコノミー[7]と呼ばれる行動が広がり、真偽に関係なく注目されることを意図した情報が氾濫している。今後、生成AIの進展によりだます手段が高度化・複雑化し、SNSでは真偽が判断できない情報が増加し続けることが予想される。問題が発生してから後追いで対策するだけではなく、将来起こる可能性がある問題にもプロアクティブに対処することも必要である。今後、デジタル社会で起こっている問題に対する人間の認知特性の理解と、それに基づく効果的な対策技術を創出することが求められる。

- **コグニティブセキュリティー特有の研究倫理**

人間の認知への攻撃を対象とした研究では、調査対象ユーザーへの侵襲性や不可逆性といった倫理的リスクが伴う可能性がある。日常的な意識や行動の把握などを目的とした一般的なユーザー調査と比較すると、認知への攻撃を扱う調査は心理的な侵襲性が高く、参加者の認知に持続的あるいは不可逆的な影響を及ぼす恐れがある。例えば、陰謀論をテーマとしたユーザー調査の後に、参加したユーザーの一部がその内容を調査後も信じ続けてしまう事例が報告されている[10)]。このようなリスクを回避して調査を行うためには、認知への悪影響を最小化しつつ信頼性ある結果を得る方法論を確立し、倫理的に許容される調査設計の議論が不可欠である。

また、認知に対する特定の脆弱性（例えば、特定の情報や状況下で人が誤った判断をしやすくなるような特性）を研究によって発見した場合、その情報をどのように開示すべきかという課題もある。サイバーセキュリティー分野では、発見した脆弱性を関係者に適切に通知し、論文発表の前に対策が講じられるようにする「責任ある情報開示（responsible disclosure）」が推奨されている。情報システムではセキュリティーパッチなどで比較的短期間で修正ができることを前提としているが、認知の脆弱性は、セキュリティーパッチなどにより容易に修正できるものではなく、技術的な情報開示とは本質的に異なる倫理的課題が存在する。一方、現実的には、こうした認知の脆弱性がいずれ攻撃者によって発見・悪用される可能性も考えられるため、研究者が過度に情報を秘匿してしまうことは、社会にとって有効な対策を検討する機会を失うリスクにつながる。現時点では、研究コミュニティーで、認知の脆弱性に関する知見をどのように扱い社会的に責任あるかたちで開示すべきかについての倫理的スタンスは十分に議論されていない。今後、この分野における倫理的な合意形成やガイドライン作成に向けた検討を積極的に進めることが求められている。

- **多様なユーザーを対象とした調査**

近年、調査対象の偏りが、この研究分野の重要な課題として指摘されている。これは過去の研究成果の一般化可能性（Generalizability）に疑問を投げかけるものである。例えば、WEIRD（Western：西洋の、Educated：教育水準の高い、Industrialized：工業化された、Rich：裕福な、Democratic：民主主義の）社会の人々に調査が大きく偏っていることや[11)]、人口統計学的要因（Sociodemographic Factors）と調査結果との関係性の分析が不十分であることが問題として指摘されている[12)]。

7　人々が情報を消費する際に支払える「注目」や「関心」が経済的価値を持つようになった経済モデル[9)]。

トピックス「偽・誤情報対策のためのツールボックス」で述べたように、偽・誤情報対策の研究では、多くの研究が実施され、一定の条件下でそれぞれの有効性が示されている。しかし、研究の対象は欧米に集中しており、日本を対象とした研究は少ない。欧米を対象とした結果が他の地域でも有効なのか、文化的・社会的背景による効果の違いがあるかについては十分に示されていない。また、近年では、「at-riskユーザー」[8]と呼ばれる人々に対する研究が盛んに実施されている。これは、従来のユーザブルセキュリティーの研究の対象が暗黙的に欧米在住者・健常者・十分な教育を受けている人々に偏っていたことを意味している。

国や地域、性別・年齢などの属性が異なるさまざまなユーザーを対象として、適切なセキュリティー・プライバシー技術を設計するには、多様なユーザー群を対象とした調査・分析を行い、異なる文化的・社会的背景や属性を考慮した技術や制度を提案することが不可欠である。

- **ユーザー募集のための環境整備**

本領域の研究では、ラボ実験、アンケート、インタビューなどのユーザー調査によって課題の発見や開発した技術の有効性評価が実施されている。調査対象のユーザーは、クラウドソーシングサービスを利用して募集されるが、クラウドソーシングサービスに登録しているユーザー層と調査対象として想定しているユーザー層との間に乖離が発生しやすい。登録者の属性分布やエンゲージメントの程度などの特性は、クラウドソーシングサービスによって大きく異なることが、近年、複数の研究で明らかにされている[13)-15)]。これは、同じ調査であっても、ユーザー募集に使用するクラウドソーシングサービスの選択自体が研究結果に影響を及ぼし得ることを示唆している。

欧米には研究目的が明示され、研究目的に賛同するユーザーが多く登録されているクラウドソーシングサービスが存在するが、日本にはそのようなサービスがないためフリーランサーが登録するサービスなどが利用されている。また、近年の傾向であるエキスパートユーザーやAt-riskユーザーを対象にした研究では、適切なユーザーを集めること自体が容易ではないため、研究者の身近にいる人々を調査対象のユーザーとする便宜的標本抽出法（Convenience Sampling）が用いられるが、母集団の代表性を確保することが難しい。今後、本領域の研究を進めるためには、日本でも調査対象となるユーザーを安定的に確保できる環境を整備することが必要である。

- **ユーザー調査手法の確立**

ユーザー調査では、セキュリティーやプライバシーに関する調査を実施する際、調査対象のユーザーにさまざまなバイアスがかかりやすい状況があることが知られている。例えば、サーベイやインタビューでは、本来とは異なる社会的に望ましいとされる回答をする「社会的望ましさバイアス（Social Desirability Bias）」や、質問の言い回し（Wording）によるバイアスが発生しやすい。また、ラボ実験では、自身が観察されていることを意識したり、研究者の期待に応えようとして普段と異なる行動を取る「観察者効果（ホーソン効果：Hawthorne Effect）」が発生したり、普段とは異なる人工的な環境で作業をすることで本来の自然な行動が観測できず、「生態学的妥当性（Ecological Validity）」が低い事象が発生したりする。例えば、フィッシングメールに対する反応を調査する場合、調査前にユーザーにフィッシングメール判別であることを説明しすぎると、普段以上に警戒した行動を取りやすくなり、普段の環境での反応を観測することが難しくなる。このようなバイアスが少ない調査手法の設計や、生態学的妥当性を向上するための調査手法の確立が求められる。

また、フィッシングなどのサイバー詐欺に関する調査は、テキスト情報など単一のモダリティに基づく、比較的単純なワンステップの攻撃シナリオを対象とするものが多い。そのため、実際のデジタル空間において

8 「at-riskユーザー」とは、セキュリティー・プライバシー（さらにはセーフティー）に関する被害に遭いやすい人々をさす。例えば、情報リテラシーが低いユーザーや被害によるリスクが大きいユーザーなどがあげられる。

ユーザーが直面するような複数の情報源が絡み合う状況や多段階にわたる詐欺攻撃（例えば、SNSやメッセージアプリ、Webブラウザーを横断する連携的な誘導）について、ユーザーの自然な認知・行動を踏まえた調査設計や調査手法を確立することが必要である[16)]。

- **大規模言語モデルを用いたユーザー調査の可能性**

近年、大規模言語モデル（Large Language Model：LLM）の技術的な進展により、従来、手間やコストがかかったり、ユーザーへの身体的・心理的な負担を伴ったりしていたユーザー調査で、人間（ユーザー）をLLMで代替する調査手法が増加しつつある。この調査手法では、特定の属性や状況を設定したLLMに対して質問やタスクを提示し、その反応を分析するものであり、「シリコンサンプリング（Silicon Sampling）」とも呼ばれている[17)]。例えば、誤情報へのだまされやすさを評価する研究では、年齢・教育レベル・政治的立場などの属性を設定したLLMを「仮想的な参加者」として、さまざまな質問やタスクを提示して反応を分析する試みが行われている[18)]。このようなアプローチは、調査の倫理的・実務的な制約を緩和し、人間では実施することが難しいセンシティブな調査項目を含む研究や仮説的なシナリオでの調査を実施できる可能がある。一方で、LLMを人間の代替として調査に活用することには慎重な姿勢も求められる。例えば、LLMの反応が「平均化された人間の意見」なのか、「特定の文化的・社会的バイアスを反映したステレオタイプ」なのか、「実際の人間の意思決定」なのかなど、LLMの反応が何を意味しているのかについて、さらなる実証的証拠の蓄積と理論的枠組みが必要であることが指摘されている[19)]。また、質的研究にLLMを活用することには批判的な視点がある。質的研究が本来重視する「共創的な文脈に根ざした意味づけのプロセス」は、モデルとしてのLLMには本質的に欠けているとされ、知識生産における主体性や方法論の正当性が損なわれる可能性があることが指摘されている[20)]。

- **OSS開発のセキュリティー研究**

近年のプロダクト開発においてOSSは基本的な機能を実現する上で欠かせない存在であるが、OSSを開発するプロジェクトに対する攻撃も確認されている。攻撃コードや脆弱性がOSSに混入しやすいことが明らかになっており、セキュアな開発を行うためには、開発段階におけるコードを悪意のある開発者から保護することが課題になっている。2020年から2021年にかけて、Linux Kernel開発コミュニティーに対して研究者が「善意の開発者」を装って脆弱なコードをコミットする実験を実施したことが大きな問題になった[21)]。この研究者はOSSにおけるコードレビュープロセスの問題点を抽出することを目的としていたが、開発コミュニティーには受け入れられず、開発コミュニティーは研究者によってコミットされたすべてのコードを再検証するために膨大な人的コストを負うことになった。

こうした「開発者への秘密裏あるいは暗黙的な介入」による調査研究の問題が広く認知されたことにより、近年の開発者に対する調査研究では、研究者と開発者の信頼関係や調査方法の透明性を重視した協働的・観察的アプローチが主流となっている。具体的には、開発者に対するオンラインアンケートやインタビューによる調査、あるいはGitHubなどのコードリポジトリに公開されているコードの分析が広く実施されている。しかし、アンケートやインタビューは、開発者の意図や文脈を深く理解できが、調査対象とする開発者の人数に限りがあるため大規模な調査には適さない。公開されているコードの分析は、広範な開発者を対象として調査することが可能であるが、開発者がなぜそのように判断・行動したのかという背景や認知プロセスを明らかにすることが難しいといった利点と限界がある。今後はこれらの調査手法を補完的に組み合わせた「混合研究法的アプローチ（Mixed-methods Approach）」によって、OSS開発とセキュリティーの包括的な調査と課題の抽出、さらには対策技術の設計・提案が期待されている。

3.4 各国・地域の取り組み

【日本】

（基礎研究）

・国内ではユーザブルセキュリティーの研究コミュニティーが2017年に立ち上がり、大学や企業の研究発表数も増加傾向にある。

・著名な国際学会で論文が採録されるようになってきている。SOUPS（2021：1件、2022：3件、2024：1件）、USENIX Security Symposium（2023：1件、2024：2件）。EuroUSEC 2021 Best Paper Award受賞（早稲田大学/NTT）。

・サイバーセキュリティー研究倫理について、国内学会でチェックリストの整備や相談窓口の設置などサポート体制の充実が確認できる。

（応用研究・開発）

・ユーザーの行動観測やユーザーに対する注意喚起などを実施するいくつかのプロジェクトが始動しており、今後の研究成果や社会実装が期待できる。例：Web媒介型攻撃対策技術の実用化に向けた研究開発（WarpDriveプロジェクト）（NICT）、am I infected?（横浜国立大学）、IoT機器を悪用したサイバー攻撃防止に向けた注意喚起の取り組み（NOTICE）（総務省およびNICT）

［評価＊　基礎研究：現状○／トレンド↗、応用研究・開発：現状△／トレンド↗］

【米国】

（基礎研究）

・米国はユーザブルセキュリティーの黎明期から研究分野をけん引・進展させてきた。

・中心的な研究グループが属するCarnegie Mellon University（CMU）Cylabや、そのOBとOGの多くが米国の各大学（メリーランド大学、シカゴ大学など）で研究チームを作り、本分野をけん引している。

（応用研究・開発）

・ユーザブルセキュリティーの研究成果はNISTのガイドライン（NIST SP800-63Bなど）などに取り入れられて、米国だけでなく、欧米や日本などでも広く参照されている。

・Carnegie Mellon University（CMU）Cylabの「Privacy Nutrition Labels」研究の知見に基づいてApple “Privacy Nutrition Labels”（2020）やGoogle “Data Safety Section”（2022）が導入された。

・SBOMの仕様策定や普及推進活動が活発に行われている。

［評価＊　基礎研究：現状◎／トレンド→、応用研究・開発：現状◎／トレンド→］

【欧州】

（基礎研究）

・GDPRなどの各種法制度を後押しに、ここ数年で多数の研究成果をあげている。

・ユーザブルセキュリティーに関して有力な研究グループが増加しており、英国に加えて、ドイツの複数の研究グループによる成果が顕著である。

（応用研究・開発）

・GDPRなどによるプライバシーの規制は、プライバシーポリシーやCookieなどインターネット上でのビジネス活動に大きな影響を与えている。EUに限らず、米国や日本などに対してもビジネス/ 法規制の面で大きな影響を与えている。

［評価＊　基礎研究：現状◎／トレンド↗、応用研究・開発：現状◎／トレンド↗］

【中国】

（基礎研究）

・著名国際学会の採択論文実績が増えてきている。USENIX Security Symposium 2024では4本採択。

（応用研究・開発）

・顕著な成果は見られない。

［評価＊　基礎研究：現状〇／トレンド↗、応用研究・開発：現状×／トレンド→］

【韓国】

（基礎研究）

・ユーザブルセキュリティーに関する国際学会での発表がいくつか確認できる。

（応用研究・開発）

・顕著な成果は見られない。

［評価＊　基礎研究：現状△／トレンド→、応用研究・開発：現状×／トレンド→］

＊基礎研究：大学・国研などでの基礎研究の範囲、応用研究・開発：技術開発（プロトタイプの開発含む）の範囲
現状（日本の現状を基準にした評価ではなく、CRDSの調査・見解による評価）：
◎ 特に顕著な活動・成果が見えている　〇 顕著な活動・成果が見えている、
△ 顕著な活動・成果が見えていない　× 特筆すべき活動・成果が見えていない
トレンド（ここ1～2年の研究開発水準の変化）：↗上昇傾向、→現状維持、↘下降傾向

参考文献

1）国立研究開発法人科学技術振興機構研究開発戦略センター『俯瞰ワークショップ報告書 コグニティブセキュリティー研究動向』（2024年3月）, https://www.jst.go.jp/crds/report/CRDS-FY2025-WR-04.html, （2025年11月25日アクセス）.

2）国立研究開発法人科学技術振興機構研究開発戦略センター『科学技術未来戦略ワークショップ報告書 コグニティブセキュリティー』（2025年9月）, https://www.jst.go.jp/crds/report/CRDS-FY2025-WR-04.html, （2025年11月25日アクセス）.

3）国立研究開発法人科学技術振興機構研究開発戦略センター『研究開発の俯瞰報告書 システム・情報科学技術分野（2024年）』（2024年9月）, https://www.jst.go.jp/crds/report/CRDS-FY2024-FR-03.html, （2025年11月25日アクセス）.

4）U.S. Department of Homeland Security, “The Menlo Report: Ethical Principles Guiding Information and Communication Technology Research, August 2012,” https://www.dhs.gov/sites/default/files/publications/CSD-MenloPrinciplesCORE-20120803_1.pdf, （2025年11月25日アクセス）.

5）Anastasia Kozyreva, et al., “Toolbox of individual-level interventions against online misinformation,” *Nature Human Behaviour* 8, （2024）: 1044-1052, https://doi.org/10.1038/s41562-024-01881-0.

6）Yuwei Chuai, et al., “Did the Roll-Out of Community Notes Reduce Engagement With Misinformation on X/Twitter?” *Proceedings of the ACM on Human-Computer Interaction* 8, no. CSCW2, 428 （2024）: 1-52, https://doi.org/10.1145/3686967.

7）Federal Trade Commission, “FTC, ICPEN, GPEN Announce Results of Review of Use of Dark Patterns Affecting Subscription Services, Privacy,” https://www.ftc.gov/news-events/news/press-releases/2024/07/ftc-icpen-gpen-announce-results-review-use-dark-patterns-

affecting-subscription-services-privacy,（2025年11月25日アクセス）.

8）Mark Leiser, “Psychological Patterns and Article 5 of the AI Act,” *Journal of AI Law and Regulation* 1, no. 1（2024）: 5-23, https://doi.org/10.21552/aire/2024/1/4.

9）Thomas H. Davenport and John C. Beck, “The Attention Economy,” *Ubiquity* 2001, 6,（2021）: 1-es, https://doi.org/10.1145/376625.376626.

10）Sander van der Linden, “The conspiracy effect: The truth is our there,” in *Foolproof: Why Misinformation Infects Our Minds and How to Build Immunity*（New York: W. W. Norton & Company, 2023）, 46-62.

11）Ayako A. Hasegawa, Daisuke Inoue and Mitsuaki Akiyama, “How WEIRD is Usable Privacy and Security Research?” *33rd USENIX Security Symposium（USENIX Security 2024）*. Philadelphia, PA, USA, 2024-08-14/16, USENIX Association, 2024, 3241-3258, https://www.usenix.org/conference/usenixsecurity24/presentation/hasegawa,（2025年11月25日アクセス）.

12）Miranda Wei, et al., “SoK（or SoLK?）: On the Quantitative Study of Sociodemographic Factors and Computer Security Behaviors,” *33rd USENIX Security Symposium（USENIX Security 2024）*. Philadelphia, PA, USA, 2024-08-14/16, USENIX Association, 2024, 7011-7030, https://www.usenix.org/conference/usenixsecurity24/presentation/wei-miranda-solk,（2025年11月25日アクセス）.

13）Derek A. Albert and Daniel Smilek, “Comparing attentional disengagement between Prolific and MTurk samples,” *Scientific Reports* 13, no. 1, 20574（2023）, https://doi.org/10.1038/s41598-023-46048-5.

14）Benjamin D. Douglas, Patrick J. Ewell and Markus Brauer, “Data quality in online human-subjects research: Comparisons between MTurk, Prolific, CloudResearch, Qualtrics, and SONA,” *PLoS One* 18, no. 3, e0279720（2023）, https://doi.org/10.1371/journal.pone.0279720.

15）Laura Lascau, et al., “Monotasking or Multitasking: Designing for Crowdworkers’ Preferences,” *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. Glasgow, UK, 2019-05-04/09, ACM, 2019, 419, 1-14, https://doi.org/10.1145/3290605.3300649.

16）Pavlo Burda, Luca Allodi and Nicola Zannone, “Cognition in Social Engineering Empirical Research: A Systematic Literature Review,” *ACM Transactions on Computer-Human Interaction* 31, no. 2, 19（2024）: 1-55, https://doi.org/10.1145/3635149.

17）Lisa P. Argyle, et al., “Out of One, Many: Using Language Models to Simulate Human Samples,” *Political Analysis* 31, no. 3（2023）: 337-351, https://doi.org/10.1017/pan.2023.2.

18）Weicheng Ma, et al., “Simulated Misinformation Susceptibility（SMISTS）: Enhancing Misinformation Research with Large Language Model Simulations,” *Findings of the Association for Computational Linguistics: ACL 2024*. Bangkok, Thailand, 2024-08-11/16, ACL, 2024, 2774-2788, https://doi.org/10.18653/v1/2024.findings-acl.162.

19）Monojit Choudhury, et al., “The promise and pitfalls of generative AI,” *Nature Reviews Psychology* 4, no. 2（2025）: 75-80, https://doi.org/10.1038/s44159-024-00402-0.

20）Shivani Kapania, et al., “Simulacrum of Stories: Examining Large Language Models as Qualitative Research Participants,” *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. Yokohama Japan, 2025-04-26/05-01, ACM, 2025, 489, 1-17, https://doi.org/10.1145/3706598.3713220.

21）Thorsten Holz and Alina Oprea, “IEEE S&P’21 Program Committee Statement Regarding The ‘Hypocrite Commits’ Paper,” *42nd IEEE Symposium on Security and Privacy（SP）*. 2021-05-

24/27, IEEE, 2021, 1-3, https://www.ieee-security.org/TC/SP2021/downloads/2021_PC_Statement.pdf,（2025年11月25日アクセス）.

謝辞

本報告書の作成にあたっては、学協会、大学、民間企業、公的機関など、さまざまな方にご協力を賜った。情報提供者の一覧を下記URLに示す。
https://doi.org/10.82643/crds-fr-s009

本書について、著作権法で認められる範囲でのご利用についてはJSTのポリシー（https://www.jst.go.jp/site_policy.html）に従い、以下のクレジットを明記してください。本書に関するお問い合わせはcrds@jst.go.jpまでご連絡ください。

Regarding the use of this report within the scope permitted by copyright law, please follow the policy outlined here（https://www.jst.go.jp/EN/copyright.html）and clearly indicate the following credit. For inquiries regarding this report, please contact crds@jst.go.jp.

国立研究開発法人科学技術振興機構　研究開発戦略センター, 研究開発の俯瞰報告書　システム・情報科学技術分野～領域別動向編～（2026年）, セキュリティー・トラスト，人 ・社会とセキュリティー, 令和8年1月, CRDS-FR-S304-202601.
https://doi.org/10.82643/crds-fr-s-st-hsas

Center for Research and Development Strategy, Japan Science and Technology Agency, Trends Report by Discipline, Systems and Information Science and Technology Field（2026）, Security and Trust, Human and Social aspect in Security, January 2026, CRDS-FR-S304-202601.
https://doi.org/10.82643/crds-fr-s-st-hsas