URL https://journal.ntt.co.jp/article/35347
DOI https://doi.org/10.60249/25081101

NTT社会情報研究所
上席特別研究員
秋山満昭 Mitsuaki Akiyama

セキュリティに関する人間中心・オフェンシブ・社会技術的観点から，認知にかかわる脅威に対抗する「コグニティブセキュリティ」の確立に向けて取り組む

最近，ランサムウェアによる情報流出，DDoS（Distributed Denial of Service）攻撃による交通機関混乱，生成AI（人工知能）悪用の不正プログラムによる偽契約・転売，SQLインジェクション攻撃や不正アクセスによるデータ改ざん，ディープフェイクを悪用した詐欺等，サイバー攻撃による被害が国内外で発生しています．こうした被害に遭遇しないために，サイバー攻撃そのものを食い止める手段，サイバー攻撃の影響・波及を極小化する技術，人間の心理や行動による防御といった多角的な側面からの対応が必要となります．「人間中心のセキュリティ」「オフェンシブセキュリティ」「社会技術的情報セキュリティ」の観点から研究に取り組んでいる，NTT社会情報研究所 秋山満昭上席特別研究員に，サイバーセキュリティに関する3つのテーマとその共通領域における新たな研究，自分の経験や知識だけを頼りに固執しようと思わず，疑うことやアップデートすること，相互理解，モチベーションの共有を図るという考え方を伺いました．

「人間中心のセキュリティ」「オフェンシブセキュリティ」「社会技術的情報セキュリティ」の3つのテーマを柱とし，テーマ横断的な研究にも取り組む

現在，手掛けていらっしゃる研究について教えていただけますでしょうか．

ユーザの行動・意思決定や対策に着目するセキュリティ研究「人間中心のセキュリティ（Human-Centered Security）」，攻撃手法や攻撃者の戦略に着目するセキュリティ研究「オフェンシブセキュリティ（Offensive Security）」，情報の生成・流通・操作・受容における社会・技術の相互作用に着目するセキュリティ研究「社会技術的情報セキュリティ（Socio-Technical Information Security）」の3つのテーマを柱として研究しています（**図1**）．

前回（2022年6月号）では，①Webリホスティングサービスの脅威発見，②開発者や開発プロジェクトに着目したセキュアなソフトウェア開発，③英語ノンネイティブのフィッシングメール対策，④ユーザスタディの正しい方法論，⑤ソーシャルメディアにおける偽・誤情報の拡散メカニズム分析について紹介しました．これらは，3つのテーマの中で個々に取り組んできた研究ですが，現在は，特に「人間中心のセキュリティ」，「社会技術的情報セキュリティ」として個々に取り組む研究のほか，「オフェンシブセキュリティ×人間中心のセキュリティ」「社会技術的情報セキュリティ×人間中心のセキュリティ」といったテーマ横断的な研究にも取り組んでいます．

具体的には，「人間中心のセキュリティ」においては以下の2つがあります．

⑴　セキュリティ・プライバシー分野におけるユーザ調査研究の地理的偏りを定量的に分析

人を中心とする研究分野では，ユーザ調査を通じて心理特性や行動特性を解き明かしますが，研究対象の人が西洋偏重であり，

対象が地理的に偏ったこれまでの研究では，その結果の全人類共通性，地理的な違いの有無，その相違点，などの観点の深い分析や洞察が十分ではなく，全体像は明らかになっていませんでした．

そこで，研究分野に対して既存論文を体系的かつ包括的に検索・評価・統合する，体系的文献調査手法に基づいて，人を中心とするセキュリティ・プライバシー研究論文715本を特定し，参加者の居住国・属性・募集方法・研究手法・研究トピックに関して，複数の分析者による評価者間信頼性を保証した方法で分析しました．

本調査により，この分野の期間中の非西洋の人々が対象になるユーザ調査標本数の偏りが大きい（西洋偏重傾向）ことが明らかになりました．一方で，HCI（Human Computer Interaction）分野における同様の調査（2016～2020年の発表論文が対象）では，非西洋の国の標本数の偏りが緩和される傾向にあることが知られており，セキュリティ・プライバシー分野における偏りのほうがより顕著な傾向にあることが明らかになりました．また，世界人口比率に基づいた各国の調査度合いを調査した結果，米国，英国，ドイツなどの西洋（Western）の国々は世界人口比率に対して過多に調査されていることが分かりました．一方で，日本を含むアジアおよび中東・アフリカ・南米などの非西洋の大部分の国々では世界人口比率に対して調査が不十分であることが分かりました（**図2**）．この結果は，当該研究分野における調査対象の偏りの是正に向けた課題提起であると同時に，今後の研究の方向性を示唆するものでもあります．

本調査で判明した西洋偏重を解消し，多様な人々を理解するためのユーザ調査研究方法として，非西洋の人々に対する複製研究の推進により，「研究結果の一般化可能性および地理や文化による人々の差異を明らかにする」こと，「ユーザ調査の対象となる人々の国で活用されているローカルのクラウドソーシングサービスを

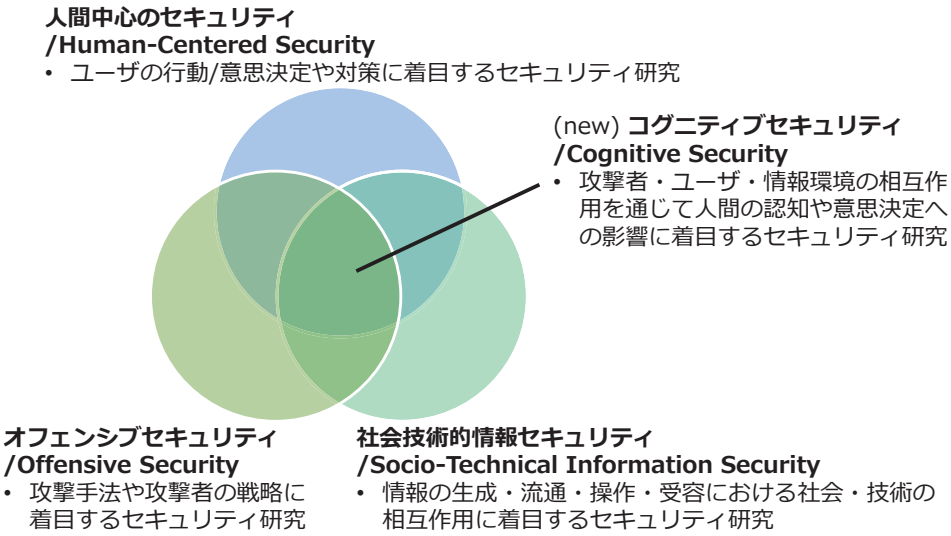

図1　研究テーマの分類

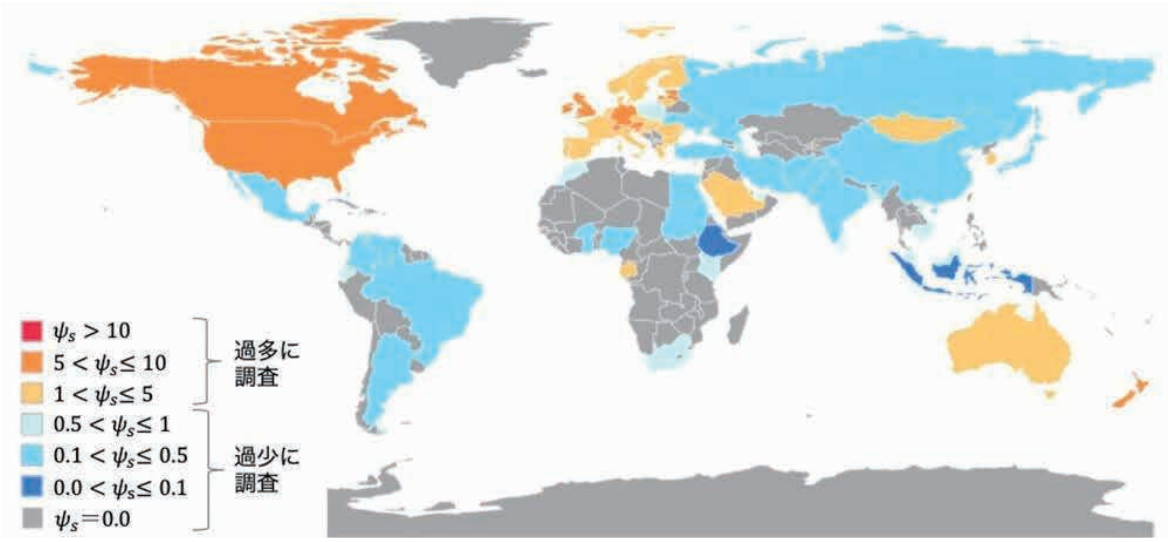

図2　世界人口に占める比率で正規化した各国の調査度合い（Ψs）

活用する，および，ローカルの言語・文化・環境を熟知した研究者との協業によって研究者のダイバーシティを向上させる」ことによる地理的・言語的障壁の克服を提案しました．

この調査・研究は国立研究開発法人情報通信研究機構（NICT）と共同で実施され，2024年 8 月14～16日に米国フィラデルフィアで開催されたサイバーセキュリティの最高峰国際会議の 1 つである「USENIX Security 2024」で発表しました[1]．

⑵ “セキュア開発ガイドライン”の日米における産業の開発現場における活用実態や運用上の課題の調査

安全なソフトウェアを開発することは，情報社会の根幹を支えるうえで不可欠であり，そのための設計や実装に関する方法や指針を記載した文書が“セキュア開発ガイドライン”とされています．既存の学術研究の多くは一般に公開されているパブリックなセキュア開発ガイドラインだけに着目している一方で，産業の開発現場における活用実態や運用上の課題はほとんど調査されてきませんでした．これを調査した結果，実際の産業の開発現場では，各社が独自に定めている自社製のセキュア開発ガイドラインのほうが広く利用されていることが明らかになりました．加えて，既存研究にて提唱されているセキュア開発ガイドラインの効果的な運用方法（例：ガイドラインに準拠して開発が行われたかの検査など）は，開発者の所属組織の性質によって実践可能性が大きく異なり，小規模な開発プロジェクトや受託開発では，開発者に与えられた裁量やコストの制約等により，特に実践が難しい場合があることも明らかにしました．

これらの調査結果は，セキュア開発ガイドラインに関する学術と産業とのギャップを鮮明にするものであり，安全なソフトウェアの開発を促進するうえでの課題の明確化に寄与しました．今後，調査結果を踏まえて安全なソフトウェア開発を推進することで，ユーザが安心して利用可能なセキュアなシステム・サービスの展開につながることが期待されます．

この成果は，HCI分野におけるトップ国際会議である，「The ACM CHI Conference on Human Factors in Computing Systems (CHI) 2023」（2023年 4 月23～28日）で採択，発表されました[2]．

「社会技術的情報セキュリティ」については，ソーシャルメディアにおけるスパムURLやフェイクニュース・偽情報の拡散などに悪用される偽アカウント（Sybil：悪性ボットの集団）攻撃対策技術の研究です．

従来，ソーシャルメディアにおけるスパムURLやフェイクニュース・偽情報をSybilから送信する攻撃に対して，アカウントの活動パターンなどから容易にボットを検知でき，対策を講じることが可能でしたが，最近では，有害投稿に無害投稿を織り交ぜるなどして正規のユーザの特徴・行動を模倣したり，人間が時折操作して他のユーザと交流をしたりなど，検知を回避しようとするボットが出現し，検知が困難となってきています．

そこで，正規のユーザが意図的に悪性ボットをフォローすることは稀なため，①必然的にボットと正規ユーザ間のつながりは疎となりやすい，②アカウントどうしのネットワークにおいてボットと正規ユーザ間の構造的なギャップを上手く特定することで悪性ボットの検知が可能，③正規ユーザのフォロー行動を攻撃者が操作することは一般に困難なため検知回避もされにくい，といったネットワークの特徴を利用して，Sybil検知問題をグラフ信号処理における「信号復元問題」に落とし込むことで，既存のさまざまな検知手法を理論的に比較・分析することを可能にする方法論を確立しました（**図 3**）．そして，この方法論に基づいてSybil検知手法が高い性能を発揮するための要件を特定し，この要件を満たす新たな検知手法を提案しました．また，既存手法と比較して，提案手法はグラフ構造によらずに安定して高い性能を発揮することを数値実験によって確認しました．

この成果は，2023年 1 月にセキュリティ分野のトップ国際論文誌である「IEEE Transactions on Information Forensics and Security (TIFS)」に掲載されました[3]．

テーマ横断的な研究ではどのような取り組みをされているのでしょうか．

「オフェンシブセキュリティ×人間中心のセキュリティ」については，Webブラウザパーミッションメカニズムの解明です．

Webブラウザのパーミッション機構は，Webページがコンピュータやスマートフォンに対して行う操作を制御するための仕組みです．これにより，例えば音声や映像を利用するWebページにおいて，マイクやカメラにアクセスすることを許可するかどうかを選択することができます．このパーミッション機構は，多種多様なWebブラウザごとに実装差異があることに着目し，その実装差異を効率的に評価する方法，その実装差異に基づく新しいセキュリティ・プライバシー脅威を明らかにしました．

通常，Webブラウザは複数種類並列で使われることがなく，Webブラウザごとのパーミッション機構の実装差異には気付きません．実装差異は，ミス等による誤実装やバグのケースがあり，これに起因してセキュリティ・プライバシーの脅威となります．そこで，デバイス上で動作しているさまざまなWebブラウザを横並びに動作させて評価する環境を構築して，さまざまなOS上で動くWebブラウザの網羅的な比較評価を可能としました．これにより，パーミッション機構の多数の誤実装・バグを発見するとともにそれらを悪用したフィッシングやトラッキングなどの新しいセキュリティ・プライバシー脅威を明らかにしました．さらに，発見した脅威がユーザの認識や行動にどのような悪影響をもたらすかをユーザ調査を通じて定量化しました．これら結果に基づいて，Webブラウザのパーミッション機構に関してWeb標準化団体に対してベストプラクティスの共有や標準化を提言しました．

この成果は，2023年 2 月28日～ 3 月 2 日に開催された，サイバーセキュリティの 4 大トップ国際会議の 1 つである「The Network and Distributed System Security Symposium (NDSS) 2023」にて論文が採択されました[4]．

「社会技術的情報セキュリティ×人間中心のセキュリティ」としては，情報の背後にある“悪意（感情操作による情報蔓延）”

をユーザに知覚させる研究です．

誤情報や感情を操作するコンテンツは，AI（人工知能）による投稿が一般化する一方で，ソーシャルメディアプラットフォームがファクトチェック機能を縮小しており，さらにはファクトチェックそのものの限界もあり，公衆衛生や適切な意思決定に対するリスクがますます高まっています．こうした課題への対処法の１つとして，感情を操作するような表現がコンテンツに含まれるかどうかを検知し，それをユーザに提示することによって，不注意に信じ込んだり他人にシェアしたりする行動を抑制する効果が期待されています．私たちは，この「感情的な操作表現に対する警告ラベル」の有効性を検証しました．具体的には，健康関連のソーシャルメディア投稿において，情報の正誤や警告ラベルの有無がユーザの反応に与える影響を調査しました．その結果，感情的な操作表現を含むコンテンツに対しては，その内容が正しいかどうかにかかわらず，警告ラベルが内省的な行動（不用意なシェアを控えるなど）を促す効果を持ち場合があることが明らかになりました．これらの結果を踏まえ，誤解を招く，あるいは操作的なコンテンツの影響を抑えるためのプラットフォーム戦略や，今後の研究に向けた指針となる提言を行いました．

こうしたトップ国際会議・学会における論文採択が高く評価されて，2024年にNDSS，2025年にUSENIX Securityにおいてプログラム委員に就任しました．日本からのプログラム委員はほとんどおらず，NDSSでは30年以上の歴史の中で日本人として初めてのプログラム委員就任だそうです．また，世界有数の研究者の中でも特に優れた貢献（品質の高い査読，委員内の議論の牽引，献身的なシェパード活動）が認められ，NDSS 2025 Distinguished Reviewerを受賞しました．

さて，これらの研究が進むに伴い，例えばオフェンシブセキュリティで新しい脅威を見つけたときに，脅威のユーザに対するインパクトの調査が必要になり，情報の流通に際しては単に脅威の伝搬のメカニズムだけではなく，ユーザの判断に関する部分の調査も必要になるなど，３つのテーマの重複や連携を考慮していく必要が出てきます．そこで，３つのテーマが重複した部分が，攻撃者・ユーザ・情報環境の相互作用を通じて人間の認知や意思決定への影響に着目するセキュリティ研究，コグニティブセキュリティ（Cognitive Security）という新しいテーマです（図１）．

従来，多様なユーザ属性を考慮し，適切な意思決定や組織のポリシー策定のサポート技術や理論構築の研究において，心理学の方法論や学際的実施が推奨されてきたにもかかわらず，心理学的手法や認知に着目する研究はかなり少ない現状にあります．そこで，「認知」に着目して，個人を標的として短期的に行われるサイバー詐欺・ソーシャルエンジニアリングや，社会や組織を標的として認知・意思決定・行動そのものを操作する目的の偽・誤情報やアルゴリズムによる操作等の「認知にかかわる脅威」（**表**）に対抗していくテーマとして“コグニティブセキュリティ”を設定しました．

具体的なアプローチとして，①攻撃手法がどう影響を与えるか（認知能力・認知プロセスの理解），②認知に対する介入が自律的な意思決定にどう影響を与えるか（認知能力・認知プロセスに対する対策），③集団になることで，どう認知やその介入効果が変化するのか，あるいはしないのか（集団としての認知の理解と対策）といった観点で，「敵対的な状況下において自律的な意思決定を維持することで，認知・情報操作の影響から個人・組織・国家/社会を守るための方法論と実践」という課題解決をめざす，情報科学だけではなく認知心理学等も関係する学際的なテーマです．

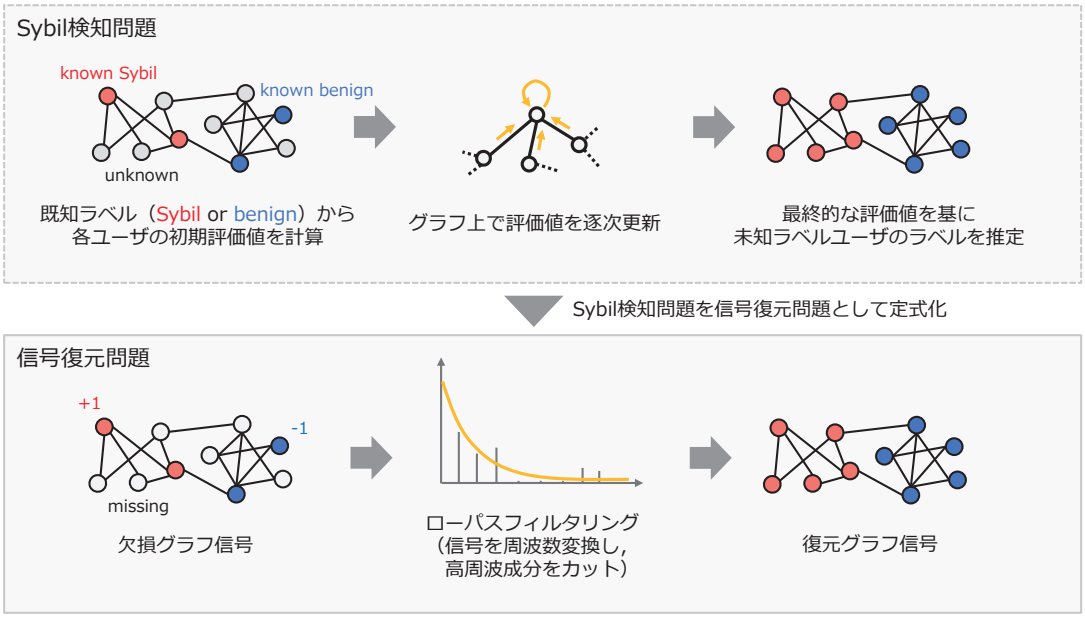

図３　Sybil攻撃対策技術

表　認知にかかわる脅威の分類

脅威の種類	具体例	主な影響内容	影響のレベル
サイバー詐欺・ソーシャルエンジニアリング	フィッシング，テックサポート詐欺，ビジネスメール詐欺	金銭や個人情報の搾取	個人
操作的デザイン	ダークパターン，FOMO（Fear of Missing Out）	ユーザが望まない選択・行動への誘導	個人
アルゴリズムによる操作	フィルターバブル，AI/LLMによる偏った・誤った回答	偏った意見・価値観の形成	個人～集団
偽・誤情報の拡散	ソーシャルメディア上での虚偽の発言，虚偽情報の共有	社会的分断の助長，民主主義の機能不全	個人～集団

自分の経験や知識だけを頼りに固執しようと思わず，疑うことやアップデートすることが重要

研究者として心掛けていることを教えてください．

通常，自分の経験や知識を活かして研究を進める機会が多く，場合によってはそれを無意識に絶対視することもあると思いますが，逆に自身の経験や知識が邪魔をすることがあることに気付きました．そこから，自分の経験や知識だけを頼りに固執しようと思わず，疑うことやアップデートすることが重要だということを意識しています．

私の研究は，しばしば人間を対象とした実験を行うことがあります．あるとき，日本で日本人を対象に実験を行った結果を論文にしたのですが，査読において，「なぜ日本で実験をするのか，日本の実験の結果が一般化できるのか」と指摘され，論文がリジェクトされたことがありました．この経験から，「国際論文として採択されるには，日本ではなく欧米を対象に調査するのが当たり前だ」と思い込むようになりました．その結果，日本に住んでいるにもかかわらず，欧米の人々を対象に英語で調査をして，結果を論文にまとめて発表するという時期が続きました．

その一方で，日本を拠点に研究をしているにもかかわらず，日本の実態を直接扱えないことに矛盾を感じ，フラストレーションがたまっていきました．さらに，研究のトレンドとしても，欧米人ばかりを対象とした調査が当然のように行われている状況に，違和感を覚えるようになりました．これがきっかけとなり，前述の「セキュリティ・プライバシー分野におけるユーザ調査研究の地理的偏りを定量的に分析」をテーマとした論文の執筆につながりました．この研究を通じて，たとえ定量的な調査であっても調査対象の環境・状況に偏りがある場合，その研究分野自体の普遍性に対する問題となり得ることを示しました．

これを機に，自分のこれまでの経験や知識だけを頼りにしようと思わずに，ときには疑うことやアップデートすることが重要だということを実感しました．

勇気と好奇心をもって専門の外に出て，相互理解，モチベーションの共有を図る

後進の研究者へのメッセージをお願いします．

自分の専門分野だけでは到底解決できない社会的課題があります．こうした課題に取り組むためには，学際的アプローチが必要になります．学際的研究を推進するには，自身の専門分野から外に出る勇気を持つことが重要です．ある分野の専門家でも，一歩外に出ると初学者になるので怖いという気持ちを抱きます．それに対しては，研究者としての好奇心で乗り越えることができます．そして，他の分野の専門家と連携するときには，共通点を見つけ，お互いに理解し合うことができます．逆に，相違点についても，それを興味深いものとしてとらえることで，理解を深めることができるはずです．さらに，モチベーションを共有することが大切です．たとえ専門分野が異なっても，同じモチベーションを持つことで，経験・知識の違いを超えて協働することが可能になります．

私の研究テーマは学際的なものが多く，その経験をとおしてこれらを実感しています．近年，研究分野の細分化が進んできており，こうした社会課題ばかりではなく，日常の研究においても学際的領域に踏み込む機会は多くなっています．だからこそ，自身の専門の外に出て，相互理解，モチベーションの共有を図ってください．

■参考文献
(1) A. A. Hasegawa, D. Inoue, and M. Akiyama: “How WEIRD is Usable Privacy and Security Research?,” USENIX Security Symposium 2024, Philadelphia, U.S.A., August 2024.
(2) F. Kanei, A. A. Hasegawa, E. Shioji, and M. Akiyama: “Analyzing the Use of Public and In-house Secure Development Guidelines in U.S. and Japanese Industries,” ACM CHI 2023, Hamburg, Germany, April 2023.
(3) S. Furutani, T. Shibahara, M. Akiyama, and M. Aida: “Interpreting Graph-Based Sybil Detection Methods as Low-Pass Filtering,” IEEE Transactions on Information Forensics and Security, Vol. 18, pp.1225-1236, 2023.
(4) K. Nomoto, T. Watanabe, E. Shioji, M. Akiyama, and T. Mori: “Browser Permission Mechanisms Demystified,” NDSS 2023, San Diego, U.S.A., Feb. -March 2023.