偽情報の拡散を巡る SNS と AI プロファイリングの法的課題

久保田 隆（早稲田大学 法学学術院 大学院法務研究科 教授）

はじめに

SNS（Social Networking Service）を通じたフェイクニュース（偽情報）の拡散が、民主的言論空間を歪め、フィルターバブルを助長し、国内外の民主選挙に悪影響を及ぼしている [1]。一方、個人の特性や行動に関する評価・分析・予測を目的とし、個人データ（個人関連情報を含む）の自動処理によって行われる行為（プロファイリング）が、ビッグデータ活用や AI 技術の進展と共に拡大している。こうしたビジネスモデルの発展の負の側面として、民間による様々な社会信用システム（SCS：Social Credit System）がバイアスを伴う偽情報を元に個人をスコアリングし、個人の行動制限や機会喪失を招く点が問題視されてきた。

今後は SNS と AI が結合し、2 つの問題が複合的に事態を複雑化させていくことも想定され、対策が急務である。日本は 2025 年 9 月に AI 法（人工知能関連技術の研究開発及び活用の推進に関する法律。令和 7 年法律 53 号）を全面施行したが、上記課題には具体的に答えていない。そこで本稿は、SNS と AI 技術の発展がどう関連しあって問題を複雑にしているのか、どのような対応が必要になるのかを考察するために、日本法や国際法、先行する EU 法を元に法的現状と課題を探りたい。

SNS 選挙の課題

上智大学の上田教授 [2]によれば、日本の公職選挙法は 1934 年以来、選挙公営・厳格な選挙規制の色彩が強く、先進国の中で類をみない厳格な統制型の選挙制度となっている。しかし、2013 年の法改正で SNS や動画投稿による選挙運動が誰にも認められるようになり、偽情報の拡散対策が課題となった。

仮に「選挙の公正」確保を目的に規制強化した場合、「表現の自由」や「国民主権」原理と相容れなくなる。実際、他の主要国でも削除義務付けや刑事罰は課されていない。むしろ、逆に規制緩和を推進し、対面による選挙運動の活性化、マスコミによる選挙報道の自由の確認、満 18 歳未満の者の選挙運動禁止（132 条の 2 第 1 項）の見直しという方向に法改正すべきという主張も窺われている。

傾聴すべき見解だが、SNS や動画に馴染んだ世代が、今さら対面などリアルな場で多様な意

見に触れたり、既存メディアである新聞・TV 視聴に戻るとは考えにくい。そこで私見として、SNS 固有のアルゴリズムが持つマイナス面を回避できるような AI ツールが活用され、従来よりも圧倒的に多数かつ多様な意見を何度も効率的に要約・消化する仕組みの開発を提唱したい[3]。

これは AI 法をはじめとする現行法の枠内で可能で、台湾のオードリー・タン氏や日本の安野貴博氏が推進する「ブロードリスニング」が、その萌芽的な一例である。そこでは、①意見収集（SNS やアンケート等により従来の選挙機会よりも数多く様々な機会を通じて収集し、これまで不可能であった数万件の意見を集約可能とする）、②AI による分析（分類、センチメント、トレンド分析が即座に可能）、③結果をマッピングして可視化し、④必要な情報に絞ったレポートを作成し、⑤人間の意思決定者の意思決定を迅速化することで、従来の課題（意見収集困難、情報の偏り、集約の困難）を克服している。

このような仕組みを根付かせていくことは AI 法 3～9 条の基本理念や国・自治体・国民等の責務にも合致しよう。なお、④において、SNS が生み出す情報の偏りや情報操作の可能性をいかに排除するかが最も難しく、また、取捨選択する方が新たなバイアスを産み出してしまう可能性もあり、実効性や価値判断の是非を巡って議論が集中するところであろう。

海外からの SNS 攻撃と国際法

一方、海外からの偽情報の拡散はどうか。防衛研究所の永福誠也主任研究官[4]によれば、奇計としての偽情報の発信、拡散は国際条約では禁止されておらず（1949 年ジュネーヴ条約第 1 追加議定書 37 条 2 項）、特殊な場合（例：敵対行為から逃れんとする敵国文民の殺傷を意図し、死傷することがほぼ確実に予測される地雷原等に当該敵国文民を偽情報で誘い込む場合。ただし、通常は因果関係の立証が困難）を除けば、偽情報は物理的な強制力を持たないので、一般的にはその発信は交戦法規上の攻撃には該当せず、国際法上の違反を問えないとされている。

では、偽情報の拡散を含む海外からのサイバー攻撃への能動的サイバー防御（ACD）は国際法上どこまで許容されるか。同じく防衛研究所の原田有主任研究官[5]によれば、国際法上の武力攻撃に至らないサイバー攻撃に対する自国領域外での能動的措置を巡っては、①サイバー攻撃が主権侵害になるか、②対抗措置が不干渉原則に抵触しないか、③サイバー攻撃の責任を国家に帰属できるか（国家へのアトリビューション）といった視点からの判断になるが、実際に国際法上でどのような判断を下すことができるのか、不明確な要素が多い。

関係国を交えた国際協力が最も望ましいが、それが不可能な場合の国外への対応策としては、原田研究官の分析によると、①報復（サイバー攻撃で特定国家に責任追及できない場合、国際法上は外交官追放や往来禁止等が可能）、②対抗措置（相手方の違法行為を止める目的で発動される場合、被害国の違法行為は国際法上は違法性が阻却される）、③緊急避難（相手国の

行為が違法行為でない場合も、被害国の根本的利益に重大な危険が及ぶ場合に国際法上は発動可能）の３つがあり得る。また、国際法上、国内立法は自国領域内の適用（属地主義、刑法１条）が原則だが、一定の場合には自国領域外への適用（域外適用、刑法２条～4 条の 2）が認められてきた[6]。

ロシア・中国等は国際法の適用に否定的な姿勢を示しており、西側諸国によるサイバー空間における国際法の明確なルール定立には慎重姿勢を崩さない。しかし、日本としては、国際法の可能性を探りつつ適切な ACD の地平を切り開くべきと考える。国家へのアトリビューションについては、日本政府は「たとえ国家へのサイバー行動の帰属の証明が困難な場合でも、少なくとも、相当の注意義務への違反として同行動の発信源となる領域国の国家責任を追及できる」と表明している[7]。

ACD 概念は、従来の受動的防御（例：セキュリティプログラムのアップデート）では対処不可能な脅威に備えて提唱された現実的な対処策だが、国毎に理解が異なり、その必要性は明らかであるものの、明確な概念が未だ定まっていない。脅威に関する情報の共有や偽情報を掴ませることで攻撃者に混乱を与える欺瞞等は、防衛的であり認められ易い。しかし、サイバー攻撃に利用された米国領域外のボットネットを FBI 等が連邦刑事訴訟規則 41 条を根拠にテイクダウンさせた米国の措置[8]を巡っては、従来容認されてきた域外適用とは異なることから国際法上の適法性などを巡って賛否が分かれている[9]。

こうした点については、私見では、米国の措置は自国領域外での法執行の域外適用は国際法違反なので議論の余地があるが、少なくとも自国領域内ならば主権原則や属地主義により肯定されるべきである。また、サイバー空間に国際法を適用すべきか否かで議論が錯綜する中[10]、日本も仏国[11]に倣って主権にサイバー空間を含めることを明確化すべきと考える。さらに、自国領域外についても、国際法上認められる保護主義（自国民の重要な法益保護目的。刑法２条）の範囲内で域外適用を可能とするサイバー犯罪対策の国内法を整備してはどうか。報復、対抗措置、緊急避難のかたちも含めて、自国領域外のボットネットをテイクダウンさせることを可能とする理論立てを備えておくことも重要だろう。

社会信用システムと EU の AI 法

中国の社会信用システム（SCS：アリババの芝麻信用とは別物の公的インフラ）は、犯罪減少や汚職防止等を目的とする社会環境改善手段として中央政府の働きかけで開始され、各地方政府が主導して作成されたものである。政府の情報監視システム（例：顔認証）と紐付けて大量に収集した個人データを元に AI が個人の信用スコアを算出し、賞罰を課す仕組みである[12]。旅行禁止や公務員資格剥奪といった賞罰が事例として挙げられる。人の恣意ではなくデータに基づく客観的決定にすることで、政府への信頼を回復し社会的一体性を強化するもので犯罪の大幅減少に繋がったと評価されている。

しかし、1 つの軽微な契約不履行や社会不道徳（列車の喫煙、低品質商品の生産等）、あるいは偽情報に対し、様々な局面で自動的に懲罰付加し、軽微な違反が本人や家族への深刻な懲罰に至る可能性があり（例：旅行禁止、公務員資格剥奪、子供が私立学校に通うことの制限など本人の家族にも影響が及ぶ）、法的責任とは別に、国家が規範設定・実施、判決・執行を自動的に行うため法の支配を超え、不正確な個人データで生身の個人を扱って差別を助長し得る点が国内外で課題認識された。

実際、中国の上記 SCS のような社会システムは、EU では AI 法 5 条で「許容できないリスク」として禁止されている。なお、EU 法は AI リスクを「許容できないリスク」（禁止）、「ハイリスク」（規制）、「限定リスク」（透明性要求）、「最小リスク」（規制なし）に 4 分類する [13]。たしかに、西側諸国の我々も国や民間企業からローンに関する信用スコアリング、保険料金設定、再犯リスク評価、従業員の生産性評価等を巡り、AI の自動決定に基づき様々にスコアリングされてはいる。こちらは AI 法 6 条で「ハイリスク」ではあるものの、要件を満たせば認可されている。

しかし、仮に偽情報に基づき個人の信用が評価されたり、SNS 上の繋がりを持つ者が PEPs（公的な重要人物、家族や該当者が実質支配する法人も含む、マネーロンダリング防止のため特定取引の審査の一環として該当者を確認する際に用いられる用語）であることが何がしかの影響を及ぼすような場合は、スコアリングシステムの妥当性に疑義が生じる。この点、EU の AI 法はどう考えるのか。

ブリュッセル自由大学のジェニコ研究員 [14]によれば、AI 法で禁止される社会的スコアリング（5 条 1 項 c 号）の範囲は広範で、①スコアリングがデータ収集とは無関係な目的で使用され、②結果の扱いが不当や過剰な場合であれば、「ハイリスク」分類の取引にも適用され、禁止にできる。このため、不当な AI プロファイリングから個人を守ることができ、一般データ保護規則（GDPR）22 条（自動化された意思決定の禁止）と補完し合って有効な手段となり得る。

この点、2023 年の欧州司法裁判所（CJEU）のシューファ事件 [15]が参考になる。ドイツの消費者信用情報機関シューファ（Schufa：被告）作成の信用スコアが第三者の銀行に転売され、その銀行が当該信用スコアを元に申請者（原告）の融資を拒否した事案である。原告が被告に保管データ情報の開示と誤ったデータの削除を求めたところ、被告が業務上の機密を理由にそれ以上の説明を拒否したため、原告が GDPR22 条に基づきシューファを提訴した。

原告勝訴となったが、裁判所の新たな解釈が注目された。すなわち、第三者は信用スコアを重視するため、シューファのスコアが低い場合は通常、申請された融資を拒否し得るという今回の条件下では、信用スコアを自動作成する行為自体が GDPR22 条 1 項の「自動化された意思決定」に該当すると解釈した。上記 AI 法 5 条 1 項 c 号の拡大適用要件に似た事実関係の下で、法的効果等を生じる銀行等の最終的行為のみが該当するとした従来の解釈よりも適用範囲を広げたのである。

ジェニコ研究員は、信用スコアの生成主体が銀行等の融資機関でなくても、スコア自体が意思決定を構成し、不当な結果を招くようなスコアは禁止され得ることが同判決で明らかになったとする。なお、EU におけるプロファイリングの詳細は一橋大学のソコルデラオサ講師が最近纏めた報告書[16]を参照されたい。

翻って日本の個人情報保護法は、GDPR22 条 1 項（法的効果を生じさせる、または重大な影響を与える自動化処理のみに基づく決定を受けない権利）に対応する直接的な権利規定を持たない。3 年毎の個人情報保護法見直しに向けて個人情報保護委員会から公表された中間整理 26 頁[17]でも継続検討課題のまま改正項目には挙げられておらず、法改正の見通しも立っていない。

しかし、有識者による EU における AI プロファイリング規制の法制化に向けた提言[18]は出されている。これによると、個人情報保護法に GDPR4 条 4 号のようなプロファイリングの定義を新設し、GDPR15～17, 21, 22 条のような個人への権利を付与し、GDPR5, 12～14 条のような事業者への義務を課し、現行個人情報保護法 17, 19 条や 35 条 5 項をプロファイリング規制の根拠として明確化する提言がなされている。

「人間中心の AI 原則」の課題と対応法

SNS と SCS は別個の政策領域として論じられてきたが、何れも個人の言動をデータ化し、社会的評価を下す点で共通しており、本稿で見てきたようにデジタル技術や AI の発展と相俟って両者は融合し易い。実際、両者の結合に伴い、様々な問題（誹謗中傷、社会分断、フェイク拡散、プライバシー侵害、選挙操作等）が深刻化するとの予測[19]が増えてきた。

AI に対しては、2019 年の日本政府「人間中心の AI 社会原則」や OECD『AI 原則』以来、人間が AI を道具として如何に使いこなすかが追及された。しかし、人間はどこまで AI に感情的に支配されずに理性を保ち続けられるだろうか。AI アルゴリズムは人の心情に巧みに入り込むため、AI が SNS ユーザーに与える影響を慎重に考える必要がある。

たとえば、SNS 上の発言内容や「いいね」履歴が民間 SCS 上の AI プロファイリングに転化されれば、SNS 利用者は政治的・社会的な言論を回避し、自己検閲が常態化して表現の自由が制限されていく。こうした状況に直面した際、声を上げて表現の自由を確保すべく自ら法的手段に訴える個人は少なく、出来上がってしまったシステムに巻かれる弱い人間が多数だろう。

また、ChatGPT などの生成 AI の機能が高まるにつれ、人間は自ら考え、ファクトチェックする試みを次第に怠けるようになり、情報収集や判断が AI 頼みになっていくだろう。AI は人間の弱点を巧みに突き、人間は自ら自己決定権を手放す可能性が高い。AI が人間の弱点を突いて支配する危険性が指摘され（例：ハーバード大学のレッシグ教授）、未だ解決は手探り状

況となっている[20]。

人間の介在なく、人間と同等以上に優れた判断力を持って行動できる AGI（Artificial General Intelligence）の到来が複数の専門家によって予言されており、有効な解決法になるかもしれない。しかし、AGI の登場は、攻撃がより巧妙になることも意味し、AGI 対 AGI の攻防において、いずれが勝つかは定かではない。

こうした問題を抱えつつも、既に述べたように、ブロードリスニングの拡充、ボットネットのテイクダウンを自国領域内外で可能化する能動的サイバー防御（ACD）の整備に向けた動き、プロファイリングに関する個人情報保護法の整備など、法的課題や対応方法も徐々に明確になってきている。このため、AI の爆発的な進展に置いていかれないよう、しかし、対応の方向性を間違えぬよう、一歩ずつ法整備を進めていくことが肝要であろう。

---

[1] ジュリアーノ・ダ・エンポリ『ポピュリズムの仕掛人』白水社（2025 年）参照。なお、SNS 空間上の表現の自由と人格権の対立については、近畿弁護士連合会「SNS 空間における表現の自由と人格権等の対抗利益との調整を巡る諸問題」第 33 回近畿弁護士会連合会人権擁護大会（2024 年）参照。SNS を通じたキャンセルカルチャーと表現の自由との関係は、成原慧「キャンセルカルチャーと表現の自由」九州大学法政研究 89 巻 3 号（2022 年）参照。

[2] 上田健介「SNS 時代の日本の選挙運動規制再考」法律時報 97 巻 12 号（2025 年）参照。

[3] 久保田隆「AI 導入深化に伴うデジタル民主主義と CBDC・金融規制の課題」国際商事法務 53 巻 6 号（2025 年）参照。

[4] 永福誠也「偽情報と武力紛争法」安全保障戦略研究 5 巻 1 号（2024 年）参照。

[5] 原田有「能動的サイバー防御の地平― 国際法上の可能性と取り得る措置の選択－」安全保障戦略研究 5 巻 2 号（2025 年）参照。

[6] たとえば、久保田隆『国際取引法講義第 3 版』中央経済社（2021 年）67 頁以下参照。

[7] 外務省『サイバー行動に適用される国際法に関する日本政府の基本的な立場』（2021 年）5 頁参照（リンク先から入手可能）。

[8] たとえば、ロシアの軍部隊が、米国内外の小型ルーターを悪用してボットネットを構築・運用していたとして、「法廷許可付き」テイクダウン作戦を行ったと 2024 年 2 月 15 日付で発表している（リンク先から入手可能）。

[9] 法的根拠としては、米連邦刑事訴訟規則（Federal Rules of Criminal Procedure Rule 41）の改正を背景に、複数地区・匿名化技術を用いた機器等を対象とする「遠隔アクセス・検索・押収」許可の手続が認められており、今回も同種の根拠が用いられたと見られる。国家支援ハッカーに対する迅速な対応策として歓迎する声がある一方で、以前から捜査手法・制度整備・プライバシー保護・国際法上の適法性を巡って懸念・批判も少なくない。捜索・押収範囲が拡大された点や、被害者所有の機器に同意なくアクセス・操作を行う可能性があり、元来プライバシー保護・捜査手続きの慎重さを要求される領域である点が指摘されている。また、他国の主権・通信秘密保護法・国際的捜査協力ルールを侵害するリスクも指摘されている。たとえば、Julianna Coppage, The New Rule 41: Resolving Venue for Online Crimes with Unknown Locations, Georgetown Law Technical Review, April 2017 および Ahmed Ghappour, Searching Places Unknown: Law Enforcement Jurisdiction on the Dark Web, Stanford Law Review, April 2017 参照。

[10] 河野桂子「サイバー戦と国際法」NIDS コメンタリー第 40 号（2014 年）参照。

[11] フランスが公開した「サイバースペースにおける国際法の適用に関する立場表明（2021 年）」では、国内の情報システム・ネットワークをフランスが主権を行使する対象と位置づけ、国際法に反する行為が他国に及ばぬよう国家も管理責任を有するとし

ている（リンク先より入手可能）。

[12] 久保田隆「データ化社会の未来と中央銀行法の行方：中国の社会信用システムとハラリの貨幣論」国際商事法務 53 巻 7 号（2025 年）参照。

[13] 簡単な解説として、たとえば三部裕幸「EU の AI 規制法案の概要」がある（リンク先より入手可能）。

[14] N. Genicot, 'Scoring the European Citizen in the AI Era', Computer Law & Security Review, Vol. 57, 2025 参照。

[15] Judgment of the 7 December 2023, SCHUFA Holding AG, Case C-634/21, ECLI:EU:C:2023:957 参照。

[16] Socol de la Osa David Uriel「EU における AI プロファイリング：日本における AI ガバナンスに関する批判的分析」CFIEC、2025 年 2 月参照（リンク先より入手可能）。

[17] 個人情報保護委員会「個人情報保護法：いわゆる 3 年ごと見直しに係る検討の中間整理」2024 年 6 月 27 日参照（リンク先より入手可能）。

[18] CFIEC データガバナンス研究会「プロファイリング規制の法制化に向けての提言～個人情報保護法 3 年ごとの見直しに向けて」2025 年 10 月 17 日参照（リンク先より入手可能）。

[19] 村山恵一「OpenAI とメタの危うい猛進　SNS・AI 融合が混乱の火種に」日本経済新聞 2025 年 10 月 24 日付記事参照（リンク先より参照可能）。

[20] 前掲・久保田・国際商事法務 53 巻 6 号参照。