

偽情報にどう向き合おうか

—— 憲法・メディア法の視点から ——

水谷瑛嗣郎

1. 導入——偽情報と法規制

インターネット上の偽情報の拡散に対処する法制度は、近年になって各国で進められつつある。EUによる偽情報対策行動規範 (The 2022 Code of Practice on Disinformation) 及びデジタルサービス法はその典型例といえる。前者は、二〇一八年に策定され、オンライン・プラットフォーム (PF) 事業者をはじめとする各事業者のコミットメントを得た行動規範を、欧州委員会のガイダンスに従って強化したものである。偽情報の発信者や事業者を直接的に規制、処罰するものではなく、署名した事業者に対して、偽情報のデマネンターション、政治広告の透明性確保、ユーザー・エンパワーメント、ファクトチェック組織の強化とい

った各種対策を要請している。加えて、デジタルサービス法の制定により、超巨大オンライン・プラットフォーム (VLOP) 事業者に指定されている事業者にとっては、同法四五条の共同規制フレームワークのもとに組み込まれることが目指されている⁽¹⁾。

これに対して、アメリカ合衆国においては、よく知られているように憲法修正一条に基づく表現の自由の重視と通信品位法二三〇条による免責規定の存在が障壁となり、基本的に連邦レベルでの偽情報対策の法規制は存在していない⁽²⁾。ただし、州レベルでは規制が試みられている。例えば、二〇二四年九月にカリフォルニア州のギャビン・ニューサム知事によって、選挙時における三つのデープフェイク規制法が署名された。

そのうちの二つである AB2655 は、選挙期間中に巨大 PF 事業者に対して、選挙関連の「実質的に欺瞞的なコンテンツ (materially deceptive content)」を特定し、削除 (§20513(a)) またはラベリングする (§20514(a)) 義務を設けている。また同法は、カリフォルニア州在住のユーザーに対して、削除またはラベリングされるべきコンテンツを通報するための「容易にアクセス可能な手段を提供」しなければならないとされる (§20515(a))。もう一つの AB2839 は、「個人、委員会またはその他の団体」に対して、「悪意を持ち、または故意に」、例えばカリフォルニア州の連邦・州・地方公職者候補が実際に発言していない内容を発言しているかのように見せかけた「実質的に欺瞞的なコンテンツを含んだ広告またはその他の選挙運動用のコミュニケーション」を配信することを禁止した (§20012(b)(1))。この AB2839 に関しては、連邦地裁に予備的差止め請求 (preliminary injunction) がなされ、連邦地裁が同法は修正一条違反の可能性が高いとして、差止めを認めている⁽³⁾。

他方で、アジア圏に目を向ければシンガポールにおいて、二〇一九年に「オンライン上の虚偽情報及び操

作からの保護法 (Protection from Online Falsehoods and Manipulation Act, POFMA)」が制定された。同法は、安全保障を損なわせる、公衆衛生を損なわせる、各種選挙の結果に影響を与える等の可能性がある虚偽事実の発信を禁止 (七条一項) しており、さらに Facebook や Twitter などの PF 事業者を、「インターネット仲介事業者 (the internet intermediary)」と位置づけ、虚偽事実が流通していた場合には事業者に対して特定是正命令 (Targeted Correction Direction) 等を出すことができる (二一条一項)⁽⁴⁾。さらに、PF 事業者に対しては、オンライン政治広告の透明性、信頼できるオンライン情報源の優先、オンラインアカウントの悪用防止および対策に関する三つの行動規範が策定されている⁽⁵⁾。ただし、司法救済の不十分さ等の理由によって、当該法律は、法の支配の原則を解体してしまうとの指摘もある⁽⁶⁾。

2. 虚偽と表現の自由

このように各国では偽情報に対する対応を進める一方で、表現の自由との衝突が問題となる。翻って日本においては、これまで虚偽の言論にも、表現の自由に

おける一定程度の保護が付与されてきた。例えば、最高裁は、虚偽の事実の摘示による名誉毀損について、「事実が真実であることの証明がない場合でも、行為者がその事実を真実であると誤信し、その誤信したことについて、確実な資料、根拠に照らし相当の理由があるときは、犯罪の故意がなく、名誉毀損の罪は成立しない」と判示している⁽⁷⁾。加えて、北方ジャーナル事件における谷口裁判官の意見では、「誤った言論にも、自由な討論に有益なものとして積極的に是認しうる面があり、真実に反する言論にも、それを保護し、それを表現させる自由を保障する必要性・有益性のあることを肯定しなければならない」と判示されているところである⁽⁸⁾。

虚偽の言論が憲法上、一定程度の保護を受けるのは、はいくつか理由が考えられる。消極的な理由としては、こうした虚偽の言論を保護せず、違法化してしまうことにより、自由な表現活動に対する萎縮効果が生じる可能性を除去するため、さらに虚偽の言論を取り締まる法規制が政府によって（例えば反政府的発言の恣意的な摘発に利用される等）濫用される可能性を警戒したものと見える。他方で、より積極的理由としては、い

わゆる思想の自由市場の中で虚偽の言論が是正されていくというプロセスの有用性があるためということが考えられる。

もつとも、現行法上、偽情報が違法とされる場合がある。例えば、偽情報による権利侵害については、虚偽の事実摘示による名誉毀損、なりすましによる肖像権侵害等によって規制されており、これ以外の違法な偽情報として、偽計業務妨害罪（刑法二三三条）、虚偽事項公表罪（公職選挙法二三五条二項）などの規制が存在している。問題は、こうした権利侵害及び違法な偽情報に該当しえないもの、何らかの有害性を含む偽情報、例えば虚偽事項を含むヘイトスピーチ、公衆衛生を害するような虚偽の医療情報等について、どのように対応すべきかであろう。

この点、総務省に設置された「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「とりまとめ」（令和六年九月）では、偽情報に対するいくつかの対策が検討されている。まず「情報流通の健全性確保に向けた基本的な考え方」（第五章）として、「表現の自由と知る権利の実質的保障及びこれらを通じた法の支配と民主主義の実現」、「安心かつ安全

で信頼できる情報流通空間としてのデジタル空間の実現、「国内外のマルチステークホルダーによる国際的かつ安定的で継続的な連携・協力」という三つの高次の基本理念が掲げられ、それに基づく情報の発信・伝送・受信それぞれに関する基本理念が設定された。そのうえで、政府や地方公共団体、伝統メディア、ファクトチェック組織、情報伝送PF事業者などのステークホルダーにそれぞれ求められる役割が示されている。

「とりまとめ」では、特に情報伝送PF事業者に対しては、制度的対応も検討されている（第七章）。まず対応すべき偽・誤情報については、①「検証可能な誤りが含まれていること」、②「当該情報そのものがある権利侵害性その他の違法性や客観的な有害性（及びその明白性）」・「当該情報が流通・拡散することによる社会的影響の重大性（及びその明白性）」・「誤りが含まれていることについての検証の容易性（誤りが含まれていることの明白性）」という「各要素の有無・軽重に照らし、具体的な方策との関係で比例性が認められること」が範囲とされた。

そのうえで、偽・誤情報を、①「他人の権利を侵害する違法な偽・誤情報」、②「行政法規に抵触する違

法な偽・誤情報」、③「権利侵害性その他の違法性はないが有害性や社会的影響の重大性が大きい偽・誤情報」に区分し、③については、表現の自由とのバランスから、政府による削除要請等は設けられず、「影響評価・軽減措置の確実な実施を求める枠組みの活用を含め、情報伝送PF事業者による取組を促す観点が必要」とされている。

ここでいう影響評価・軽減措置とは、「将来にわたる社会的影響を事前に予測し、その結果を踏まえて、影響を軽減するための措置（サービスマーケティングの変更、利用規約等の変更、コンテンツモデレーションの方法・プロセスの変更、レコメンデーション機能の変更等を通じた措置）」を指している。こうした影響評価・軽減措置のために重要となるのが、その評価のための指針などを設定するための民産学官によるマルチステークホルダープロセスである。この点についても「とりまとめ」では、「国内外の民産学官のマルチステークホルダーが相互に連携・協力しながら安定的かつ継続的に議論・検討する枠組みについて、マルチステークホルダーにより構成される協議会の設置に関する制度整備も含め、具体化を進めることが適当」とされた。

3. 偽情報とプラットフォーム法

前記の「とりまとめ」とも関連するところであるが、オンライン上、特にソーシャルメディアにおける偽情報拡散に対して、ソーシャルメディアを運営するPF事業者も、自主的に定めたポリシーに基づき、ユーザー生成コンテンツを監視し、違反コンテンツを削除する、いわゆるコンテンツ・モデレーションと呼ばれる仕組みによって対処している。⁽⁹⁾

例えば Meta はコミュニティ規定において、削除対応すべき「偽情報」として、「人々に対する差し迫った暴力または実際の危害のリスクに直接つながる可能性が高いと、専門家のパートナーが判断した偽情報および検証できない噂」、「公衆衛生と安全に対する差し迫った危害に直接的な害をもたらす可能性の高い健康誤情報」、選挙や国勢調査のような「プロセスに人々が参加する能力を妨害するリスクに直接つながる可能性が高い偽情報」を削除対象として定めている他、「コミュニティ規定に違反していなくても、コンテンツが、重要な公共性を持つ事柄に関して著しい誤解を招く可能性が特に高い、デジタルで作成または加工された写真のようにリアルな画像または動画、または実

際の音声に聞こえる音声」については、情報ラベルを配置する、広告の場合は却下するといった措置をとることが示されている。⁽¹⁰⁾

他方、X (旧 Twitter) は、X ルールの中で、「選挙やその他の市民活動への参加方法について、誤った情報や誤解を招く情報」、「選挙やその他の市民活動に参加しないよう住民を脅迫、抑圧することを目的に、市民活動を取り巻く状況について、虚偽であることが立証されている情報や、誤解を招く情報を流布すること」が禁止されており、削除等の対応がとられるとされている。また COVID-19 に関しても「明らかに誤ったコンテンツ、誤解を招くコンテンツ、重大な損害(ウイルスへの暴露を増やす、公衆衛生システムに悪影響をもたらす、など)をもたらす可能性のあるコンテンツ」を禁止している。⁽¹¹⁾ 加えて、「信頼のおける情報として共有されているものの、大幅な改ざん、操作、ねつ造により、その意味が根本的に変化しており、公共の問題について広範囲にわたって混乱をもたらしたり、公共の安全に影響を及ぼしたり、深刻な損害をもたらしたりする可能性のある」メディア、例えば実在の人物を描写しているものの、AI等を用いて捏造

されているメディア等が禁止されている⁽¹²⁾。またXの特徴の一つとして、「誤解を招く可能性があるポストに、Xユーザーが協力して役に立つノートを追加できるようにすること」で、より正確な情報を手でできるようにすること」を目指した「コミュニティ・ノート」と呼ばれる仕組みが挙げられる。コミュニティ・ノートは、資格条件をクリアした登録ユーザーによってUGC (User Generated Contents) がチェックされるボランティア・ベースの対応措置であり、削除ではなく、情報(ノート)の付加を行う仕組みとなっている。この仕組みでは、単なる多数決によってノートを評価するのではなく、視点の多様性を確保するために、「過去の評価において、評価が相違することのあった協力者の間で『役に立つ』という評価が一致すること」が求められる⁽¹³⁾。

YouTubeのコミュニティ・ガイドラインにおいても「特定の種類の誤解を招くコンテンツまたは虚偽が含まれるコンテンツで、重大な危害を及ぼす可能性があるもの」が禁止されており、例えば投票妨害や候補者の適格性など、「自由かつ公正な民主的選挙に関連する特定の種類のコンテンツ」や、予防や治療など、

「特定の健康状態や物質に関して地域の公衆衛生機関(LHA)や世界保健機関(WHO)が発信するガイダンスと矛盾する情報」が挙げられている。これ以外に動画クリエイターに対して、「コンテンツについて視聴者が常に正しい情報を得られるよう、コンテンツが実物のように見える場合は、意味ある変更や合成によって作成されたコンテンツであることを開示する」ことを求めている⁽¹⁴⁾。

最後に、特に一〇代を中心とした若い世代に人気を博しているTikTokもコミュニティ・ガイドラインの「誠実性と信頼性」ポリシーの一環で、「誤情報」と「陰謀論」を禁止している。例えば、「公共の安全に危険を及ぼしたり、危機的状況または緊急事態においてパニックを引き起こし得たりする誤情報」、「ワクチンに関する誤解を招く発言……など、健康に関する誤情報」、「確立した科学的コンセンサスを阻害する気候変動に関する誤情報」、「個人を名指しして攻撃する、陰謀論」、「暴力的またはヘイトに満ちた陰謀論」は禁止されている。さらに投票等の方法・時期・場所に関する誤情報、「選挙その他の市民活動の組織化と実施に適用される法律、プロセス、および手順」に関する誤

情報、「選挙の最終結果や成果」に関する誤情報といったコンテンツが禁止対象となっている。またこれ以外の誤情報・陰謀論についても「おすすぬ」フィードの対象外措置がとられる場合がある。AI生成コンテンツや編集されたコンテンツについては、ラベル付けの対象とされている他、「政治問題、商業製品、公共にとつて重要な問題に関与する立場にある公人」をはじめとする公人を対象とした誤解を招くAI生成コンテンツ等や、自然災害・紛争等を対象とした誤解を招くAI生成コンテンツ等は禁止されている⁽¹⁷⁾。

こうしたPF事業者によるコンテンツ・モデレーションは相当程度、大規模に及ぶ。一例として、TikTokは、二〇二四年第一四半期(一〜三月)において、約一億六七〇〇万件の動画を削除している。このうちの二・七%が「完全性(※原文ママ)」と信頼性「ポリシー違反に該当し、さらにこのうち一四・二%が「選挙関連情報の信頼性」、二六・一%が「誤情報」、一二%が「合成メディアと操作されたメディア」に該当するとされている⁽¹⁸⁾。そのため、対応の多くの割合が人間の手を離れて自動化され、アルゴリズムに処理されている⁽¹⁹⁾。

このように各PF事業者が実装しているコンテンツ・モデレーションの仕組みは、各国の法規制と連動しつつも、独自の体系を構築している。PF事業者は、「ユーザーの提供するデータを採取・収集・蓄積・占有・採掘・加工することを通じ、商業的利益のためにそれらを利用して収益化を図る、価値の捕獲のための装置として重要な推進的役割」を果たしており、生まれながらにして「独占化」の傾向を有する(「プラットフォーム資本主義」)。そしてプラットフォーム資本主義には、「新封建主義へと向かう運動法則が内包されている⁽²⁰⁾」。そうした新封建主義的存在となりつつあるPF事業者は、オンライン・コミュニケーション・インフラにおいても「新たな統治者⁽²¹⁾」として、まさに「プラットフォーム法」を構築しつつあると言える⁽²²⁾。偽情報対策もまた、こうした「プラットフォーム法」の存在を抜きにして語ることはできないのである。

4. 構造(AI)×人間——「ハイプ・マシン」としてのソーシャルメディア

加えて、オンライン上の偽情報問題は、オンライン環境を形成している「構造」的な問題でもある。例え

ば、ユーザーの嗜好を予測するアルゴリズムによって形成される「フィルターバブル」は、人々を「自分ループ」に陥らせ、多様な視点を喪失させてしまうことが懸念されている。⁽²³⁾ こうした「バブル」を形成するために、PF事業者は、ユーザーを「監視」し、データを集め、AIを介して分析することにより、ユーザーの心に接近する（「監視資本主義」⁽²⁴⁾）。

また私たちのオンライン体験は、AIをはじめとする情報技術のみによって決定されるわけではない。この点、MITのシナン・アラルは、ソーシャル・ネットワーク・システムのことを「ハイブ・マシン」と呼び表している。アラルによれば、ハイブ・マシンは、デジタル・ソーシャル・ネットワーク、ハイブ・ループ、デバイス（スマートフォン）という三つの要素で構成されているが、このうち、ハイブ・ループは、レコメンド・システムをその典型とするように、AIと人間の相互作用プロセスによって情報流通が決定づけられていることを示唆した概念である。当該概念は、ソーシャルメディアを利用するユーザー（人間）と、構造に埋め込まれたAIの関係性をよく表している。AIが大量のデータから人間の嗜好を予測し、情報を

表示する一方、人間はAIにより表示された情報に何らかのアクション（*action*）を起し、そのフィードバックを受け、スルーする（*scroll*）を起し、そのフィードバックを受けてAIはさらに精度を高めていく。このように人間とAIは、どちらか片方が一方的に片方に影響を与えるわけではなく、密接に結びついたフィードバック・ループを形成しているのである。⁽²⁵⁾

加えて、これらAIを含めたPF上のアーキテクチャのデザイン（設計）やコンテンツ生成・発信プロセスの背景に、経済原理が深くかかわっていることは見逃せない。例えば、アラルもハイブ・マシンの背景に「アテンション・エコノミー」が働いていることを指摘している。情報が溢れるオンライン環境では、いまや稀少財となった私たちの「アテンション（注目）」を奪い合うことに注力し、PFやコンテンツに対するユーザーの「粘着性」を高めることに注力するビジネスモデルが加速しつつある。⁽²⁶⁾

この点、ソーシャルメディア上の偽情報の発信または流通によって、経済市場の関係者がどのように利益を享受しているのかに焦点を当てることが重要であるとの指摘もある。特にマーケットシェイピングのよう

な市場ベース理論は、こうした偽情報の発信・流通を生み出す「修正が必要な市場機構について、より完全で実行可能な理解をもたらす可能性がある」という。

例えば、偽情報によって利益を得ているのは発信者や P F 事業者だけでなく、アドテク企業やインフルエンサー、そして著名なブランドなど、多くのアクターが何らかの形で偽情報から利益を得る構造 (Ho marketplace for misinformation) がある。また P F 上の滞在時間とインタラクシオンの最大化を目指したエンゲージメント指標が偽情報の発信と拡散に影響を及ぼしており、偽情報に対処するためのコンテンツ・モデレーションやデマネイゼーション措置にも抜け穴がある。加えて、P F 事業者の民主的ガバナンスをいかに進めるべきかに関する研究もまた必要となる。こうしたそれぞれの領域においてさらなる研究を進めることによって、この課題における適切な政策立案が可能になるのである。⁽²⁷⁾

また人間と A I のハイブリッドな環境となっているソーシャルメディアでは、例えば「フィルターバブル」を促進する可能性が指摘される一方、分極化への影響力は、当初の想定よりは緩やかな可能性も指摘さ

れており、アルゴリズムの影響力についてはより長期的な視点が必要となる。つまるところ重要なのは、偽情報の発信に経済的インセンティブを与え、さらに人間と A I がハイブリッドに絡み合うことで自身の選好に捉われがちな環境に対して、「民主主義社会における思想及び情報の自由な伝達、交流の確保という基本的原理」のために必要となる「各人が、自由に、さまざまな意見、知識、情報に接し、これを摂取する機会」をいかに確保するのか、そうした民主主義ひいては表現の自由を確保するためのオンライン環境のデザインとはどのようなものかということになろう。⁽³⁰⁾

5. オンライン環境デザインの二つの方向性

こうしたオンライン環境の(憲法フレンドリーな)デザインという観点を踏まえるならば、偽情報への対応には二つの方向性があり得る。

一つは、偽情報をはじめとする違法・有害情報の流通抑制とユーザーの表現の自由のバランスを構築するため、ステークホルダーに求められるガバナンスモデルの構築であろう。特に前述のようにオンライン環境で非常に強力なコンテンツの統治者としての立ち位置

を占める巨大PF事業者をどのようにガバナンスすべきか、という点は欠かすことはできない。

ボワーズとジットレインは、インターネット・ガバナンスが、草創期の権利ベースの時代から公衆衛生モデルの時代に移行していることを指摘する。すなわち現代の偽情報の問題は、SNSにおける偽情報の投稿一つが単独で多大な影響をもたらすわけではなく、それら有害な行動の慢性的パターンの蓄積や組織的行動によって生じる「遅効性の害悪 (slow poisoning)」にある。そのため、そうした害悪への対処も、「集合的な影響をモデル化し、対処する」ための「公衆衛生」モデルへと移行しているのである。これに対して、ボワーズらは、偽情報の時代においては、関係する諸企業に「権利と公衆衛生上の利益を首尾一貫してバランスングする責任を負わせる」ために、より説明責任を重視した「プロセス」モデルへ転換すべきであると提唱している。その際には、「プラットフォームにコンテンツガバナンスに関する新たな義務を課す解決策と、コンテンツガバナンスプロセスの重要な側面をプラットフォーム自体の外に委ねる解決策の両方を考慮する必要」があるという³¹⁾。このように偽情報に対する課題

を捉えると、必要となるのは、従来の「表現規制」対「(個人の) 表現の自由」という権利を中心にした対応よりも、むしろオンライン環境の構造に対するデザインやガバナンスモデルをベースにした対応ということになる。

こうしたガバナンスモデルのもとでは、適切なリスクベースに基づいたアプローチが推奨される。というのも、共通するリスク評価指標を設計して議論することで、「潜在的な被害の重大性と発生可能性に基づいてリスクと介入を優先順位付けし、党派に関係なく普遍的な負の結果に焦点を当てた議論を行う」ことが可能となる。また「継続的なモニタリングと評価を可能にし、リスク軽減策の継続的なモニタリングと評価の仕組みを確立する」ことで、PF事業者の説明責任と透明性確保を進展させ、藁人形論法や水掛け論を減らし、より建設的なプラットフォーム・ガバナンスを構築することに繋がり得るためである。こうしたリスク・アセスメントの一環として、コンテンツ・モデル・アクション・システムの「質」を評価するためのメトリクスの開発、特にコンテンツ・モデル・アクションの自動化が進むなかでのエラー率の活用は(課題もあるが)

必要かつ急務であろう。⁽³³⁾

第二に、偽情報の削減・抑制のみならず、信頼できる情報の生成を持続可能にし、これらにユーザーが容易にアクセスしうる環境及びエコシステムを構築するという方向性が挙げられる。偽情報から様々な形態で利益を享受するエコシステムと市場をどのように修正すべきかという観点から真っ先に検討されるのは、コストをかけてコンテンツの質の担保を行う報道機関への適正な対価配分を確保するための競争政策であろう。⁽³⁴⁾ ただし重要となるのは、ここでいうコンテンツの「質」をどのように測定するかである。オンラインエンゲージメントの指標が、情報内容の正確性や公共性を考慮せずに「刺激」を重視する状況を拡大させている以上、これとは異なる視点での価値指標の開発と実装が必要とされる。例えば、国境なき記者団が開発している Journalism Trust Initiative⁽³⁵⁾ は、メディア組織のガバナンスと透明性を自己評価させ、第三者認証する野心的な試みである。この取り組みにおいては、コンテンツの質の担保を、コンテンツの生産者（パブリッシャー）のガバナンスと透明性に求めるのである。⁽³⁶⁾

こうした指標の開発は、ソーシャルメディア上にお

ける質の高い（と推定された）コンテンツの優先表示、すなわち「プロミネンス」を可能とするための必要条件となる。⁽³⁷⁾ またそれとは逆に、PF事業者によるコンテンツ・モデレーションの例外的措置に際しての識別指標としても機能しうる。⁽³⁸⁾ いずれにせよ、今後こうしたアプローチをオンライン環境で実現していくためには、現在の複雑化したデジタル広告エコシステムの透明性を確保し、PF事業者をはじめ、関与するステークホルダー（広告主、広告代理店等）の機能に応じた適切な役割分担を議論していく必要がある。⁽³⁹⁾

最後に、いずれの方向性においても、その実効性を確保するために、法規制も視野に入れた政府による一定程度の関与が必要不可欠となる。ただしその際には政府による濫用リスクを招く新たなコンテンツ規制や個別のコンテンツ・モデレーションへの介入は必要最小限度にとどめたい。PF事業者に自社の空間管理のリスクヘッジ及びクオリティの維持・向上を促すために、（第三者による関与を通じて）説明責任を課し、透明性を確保するための施策に注力することが肝要となる。

特に、インターネット関連政策の実効性を担保する

手法として、単なる自主規制にとどまらない、共同規制モデルが着目されるようになって久しい⁽⁴⁰⁾。さらにプラットフォーム・ガバナンスは、単に企業と国家の関係のみで捉えることは適切ではなく、国家、NGO、企業群といった個々のアクターによって規制がデザインされる場合もあれば、国家と企業、企業とNGOが協働して、(自主規制を含めた)規制をデザインする場合もあるため、そうした観点も視野に入れる必要がある⁽⁴¹⁾。またこうしたプラットフォーム・ガバナンスを取り巻く諸アクターの関係性という観点からすれば、「国民の知る権利」⁽⁴²⁾の充実という機能の一環として、PF事業者を「外部」から監視することでその説明責任を高めるメカニズム、通称「プラットフォーム・ビルト」⁽⁴³⁾と呼ばれるデジタル・ジャーナリズムの新領域も注目される。

本稿が取り扱った偽情報の課題は、現代のメディア環境が孕むいくつもの課題のうちの一つに過ぎない。本稿がここまで論じてきた通り、むしろ肝心なのは、民主主義や表現の自由を実現するためのオンライン環境の信頼性をいかにして確保していくか、そのために必要なエコシステム・制度・アーキテクチャのデザイ

ンはどのようなものかといった問いである。憲法及びメディア法という領域は、他の分野(行動経済学、メディア研究、計算社会科学等)と共に、この問いに答えるための一翼を間違いなく担っている。なぜなら、ラジオ・テレビジョンの発展と放送制度の関係性を例にとるまでもなく、情報技術の発展は、それを規律する法制度と共に歩んできたからである⁽⁴⁴⁾。本稿もまた、こうしたメディアを取り巻く古くて新しい問いを検討するうえでの一助となれば幸いである。

〔付記〕 二〇二四年六月二九日(土)に慶應義塾大学三田キャンパス東館八階ホールに於いて開催された慶應法学会二〇二四年度研究大会における報告内容をもとに大幅な加筆修正を施したものである。当日は会場からの質疑応答で鋭い指摘をいくつも頂戴した。この場を借りて厚く御礼申し上げます。また本稿の基になった報告及び本稿は、JSPS 科研費22K13287の助成を受けたものである。

(1) See, Rachel Griffin, *Codes of Conduct in the Digital Services Act, Technology and Regulation*, 167–187, 174 (2024).

(2) See, Corinne Tan, *Regulating disinformation on*

- Twitter and Facebook*, Griffith Law Review, 31:4, 513-536, 516-517 (2022).
- (3) *Kohls v. Bonta*, 2024 WL 4374134 (E.D. Cal. Oct. 2, 2024).
- (4) Singapore statutes online, *Protection from Online Falsehoods and Manipulation Act 2019*, <<https://sso.agc.gov.sg/Act/POFMA2019>> (最終アクセス日二〇二四年一月一四日、以下SURLはすべて同2)。
- (5) Tan, *supra* note 2, at 519.
- (6) See, Jothie Rajah, *Normalising Authoritarianism: Authoritarian Rule of Law*, in *The Palgrave Handbook of Political Norms in Southeast Asia*, 67 (Gabriel Facal Elsa Layate de Micheaux, Astrid Norén-Nilsson ed., 2024).
- (7) 最大判昭和四四年六月二五日刑集第三三卷七号九七五頁。
- (8) 最大判昭和六一年六月一日民集第四〇卷四号八七二頁・谷口裁判官意見。
- (9) コンテンツ・モデレーションの仕組みとAIの関与については、拙稿「AI×メディア×立憲主義」メディア研究一〇五号(二〇二四年)五三―六九頁を参照。
- (10) Meta Transparency Center 「Facebook トラニチー 規定：偽情報」<<https://transparency.meta.com/ja-jp/policies/community-standards/misinformation/>>。なお本稿で引用している各PFのルール内容は、いずれも二〇二四年一月時点のものであることに留意が必要である。
- (11) Xヘルプセンター「市民活動の阻害に関するポリシー」<<https://help.x.com/ja/rules-and-policies/election-integrity-policy>>；Xヘルプセンター「COVID-19のごく誤解を招く情報に関するポリシー」<<https://help.x.com/ja/rules-and-policies/medical-misinformation-policy>>。
- (12) Xヘルプセンター「信頼性」<<https://help.x.com/ja/rules-and-policies/authenticity>>。
- (13) Xトラニチーノートウェブサイト「トラニチーノートガイド：概要」<<https://communitynotes.x.com/guide/ja/about/introduction>>。
- (14) YouTube ヘルプセンター「YouTube のポリシー：選挙の誤った情報に関するポリシー」<https://support.google.com/youtube/answer/10835034?hl=ja&rel_topic=10833358&sjid=12556928393354164261-AP>。
- (15) YouTube ヘルプセンター「YouTube のポリシー：医学的に誤った情報に関するポリシー」<https://support.google.com/youtube/answer/13813322?hl=ja&rel_topic=10833358&sjid=12556928393354164261-AP>。
- (16) YouTube ヘルプセンター「YouTube のポリシー：改変コンテンツまたは合成コンテンツの使用に関する開示」

- <https://support.google.com/youtube/answer/14328491?hl=ja&ref_topic=9282435&sjid=1255692833354164261-AP>.
- (17) TikTok サイバーセキュリティガイドライン：誠実性と信頼性」<<https://www.tiktok.com/community-guidelines/ja/integrity-authenticity>>.
- (18) TikTok 透明性センター「コンテンツガイドラインの施行レポート 二〇二四年一月～三月」(二〇二四年六月二六日) <<https://www.tiktok.com/transparency/ja-jp/community-guidelines-enforcement-2024-1>>.
- (19) Robert Gorwa, Reuben Binns & Christian Katzenbach, *Algorithmic Content Moderation: Technical and Political Challenges in the Automation of Platform Governance*, Big Data & Society 7:1 (2020).
- (20) 水嶋一憲他編著『プラットフォーム資本主義を解説する』(ナカニシヤ出版、二〇二三年) 一四六、一五〇頁。プラットフォーム資本主義については、ニック・スルネック(大橋完太郎・居村匠訳)『プラットフォーム資本主義』(文書院、二〇二二年)を参照。
- (21) Kate Klomick, *The New Governors: The People, Rules, and Processes Governing Online Speech*, 131 Harv. L. Rev. 1598 (2018).
- (22) David Kaye, Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, UN Doc A/HRC/38/35, 18 June–6 July 2018, p. 3. See also, Molly K. Land, *The Problem of Platform Law: Pluralistic Legal Ordering on Social Media*, in *The Oxford Handbook of Global Legal Pluralism*, 975, 978–979 (Paul Schiff Berman ed., 2020). 「プラットフォーム法」について詳しくは、拙稿「プラットフォーム法」から見たヘイトスピーチ対策」メディア法研究二号(二〇二四年) 一三三–一五八頁。
- (23) イーライ・パリサー『フィルターバブル インターネットが隠していること』(早川書房、二〇一六年)。
- (24) ショシヤナ・スボフ(野中香方子訳)『監視資本主義』(東洋経済新報社、二〇二二年)。
- (25) シナン・アラル(夏目大訳)『デマの影響力——なぜデマは真実よりも速く、広く、力強く伝わるのか?』(ダイヤモンド社、二〇二二年) 一–四頁以下を参照。
- (26) 山本龍彦『アテンション・エコノミーのジレンマ』(KADOKAWA、二〇二四年)を参照。
- (27) Carlos Diaz Ruiz, *Disinformation on digital media platforms: A market-shaping approach*, New Media & Society, 00, 18–19 (2023).
- (28) Roee Levy, *Social Media, News Consumption, and Polarization: Evidence from a Field Experiment*,

- American Economic Review, 111:3, 831-70 (2021).
- (29) Andrew M. Guess et al. *How do social media feed algorithms affect attitudes and behavior in an election campaign?* Science, 381, 398-404 (2023). ただし、近時になつてこの論文の基になつた実験期間中に、Meta 社が緊急措置としてアルゴリズムを変更してつたことが eLetter により指摘されている点に注意が必要である。
- (30) 最大判昭和五八年六月二二日民集三七巻五号七九三頁。この考え方は、かねてより筆者が提案してきたメディア環境に対する「構造」的アプローチを基にしている。拙稿「『表現の自由』の現代的な『カタチ』：フェイクニュース問題を切り口に」論究シュリスト三三三号（二〇二〇年）二一三—三〇頁。
- (31) John Bowers, Jonathan Zittrain. *Answering impossible questions: Content governance in age of disinformation*. Harvard Kennedy School (HKS) Misinformation Review, 1:1, 2-5 (2020).
- (32) Beler Center and Shorenstein Center. *Democracy and Internet Governance Initiative: Towards Digital Platforms and Public Purpose*, 25-26 (2023) <https://shorensteincenter.org/wp-content/uploads/2023/07/DIGI-Final-Report_July5_2023.pdf>.
- (33) See, Harsha Bhatlapenumarthi, James Gresham, *Metrics for Content Moderation*. Trust & Safety Professional Association. <https://www.tspa.org/curriculum/ts-fundamentals/content-moderation-and-operations/metrics-for-content-moderation/>. See also, Sebastian Felix Schwemer, *Decision Quality and Errors in Content Moderation*. IIC 55, 139-156 (2024).
- (34) 例えば和久井理子「報道機関とデジタルプラットフォーム：独占禁止法・公正取引委員会の役割と限界、今後の展開」法律時報九六巻五号（二〇二四年）四五—五〇頁、木下昌彦「デジタル・メディア・プラットフォームの憲法理論」情報法制研究九号（二〇二一年）一六—三三頁を参照。
- (35) Journalism Trust Initiative Website. <https://www.journalismtrustinitiative.org/>.
- (36) それ以外に、CIPA News Gaurd、GDI、The Trust Project、オリミネーター、プロファイルゴッド手法が模索されている。
- (37) 拙稿「メディア環境デザイナーとしてのプロミネンス・ルール」Nextcom 二〇二三年五六号、三〇—三七頁、また NHK 受信料制度等検討委員会次世代 NHK に関する専門小委員会「次世代 NHK に関する論点とりまとめ（第二次）報告書」（二〇二三年）を参照。See, Konrad Bleyer-Simon, Eida Brogi, *Enhancing content reliability by*

Prominence: indicators for trustworthy online sources.

[Centre for Media Pluralism and Media Freedom (CMPF)], European Digital Media Observatory (EDMO), (2021).

(38) この点については、欧州メディア自由法一八条及びびりサイタル五三を参照。

(39) なお今般、開催された総務省の「デジタル空間における情報流通の諸課題への対処に関する検討会」には「デジタル広告ワーキンググループ」が設置されている。同WGには筆者も参加しているが、「質の高いデジタル広告の流通に向けた諸課題への対処の在り方」、「質の高いオンラインメディアへの広告配信に向けた諸課題への対処の在り方」、「その他のデジタル広告の流通を巡る諸課題への対処の在り方に関する事項」といった本稿でも取り上げた諸課題を検討することとなっている。

(40) See, Chris Marsden, Trisha Meye, Ian Brown, *Platform values and democratic elections: How can the law regulate digital disinformation?*, Computer Law & Security Review, 36 (2020).

(41) Robert Gorwa, *The platform governance triangle: conceptualising the informal regulation of online content*, Internet Policy Review, 82 (2019).

(42) 最大判昭和四四年一月二六日刑集第二三卷一一号一

四九〇頁。

(43) Philip M. Napoli, *The platform beat: Algorithmic watchdogs in the disinformation age*, European Journal of Communication, 36:4, 376-390 (2021).

(44) アメリカにおけるラジオ及びテレビジョンの発展に、法制度（無線法、連邦無線委員会、通信法、連邦通信委員会）がどのように関与してきたかについては、水越伸「メディアの生成 アメリカ・ラジオの動歴史」（筑摩書房、二〇一三年）を参照。