令和7年度第1回インターネット・モニタリング事業研修
5月28日(水)10:30～12:00
@兵庫県立のじぎく会館

デジタルメディア環境の諸課題と情報流通プラットフォーム対処法

慶應義塾大学メディア・コミュニケーション研究所
准教授　水谷　瑛嗣郎
博士(法学)

Keio University

自己紹介

名前	水谷瑛嗣郎（ミズタニエイジロウ）
経歴	同志社大学法学部を卒業後、慶應義塾大学大学院法学研究科へ。2016年4月より、帝京大学法学部に助教として着任。2019年4月より関西大学社会学部准教授。2025年4月より慶應義塾大学メディア・コミュニケーション研究所に准教授として着任。
専門	博士（法学）。憲法学・メディア法・情報法。インターネットやAIの普及を踏まえ、憲法学の観点からデジタルメディア環境の「デザイン」を考える。
お仕事	日本ファクトチェックセンター運営委員、放送倫理・番組向上機構放送倫理検証委員、総務省誹謗中傷等の違法・有害情報への対策に関するWG、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、同WG、総務省公共放送WG、総務省デジタル広告WGの構成員などを務める。

水谷の研究関心

01 報道の自由の持続可能性

- アテンション・エコノミー
- ピンクスライム・ジャーナリズム
- プロミネンス
- 第三者認証

02 プラットフォーム・ガバナンス

- コンテンツ・モデレーション
- レコメンド・システム
- プラットフォーム法

03 AIガバナンス

- AIのリスクと規制
- 共同統治モデル
- Constitutional AI モデル

04 デジタル広告エコシステム

- ブランド・セーフティ
- 広告の適正化

通底する憲法学的な視座：
表現の自由、報道の自由、個人の尊厳、民主政システム

法学に留まらない情報テクノロジー『デザイン』という発想

目次

１．デジタルメディア環境で生じているリスクの現状

２．「新たな統治者」としてのデジタル・プラットフォーム事業者

３．転換点としての情報流通プラットフォーム対処法

４．情プラ法を踏まえた今後の課題

1．デジタルメディア環境で生じているリスクの現状

➢偽・誤情報のリスク

- 2024年のアメリカ大統領選に際して、7月にトランプ支持を表明したイーロン・マスク氏による政治的投稿は、171億viewを稼いでいる（CCDHによる調査）*。

  →このうち、**87件の投稿がファクトチェック済みの虚偽またはミスリード情報で、それら閲覧数は20億view**に達している。

  ＋**コミュニティ・ノートがついていない****

- 2024年、イギリスにおいては、17歳の少年に子ども3人が刺殺された事件において、容疑者がイスラム教移民であるという偽情報がX上で拡散し、大規模な暴動に発展***。さらにこれを煽る捏造写真をマスク氏がリポスト****。

- 新型コロナ禍におけるワクチン誤情報がワクチン接種率と死亡者数に与える影響*****

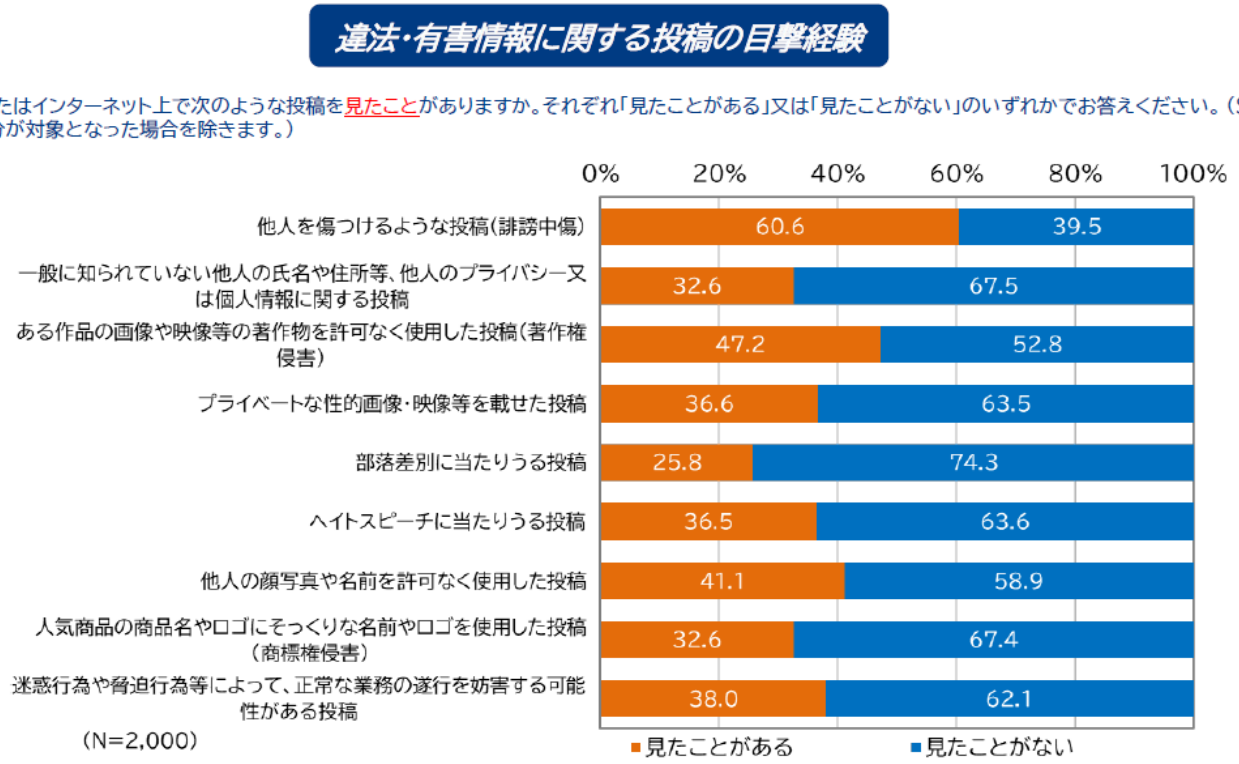
* The Center for Countering Digital Hate, *Musk's Political Posts How Elon Musk's political posts amass more views than all U.S. political campaign ads on X's disclosure dataset*, Nov. 4, 2024, <https://counterhate.com/research/musk-political-posts-x/>.
** コミュニティ・ノートがついていない理由についてはNHKの調査がある。斉藤直哉・岡谷宏基・絹川千晴「イーロン・マスク氏とトランプ次期大統領 急接近でXは？」NHK（2024年11月16日）<https://www3.nhk.or.jp/news/html/20241116/k10014639131000.html>.
*** NHK「偽情報で大規模暴動起きたイギリス SNS運営企業規制強化の方針」（2024年9月7日）<https://www3.nhk.or.jp/news/html/20240907/k10014575741000.html>.
**** BBC Website, Musk shares faked far-right 'detainment camp' for rioters post, Aug. 8, 2024, <https://www.bbc.com/news/articles/cp35w0kj2y4o>.
***** Furuse, Yuki, and Takahiro Tabuchi. "Impact of COVID-19 vaccination by implementation timing and coverage rate in relation to misinformation prevalence in Japan." Vaccine, vol. 59 127273. 20 May. 2025,

1．デジタルメディア環境で生じているリスクの現状

- **違法有害情報に関する目撃経験に関する調査***

→誹謗中傷は60.6%、部落差別は25.8%、ヘイトスピーチは36.5%の人が目撃

➡また目撃したサービスに関しては、誹謗中傷は６６％がX、部落差別投稿も５５．３％がX、ヘイトスピーチも６０．６％がXで、**いずれもXでの目撃が１位（２位はYouTube）****



* 図は、三菱総合研究所「インターネット上の違法・有害情報に関する流通実態アンケート調査」総務省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第２回資料２－１（2024年11月21日）8頁より引用。
** 前掲9頁

1．デジタルメディア環境で生じているリスクの現状

- **プラットフォームのバックラッシュ―Meta社の方針変更***

①サードパーティ・ファクトチェックプログラムの終了、コミュニティ・ノートモデルへの移行
→ファクトチェックは「偏っている」？**

②コンテンツ・モデレーションの緩和（More Speech）
→特にヘイトスピーチに関する規定が大きく緩和されている***

③政治関連コンテンツのパーソナライズ化
→政治関連コンテンツの配信を元に戻し、代わりにパーソナライズを強化する

Meta

More Speech and Fewer Mistakes

January 7, 2025
Joel Kaplan, Chief Global Affairs Officer

* Joel Kaplan, More Speech and Fewer Mistakes, Meta Website, Jan 7. 2025, <https://about.fb.com/news/2025/01/meta-more-speech-fewer-mistakes/>.
** Mosleh, M., Yang, Q., Zaman, T. et al. Differences in misinformation sharing can lead to politically asymmetric sanctions. Nature 634, 609–616 (2024).
*** Kate Knibbs, Meta Now Lets Users Say Gay and Trans People Have ‘Mental Illness’, Wired, Jan 7. 2025, <https://www.wired.com/story/meta-immigration-gender-policies-change/>.

2．「新たな統治者」としてのデジタル・プラットフォーム事業者

- ソーシャルメディアを運営するDPF事業者は、ユーザーに提供している「場」において、情報流通の環境をデザインし、管理する力（＝**情報環境形成力**）を有している。
  →その力は、**いまや国家に匹敵する（か場合によってはそれ以上）。**

- **DPF事業者のコンテンツ管理*は、私たちのユーザー体験に様々な影響を与えている。**

・**ポジティブな管理**：フィード上で、ユーザーの目につきやすい位置にコンテンツを優先表示する。（e.g.レコメンド・システム、トレンド）

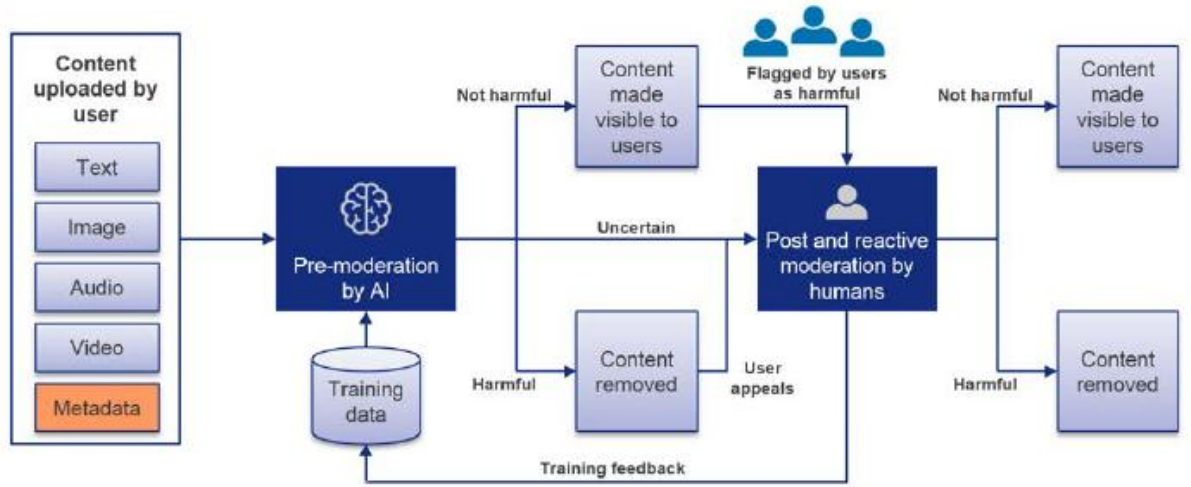
・**ネガティブな管理**：「**コンテンツ・モデレーション**」
※ただし、両者の関係は必ずしも明確に切り分けられるわけではない。

* Tim Wu, *Will Artificial Intelligence Eat the Law? The Rise of Hybrid Social-Ordering Systems*, 119 COLUM. L. REV. 2001, 2014 (2019).

2．「新たな統治者」としてのデジタル・プラットフォーム事業者

コンテンツ・モデレーション：ソーシャルメディアにおいて、ルール（コミュニティ規定等）を定め、違反する主としてUGC及びアカウントを審査し、削除等を行う一連のプロセス

→**<u>ユーザの表現の自由と有害事象のバランシング</u>**



➢ **三つのモデル***

職人モデル、コミュニティ依存モデル、**大規模工場モデル**

➢ **事前モデレーションと事後モデレーション（右上図）****

→AIと人間のモデレーターのハイブリッドシステム

- Metaは、1日当たり1億件のポリシー違反の執行措置を行っているとされる***。

→**わずかなエラー率でも、膨大な数のエラー（偽陽性／偽陰性）が起きる。**

表：TikTokの動画削除数・自動削除数・自動削除割合****

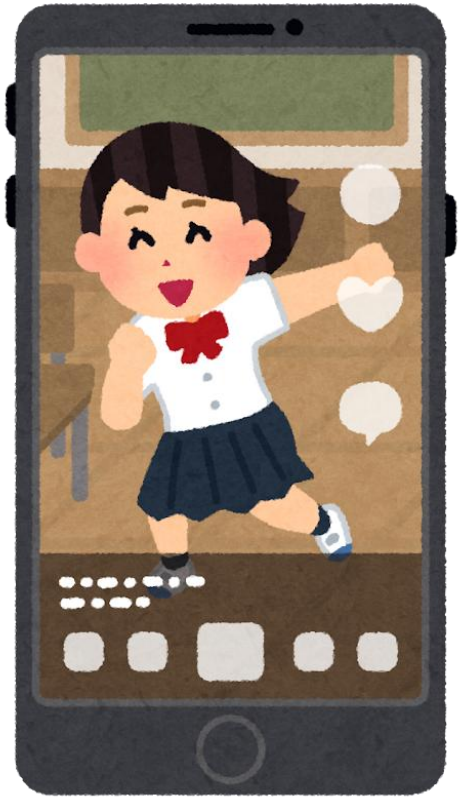
期間	動画削除総数(a)	自動削除数(b)	自動削除割合(b/a)
2021年1～3月	61,951,327	8,832,345	約14.3%
2023年10～12月	176,461,963	128,300,584	約72.7%

* Robyn Caplan, *Content or Context Moderation?: Artisanal, Community-Reliant, and Industrial Approaches*, DATA & SOCIETY REPORT (Nov. 14 2018), < https://datasociety.net/library/content-or-context-moderation/>.
** Cambridge Consultants, *Use of AI in online content moderation*, produced by Ofcom (2019), pp.35 Figure 16, < https://www.ofcom.org.uk/__data/assets/pdf_file/0028/157249/cambridge-consultants-ai-content-moderation.pdf>.
*** *See*, Over Sight Board, PAO-2021-02, < https://oversightboard.com/decision/PAO-NR730OFI/>.
**** 表は「TikTokコミュニティ・ガイドライン実施レポート2023年10月1日～12月31日期（2024年3月15日公開）」の『削除および復元された動画，LIVE配信とアカウントの総数（コンテンツタイプ別）グラフ』（最終取得日2024年5月2日，https://www.tiktok.com/transparency/ja-jp/community-guidelines-enforcement-2023-4/）をもとに筆者が作成。なお自動削除率は，小数点第4位以下を四捨五入して％で表示。

2．「新たな統治者」としてのデジタル・プラットフォーム事業者

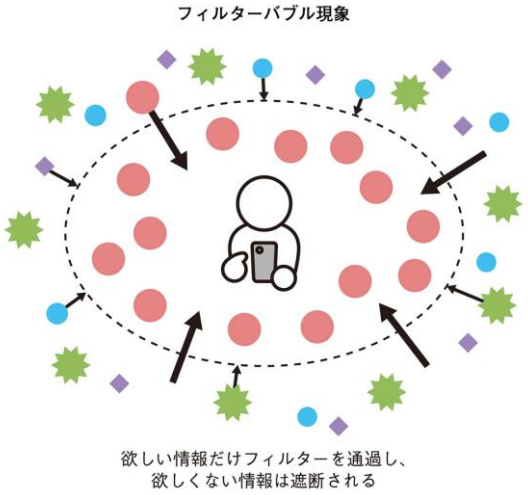
- **Instagramのフィード・アルゴリズム***

→UGCの表示ランクを決める数千種類の「**シグナル**」。投稿がシェアされた時間、使用しているのがスマホかPCか、いいね！ボタンを動画に押す頻度等、ありとあらゆる情報が用いられる。



- **TikTokの「おすすめ」****

「このような様々な要素はすべて当社の**レコメンドシステムによって処理され**、ユーザーにとっての価値に基づいて重要度が決められます。**ユーザーが最初から最後まで長い動画を見終えたかどうかといった強い指標は、ユーザーと動画の視聴者と制作者が同じ国にいるかどうかといった弱い指標よりも重要度が高く判断されます。**」



➢ **フィルターバブル（図左）*****：Facebookで行われた大規模なフィールド実験では、リベラル派と保守派の参加者に、それぞれランダムにニュース・アウトレットの購読を提供したところ、自身の態度に反するニュース（counter-attitudinal news）への接触をソーシャルメディアのアルゴリズムが制限しており、分極化を促進する可能性が指摘されている******。

⇔**分極化などの影響力は、当初の想定よりは緩やかな可能性もあるが…*******

* Adam Mosseri, *Shedding More Light on How Instagram Works*, Instagram Website, Jun. 8, 2021, <https://about.instagram.com/blog/announcements/shedding-more-light-on-how-instagram-works>.
** TikTok「TikTokが「おすすめ」に動画をレコメンドする仕組み」（2020年6月19日）< https://newsroom.tiktok.com/ja-jp/how-tiktok-recommends-videos >
***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、2016年）19頁。図は、鳥海不二夫・山本龍彦『デジタル空間とどう向き合うか 情報的健康の実現をめざして』（日経BP、2022年）78頁より引用。
**** Ro'ee Levy, *Social Media, News Consumption, and Polarization: Evidence from a Field Experiment*, American Economic Review, 111 (3): 831-70 (2021).
***** Andrew M. Guess et al. ,How do social media feed algorithms affect attitudes and behavior in an election campaign?.Science381,398-404(2023).; エコーチェンバーについては、Brendan Nyhan et al., *Like-minded sources on Facebook are prevalent but not polarizing*, Nature, 620, 137–144 (2023).

2．「新たな統治者」としてのデジタル・プラットフォーム事業者

➢ **「関心経済（アテンション・エコノミー）」**

→ユーザーの稀少な「注目」をいかに引きつけるか。

「**デジタル世界での生き残りは、粘着性（stickiness）に左右される**――企業が利用者を引きつけ、長く滞在させ、何度も繰り返し戻ってこさせる能力だ。」*。

- アテンション・エコノミー（関心経済）においては、**コンテンツの中身の正確性、情報源の信頼性、公共性よりも、いかに「脊髄反射」的な「刺激」を与え、プラットフォームに「依存」をさせられるかが重視される**。

  →「**思想内容の競争から刺激の競争**」へシフトするとの指摘**。

- 特にユーザーを「監視」し、多くの情報を集め、分析・適用することでユーザーの粘着性を高め、経済的市場を支配するプラットフォーム企業が、思想の自由市場でも強い力を持つように（**監視資本主義**）。

* マシュー・ハインドマン（山形浩生訳）『デジタルエコノミーの罠　なぜ不平等が生まれ、メディアは衰亡するのか』（NTT出版、2020年）10頁。
** 山本龍彦「思想の自由市場の落日：アテンション・エコノミー×AI」Nextcom 44巻（2020年）10頁。

2．「新たな統治者」としてのデジタル・プラットフォーム事業者

❶コンテンツに対する過少／過剰な管理

⑴コンテンツ・モデレーションの「過少」
⑵コンテンツ・モデレーションの「過剰」
⑶モデレーション・プロセスに対する透明性及び適正手続の不足
＋Meta社の方針変更

❷「誤情報の市場（the marketplace for misinformation）」

誤情報によって利益を得ているのは発信者やDPF事業者だけでなく、**多くのアクターが何らかの形で誤情報から利益を得る構造がある**との指摘*。

➡アテンション・エコノミーの弊害

❸構造上のシステミックリスク

「パーフェクト・スロットマシーン」としてのTikTok**、レコメンドシステム・シェアボタンと誤情報の拡散***

* Carlos Diaz Ruiz, *Disinformation on digital media platforms: A market-shaping approach*, New Media & Society, 0:0, 18(2023).
** Annie Goldsmith, *'The Perfect Slot Machine': TikTok's Most Addictive Design Features Are Being Cloned Across the Internet*, The Information, April 15, 2022, < https://www.theinformation.com/articles/the-perfect-slot-machine-tiktoks-most-addictive-design-features-are-being-cloned-across-the-internet>.
*** Jeff Allen, Misinformation Amplification Analysis and Tracking Dashboard, Integrity Institute, Oct. 13, 2022, <https://integrityinstitute.org/blog/misinformation-amplification-tracking-dashboard>.

2．「新たな統治者」としてのデジタル・プラットフォーム事業者

➢DPF事業者は、「そのデジタル空間ガバナンスによって、公的機関の監視に頼ることなく、自律的な準公的機能をも果たしている」という指摘*

- lex mercatoriaと並ぶlex digitalis**

→「**プラットフォーム法**」：①契約法、②実体法、③手続法、④技術法という４つの要素から構成されている***。

形態	説明	具体例
契約法	利用者と企業のそれぞれの権利と義務を規定する契約条項	利用規約
実体法	予測される規範、手順、制裁に関するコミュニケーション	コミュニティ基準や規則・ユーザーに送られた警告、通知、アクションの説明等
	コンテンツ・モデレーターとユーザー・アカウントの慣行（判決法）	どのようなコンテンツを削除または許可するか・いつ、どのようにユーザーに制裁を加えるかについての決定・不服申し立てに対する決定等
	コンテンツをモデレートし、ユーザーアカウントの停止および終了を管理するために使用される内部ガイダンスおよびポリシー	内部ポリシー及び文書・モデレーターの研修資料・会議の記録
手続法	コンテンツのフラグ立て、審査、削除、およびユーザーへの制裁の手順	ユーザーまたはアルゴリズムによるフラグ立て・モデレーターによる審査・エスカレーションの手順と基準・制裁の手順と基準・不服申し立てプロセス等
技術法	技術上およびシステム設計上の選択	アルゴリズムによる選好メカニズム・コンテンツを共有するためのツール（リツイート、シェア）・ユーザーがコンテンツをコントロールできるツール等

* Giovanni De Gregorio, Digital Constitutionalism in Europe: Reframing Rights and Powers in the Algorithmic Society 82 (2022).
** Andreas Fischer-Lescano & Gunther Teubner, *Regime-Collisions: The Vain Search for Legal Unity in the Fragmentation of Global Law*, 25 MICH. J. INT'L L. 999, 1010-11 (2004).
*** 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).

2．「新たな統治者」としてのデジタル・プラットフォーム事業者

- **DPF事業者に対するガバナンスを考えるうえで重要なプラットフォーム・ポリティクス**

①経済市場（ユーザー、広告主）
②政府関与
③社会規範
④上位DPF（GoogleとApple）

- ただし、市場や社会規範を通じたガバナンスの「不安定」さに注意が必要。
  →市場や社会規範が常に公正なガバナンスを提供するわけではない
  ＝**政府関与または民主的手続きの「不足」**

- **「新たな統治者」（ケイト・クロニック）あるいは「ビヒモス」（山本龍彦）として、DPF事業者に対する憲法的統御の法理を構築する必要性。**

- 重要なのは、**実現すべき立憲的価値（自由かつ多様な情報流通）のために、“誰に”、“どのような責務を”、“どのような方法で”、負わせるべきか。そして、そのために「（国家）法」はどのような役割を果たすべきか、という視点。**

3．転換点としての情報流通プラットフォーム対処法

➢ 日本の転換点としての「情報流通プラットフォーム対処法」＝「日本版DSA」*

同法第一条　この法律は、特定電気通信による情報の流通によって権利の侵害等があった場合について、特定電気通信役務提供者の損害賠償責任の制限及び発信者情報の開示を請求する権利について定めるとともに、発信者情報開示命令事件に関する裁判手続に関し必要な事項を定め、あわせて、**侵害情報送信防止措置の実施手続の迅速化及び送信防止措置の実施状況の透明化を図るための大規模特定電気通信役務提供者の義務**について定めるものとする。

情報流通プラットフォーム対処法全体の構造	
第二章　損害賠償責任の制限	「プロバイダ」の作為・不作為に関する損害賠償の免責規定
第三章　発信者情報の開示請求等	発信者情報の開示請求権
第四章　発信者情報開示命令事件に関する裁判手続	発信者情報開示に関する「新たな裁判手続き」として、改正法（令和3年法律第27号）が令和3年4月28日に公布され、令和4年10月1日に施行された。
第五章　大規模特定電気通信役務提供者の義務　←New！	ソーシャルメディア等の大規模DPF事業者を対象とした日本で初めての規律として、令和6年5月10日に成立。令和7年4月1日施行。

* 衆議院総務委員会・案件「特定電気通信役務提供者の損害賠償責任の制限及び発信者情報の開示に関する法律の一部を改正する法律案（213国会閣34・212国会衆15）」（2024年4月18日）内の平林晃（公明党）委員の発言 <https://www.shugiintv.go.jp/jp/index.php?ex=VL&deli_id=55161>.

3．転換点としての情報流通プラットフォーム対処法

- 旧法の民事上の事業者免責規定及び発信者情報開示の手続きを維持したうえで、総務大臣が指定する「**大規模特定電気通信役務提供者**」に、大別して二つの公法上の義務を課す。
  →(1)**権利侵害情報への迅速化規律**、(2)**送信防止措置に対する透明化規律**

権利侵害情報への迅速化規律	**送信防止措置に対する透明化規律**
被害者の申し出窓口の設定と公表 権利侵害有無に関する必要な調査の実施 **侵害情報調査専門員の選任** 14日以内の総務省令で設定された期間内の判断と申出者への通知（**迅速化義務**） 例外としての申出者への通知（**誠実対応義務**）	**コンテンツ・モデレーションの基準公表義務** モデレーション基準の規定に関する（努力）義務 基準変更義務 コンテンツ・モデレーションに関する資料作成・公表の（努力）義務 発信者への通知義務 **運用状況公表義務**
総務大臣に、報告徴収権限、規定違反への勧告・命令権限を与え、**法人等による命令違反には、最大で1億円の罰金**	

3．転換点としての情報流通プラットフォーム対処法

➢ 情プラ法を読み解くうえで重要な二つのガイドライン

(1)「特定電気通信による情報の流通によって発生する権利侵害等への対処に関する法律における大規模特定電気通信役務提供者の義務に関するガイドライン」（令和７年３月11日制定）
　→ここでは「**大規模DPF事業者義務ガイドライン**」と呼称

(2)「特定電気通信による情報の流通によって発生する権利侵害等への対処に関する法律第26条に関するガイドライン」（令和７年３月11日制定）
　→ここでは「**26条ガイドライン**」と呼称

3．転換点としての情報流通プラットフォーム対処法

➢ 対象事業者と総務大臣の指定権限

- 以下1号～3号のいずれにも該当し、総務大臣により「侵害情報送信防止措置の実施手続の迅速化及び送信防止措置の実施状況の透明化を図る必要性が特に高いと認められる」サービスを提供する事業者は、「**大規模特定電気通信役務提供者**」（以下、**大規模DPF事業者**）として指定を受ける（21条1項）。

→事業者のユーザー（発信者）規模（同１号）、侵害情報送信防止措置が「技術的に可能」であること（同２号）、総務省令で定められる「権利の侵害が発生するおそれの少ない特定電気通信役務」ではないこと（同３号）

➡**Google、Meta、X（Twitter）、LINEヤフー、TikTokが指定された**。

3．転換点としての情報流通プラットフォーム対処法

- **そもそも迅速化規律の対象となる「侵害情報**（2条6号：特定電気通信による情報の流通によって自己の権利を侵害されたとする者が当該権利を侵害したとする情報）」**とは？**

→名誉権、名誉感情侵害、プライバシー、私生活上の平穏、肖像権、氏名権、パブリシティ権、著作権・著作隣接権、商標権、営業上の利益、その他

- このうち、**部落差別に関連する「差別されない権利」*は、私生活上の平穏に該当する**（26条ガイドライン5,28-29頁）。

→「差別を引き起こすおそれのある行為にまで違法性を認めることになれば、プライバシーや名誉を保護法益とする訴訟よりも、より広範な差別行為を規制対象に含めることができ」る点が「最大の意義」**。

* 東京高判令和５年６月28日判タ1523号143頁、併せて最三小決令和6年12月4日判例集未搭載も参照。
** 金子匡良「判批：被差別部落の所在地を出版物等で公開することの違法性」新・判例解説Watch 憲法No.243（2025年3月28日）4頁。

3．転換点としての情報流通プラットフォーム対処法

➢ 権利侵害情報（侵害情報送信防止措置）に対する迅速化規律

22条1項：ユーザーのうち「**被侵害者**」（「自己の権利を侵害されたとする者」）が「侵害情報送信防止措置を講ずるよう申出を行うための方法」を定めて公表する義務（**被害者の申し出窓口の設定と公表義務**）

→①オンラインで申出ができる、②申出者に「**過重な負担**」を課さないこと、③申出受付日時が申出者に明らかになるようにすること（同2項各号）

➡当該窓口は「被侵害者が日本語による申出を行うことができる」こと、そして「インターネットを利用」して公表されていなければならない（施行規則13条1項・2項）

「過度な負担」を課さないことの例示（大規模DPF事業者義務ガイドライン3頁）
トップページから少ないクリック数でアクセスできる等、申出フォームが見つけやすいこと。
文字制限のない文章記入欄が設けられている、証拠が添付可能である等、十分に情報提供が可能な申出フォームとなっていること。
アカウント非保有者であっても申出を行うことができること。
申出先以外の第三者との関係で、申出者のプライバシー等の権利・利益の侵害を生じさせない形で、申出を行うことができること。
申出を行ったことを理由として、申出以後のサービス利用に当たって不利益を受けないこと。

3．転換点としての情報流通プラットフォーム対処法

➢ **権利侵害情報（侵害情報送信防止措置）に対する迅速化規律**

23条：22条1項の申出に対して「被侵害者の権利が不当に侵害されているかどうか」に関する調査義務（**権利侵害有無に関する必要な調査の実施**）

24条：23条の調査に際して、「専門的な知識経験を必要とするものを適正に行わせるため」の「**侵害情報調査専門員**」の選任義務

→指定DPF事業者のユーザー規模等に応じて総務省令が定めた数以上の専門員の選任義務（同2項）、選任後の届出（同3項）。

➡「**大規模特定電気通信役務ごとに一人**」（施行規則14条）

- 専門員は、「削除等のコンテンツモデレーションを実施する他の職員が判断に迷った際に、当該職員からの上申を受けて、より専門的な調査を行う」任務を帯びる。
  →そのため、「法令の知識又は文化的・社会的背景の理解の観点から、**弁護士等の法律専門家や日本の風俗・社会問題に十分な知識経験を有する者**（自然人に限る。）が考えられる」（大規模DPF事業者義務ガイドライン3-4頁）。

3．転換点としての情報流通プラットフォーム対処法

25条1項：指定DPF事業者は、23条の申出に対して、**侵害情報送信防止措置を行うかを判断して、**「申出を受けた日から十四日以内の総務省令で定める期間内」に、**申出者に対する通知を行う義務**

→措置を行わなかった場合には、その旨と「理由」を通知（同2号）

➡期間については、「**7日**」（施行規則16条）となった*。

25条2項（1項の例外規定）**前段：発信者の意見聴取（同1号）、専門員の調査（同2号）、その他「やむを得ない理由」がある（同3号）といったいずれかの場合**には、23条調査の結果に基づいて送信防止措置を行うかどうかを判断後、「遅滞なく」、25条1項の通知を行う義務

25条2項後段：上記の場合には、**それら理由（1号～3号に該当するのか）を「7日」いかにに申出者に対して通知を行う義務**

➡**誠実に対応する義務**

* この点については、総務省プラットフォームサービスに関する研究会「第三次とりまとめ」（令和6年1月）11頁。アンケート結果と、誹謗中傷の事案ではないものの、具体的な日数を示した裁判例として、楽天チュッパチャプス事件（知財高判平成 24 年 2月 14 日判タ 1404 号 217 頁）が参考とされている。

3．転換点としての情報流通プラットフォーム対処法

➢ 送信防止措置に対する透明化規律

→権利侵害情報の送信防止措置のみを対象としている迅速化規律と異なり、**大規模DPF事業者が自主的に設定している基準に基づくコンテンツ・モデレーションも射程に入れた規律**

26条1項：送信防止措置の基準公表義務

→法律の建付けでは、下記の1号～3号に加え、「自ら定め、公表している基準に従う**場合に限り、送信防止措置を講ずることができる**」となっている。

- 事業者自身が「送信防止措置を講じようとする情報の発信者」である場合（同1号）、**法令上の義務に基づいて送信防止措置を行う場合（同2号）**、緊急の必要性（同3号）

同2項：モデレーション基準についても、①送信防止措置対象情報の具体性、②サービス停止措置に関する具体的な基準、③理解容易な表現、④送信防止措置実施の努力義務を課す法令との整合性に適合的なものにする努力義務（1～4号）
同3項：1項3号の緊急の必要性から送信防止措置を行った場合には、「速やかに」基準の変更を行う義務
同4項：送信防止措置の事例集を年一回公表する努力義務

3．転換点としての情報流通プラットフォーム対処法

27条：送信防止措置を行った場合に、**発信者への通知等の義務**
＋送信防止措置と基準との関係を明示する必要性
→例外として、①事業者自身が「送信防止措置を講じようとする情報の発信者」である場合、②正当な理由がある場合（e.g. 繰り返し同じ理由で送信防止措置をされているユーザーの場合等）

28条：「毎年一回、総務省令で定めるところにより、次に掲げる事項を公表しなければならない」。
　→**運用状況の公表義務**
　①被侵害者からの申出、②申出者への通知
　③発信者への通知、④送信防止措置の実施状況
　⑤①～④についての自己評価
　⑥そのほか総務省令で定める事項

➡いわゆる「**透明性レポート**」**の公表義務。**

3．転換点としての情報流通プラットフォーム対処法

➢エンフォースメント・罰則

29条：第22条、第24条、第25条、第26条第１項若しくは第３項、第２7条又は28条に関する報告徴収権限

30条1項・2項：第22条、第24条、第25条、第26条第1項若しくは第３項、第27条又は第28条の違反に対する勧告・命令権限

➡総務大臣に報告徴収権限、勧告・命令権限の付与

35条：命令（30条2項）違反への罰則（一年以下の拘禁刑又は百万円以下の罰金）

36条：21条届出義務違反と29条報告徴収義務違反への罰則（50万円以下の罰金）

37条：法人等への罰則

①総務大臣の命令違反　or　21条届出義務違反　1億円以下の罰金

②29条報告徴収義務違反　50万円以下の罰金

38条：その他の罰金　30万円以下の過料

3．転換点としての情報流通プラットフォーム対処法

➢ 情プラ法の迅速化規律と透明化規律の関係性

- 情プラ法は「迅速化」によりDPF事業者に違法（権利侵害）情報への対応を迫り、他方でオーバーブロッキングをはじめとする問題を引き起こしていないかを「透明化」によりモニタリングするという**「両輪」または「コインの裏表」として作用***。

➢ 情プラ法による二つの転換

①旧プロ責法における個々の権利侵害に媒介する事業者（プロバイダ）にどこまで責任を負わせるべきかという観点から、大規模DPF事業者が負うべき社会的責務を公法上の義務として課すという観点への転換（cf. 放送制度）

→ただしプロバイダ免責規定はそのまま残されている。

②加えて情プラ法は、「**民間企業の構築する言論空間を含めて、アイデアの市場の最低限の管理を行う憲法上の責務が国家には存在する**」というフェーズへの転換**。

→もっとも、「**この責務（義務）がどれくらい具体的なものと考えるか、また、市場管理を（抽象的であれ）国家の義務とみるか、正当化されうる政府利益の一つに過ぎないとみるか**」が論点となり得る***。

* 総務省・誹謗中傷等の違法・有害情報への対策に関するワーキンググループ「議事概要：第11回（令和５年11月21日開催）」17頁（水谷発言及び事務局発言）、23頁（宍戸主査発言）を参照< https://www.soumu.go.jp/main_content/000917279.pdf >.

** 梶原健佑「思想の自由市場論の新展開」法律時報96巻5号18頁。

*** 前掲18頁脚注28。

4．情プラ法を踏まえた今後の課題

➢ **情プラ法に関する重要ポイント**

①24条の侵害情報調査専門員の存在

→申出⇒即削除、ではなく大規模DPF事業者に権利侵害に関する調査義務（23条：「遅滞なく必要な調査を行わなければならない」）が課せられている。

➡コンテンツ・モデレーションにおいては、いかに「**ローカライゼーション**」を実装するかが課題となっている（e.g. ヘイトスピーチ対策、Meta監督委員会）

＋日本におけるヘイトスピーチ及び「差別されない権利」の固有性

「現にインターネット上における識別情報の摘示を中心とする部落差別の事案は増加傾向にあること……等に鑑みると、**本件地域の出身等であること及びこれを推知させる情報が公表され、一般に広く流通することは、一定の者にとっては、実際に不当な扱いを受けるに至らなくても、これに対する不安感を抱き、ときにそのおそれに怯えるなどして日常生活を送ることを余儀なくされ、これにより平穏な生活を侵害されることになる**」*。

* 東京高判令和５年６月28日判タ1523号143頁

4．情プラ法を踏まえた今後の課題

➢ **情プラ法に関する重要ポイント**

②25条の迅速化規律は、迅速に削除する義務、ではない

→大規模DPF事業者に対し、**侵害情報への対処に対する「誠実な対応」を求めたもの**

＝申出に対する対応状況を申出者に誠実に説明する仕組み

③26条の基準公表の義務

→当該基準は送信防止措置を行う「**14日**」（施行規則17条）前までに公表しておかなければならない。

➡これは大規模DPF事業者による恣意的な送信防止措置を禁止したものと理解できる

4．情プラ法を踏まえた今後の課題

➢ 情プラ法に関する重要ポイント

④26条ガイドラインの列挙事項

→26条1項2号にどのような情報が該当するかを示したものだが、ここには、**大規模DPF事業者が「送信防止措置」を講ずる義務が生じるものが例示列挙されているという点が重要**。

＋しかも権利侵害情報以外の法令違反情報も列挙されている。これらについては、「**自ら提供するプラットフォーム上で特定の法令違反情報が流通していることを認識しつつも放置していた場合であって、当該不作為が作為により当該情報の流通に関与した場合と同視し得るときは、当該プラットフォーム事業者等は、当該情報を削除する等、送信を防止する措置を講ずる刑事法上の義務（作為義務）を負い、その義務違反について刑事責任を追及され得る**」（同ガイドライン25頁）とされている。

他人の権利を不当に侵害する情報の送信を防止する義務	その他送信防止措置を講ずる法令上の義務
名誉権、名誉感情侵害、プライバシー、私生活上の平穏、肖像権、氏名権、パブリシティ権、著作権・著作隣接権、商標権、営業上の利益、その他	わいせつ電磁的記録記録媒体公然陳列罪、児童ポルノ、売春防止法違反広告、出会い系サイト規正法上の児童誘引等情報、リベンジポルノ、性的姿態撮影等処罰法に基づく卑わいな盗撮画像等、規制薬物・指定薬物・指定薬物等の疑いがある物品に関する広告、薬物犯罪の実行・規制薬物の濫用を「公然、あおり、又は唆す」情報、未承認医薬品広告、医薬品等の虚偽・誇大広告、振り込め詐欺関連情報、闇バイト関連情報、金融業関係情報、消費者取引関連情報、銃刀法関連情報、ストーカー規制、不正サクセス助長行為に関する情報など

4．情プラ法を踏まえた今後の課題

➢ 情プラ法に関する重要ポイント

⑤28条の透明性レポート規定

- １～２号で挙げられている公表事項は迅速化規律の実施状況（23条、25条）を公表させるものであり、事業者が同規律に**適切に対応しているか（または過剰に対応していないか）を客観的にモニタリングするために重要**

＋また透明性レポートは各社すでに自主的に公表しているが、本規定によって**日本の文脈で公表させる**ということの意義を考えるならば、３～６号の意味も大きい。

➡特に4号が広範にコンテンツ・モデレーションの状況について、その体制も含めて公表させる（下記）。また5号に挙げられている「**自ら行った評価**」は、自身の運用状況に対する（部分的ではあるが）**セルフアセスメントの公表義務**にもなっている（But…アセスメント自体の義務ではないことに注意）。

「送信防止措置を講ずるための人的体制及び技術的措置についての定性的又は定量的な説明」に含むもの（大規模DPF事業者の義務に関するガイドライン9頁）
送信防止措置の実施に当たっての専門員の役割及び実績
送信防止措置の実施におけるＡＩ等の自動化手段の活用状況
ＡＩ等の自動化手段による送信防止措置の正確性向上のための取組
送信防止措置の対象となる情報（権利侵害情報、法令違反情報、その他削除基準に掲げられている情報）を投稿するボットの検出技術

4．情プラ法を踏まえた今後の課題

➢ **情プラ法の課題**

⑴ユーザーの表現の自由との兼ね合い

→この法律は、相応の社会的責務を果たすべき大規模DPF事業者に対して、**表現内容に踏み込まず、手続き面からの統制を図っている（＝新たに違法な表現のカテゴリーが拡張したわけではない）**

⑵DPF事業者のコンテンツ管理権と表現の自由

→DPF事業者のコンテンツ管理権を、報道機関の「編集の自由」と同じように考えるべきか？

- 「**国民の『知る権利』に奉仕する**」報道機関*
- 「**表現行為としての側面を有する**」、「現代社会においてインターネット上の情報流通の基盤として大きな役割を果たしている」検索エンジン**
- 「その利用者に対し、情報発信の場やツイートの中から必要な情報を入手する手段を提供するなどしている」SNS***

* 最大決昭和44年11月26日刑集23巻11号1490頁。
** 最決平成29年1月31民集71巻1号63頁。
*** 最判令和4年6月24日裁時1794号48頁。

4．情プラ法を踏まえた今後の課題

⑶迅速な削除とエラーのトレードオフ

→違法・有害情報に対して迅速な削除を迫れば迫るほど、エラー（過剰削除）は起こり得る。必要なのは、**過剰削除を許容できるコンテンツのカテゴリは、どのようなものかについてのオープンな議論**（例：CSAM、リベンジポルノ）

- **「選挙に関する各党協議会」における論点案**

「自民の論点案は、ＳＮＳ上で選挙関連のコンテンツの発信者が大きな収益を得ている現状について、『供託金による候補者乱立の抑止効果が薄れ、選挙運動のボランティア原則との齟齬も生じている』と問題視。**偽情報が拡散しても現行の情プラ法には即日差し止めの規定がないとして、同法改正を提案した**」*。

- **候補者や後援会が、選挙期間中に「口を塞がれ」、SNSから「追い出される」ことのリスク（cf. 兵庫県知事選時の稲村候補後援会アカウント凍結事例）。**

→**即日対応など、情プラ法の迅速化規律をさらに強化するような仕組み（e.g. notice and take down）にしてしまうと、「ハック」される危険性も**。

* 読売新聞オンライン「選挙のＳＮＳ規制、自民が論点案提示…選挙運動名目の営利行為抑止へ情プラ法改正など」（2025年2月6日）<https://www.yomiuri.co.jp/politics/20250206-OYT1T50013/>.

** Joel Kaplan, More Speech and Fewer Mistakes, Meta Website, Jan 7. 2025, <https://about.fb.com/news/2025/01/meta-more-speech-fewer-mistakes/>.

4．情プラ法を踏まえた今後の課題

➢ **発想の転換が必要**

「個別の判断を規律する試みによりコンテンツ・モデレーションをガバナンスすることは、沈みゆく船から水をかき出すためにティースプーンを使う」ようなもの*。

➡権利侵害や法益侵害に対する事後的な個別対応モデルの限界（≠不要）

- **予防的、事前規制モデルの検討**

→リスクアセスメントを通じた透明性の強化、コンテンツ・モデレーションAIのローカライゼーション、**選挙期間中のコンテンツ・モデレーションの体制づくり、インシデントが起きた際の事業者側の対応・報告義務など**

- **エラー率の国・地域ごとの偏りに関する指摘****

→コンテンツ・モデレーション・メトリクスに基づく「**質的評価**」の必要性***　Cf. デジタルサービス法15条1項(e)

* Evelyn Douek, *Content Moderation as Systems Thinking*, 136 HARV. L. REV. 526, 606 (2022).
** Dia Kayyali, An Advocate's Guide to Automated Content Moderation, Tech Policy Press, Feb. 13, 2025, <https://www.techpolicy.press/an-advocates-guide-to-automated-content-moderation/>.
*** *See*, Harsha Bhatlapenumarthy, 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/>.

ご清聴ありがとうございました。

謝辞:本報告は、JST、RISTEX、 JPMJRS23L2及びJSPS科研費22K13287の支援を受けたものです。

参考資料：情プラ法の規定

➢参考資料：28条透明性レポートの記載事項（施行規則18条２～７項）

1号項目（2項）	2号項目（3項）	3号項目（4項）
「各公表年度における申出を受け付けた件数」	・送信防止措置を講じた際の通知件数（1号） ・送信防止措置を講じなかった際の通知件数（2号） ・通知をしなかった場合の理由（3号） ・25条2項前段に基づき、送信防止措置を講じた際の通知件数（4号） ・25条2項前段に基づき、送信防止措置を講じなかった際の通知件数（5号） ・25条2項後段による、同項1号の通知件数（6号） ・25条2項後段による、同項2号の通知件数（7号） ・25条2項後段による、同項3号の通知件数（8号）	・「発信者に通知等の措置を講じた件数」（1号） ・「通知等の措置を講じなかった場合」の理由（2号）

参考資料：情プラ法の規定

4号項目（5項）

➢ 「各公表年度における**日本の利用者に関する送信防止措置の実施状況**であって、**日本の利用者に関する**」以下の事項

・「延べ発信者数の総数又は役務提供停止措置の対象となる情報の発信者の総数及びその具体的な算定方法」（1号）
・ユーザー通報による送信防止措置を講じたor講じなかった件数（2号）
・「自ら探知して送信防止措置を講じた件数」（3号）
・裁判所を除く公的機関からの送信防止措置の要請件数（4号）
・前号のうち、送信防止措置を講じたor講じなかった件数（5号）
・裁判所からの送信防止措置を講ずる判決・決定の件数（6号）
・前号のうち、送信防止措置を講じたor講じなかった件数（7号）
・役務提供（アカウント）停止措置の件数（8号）
・AIによる送信防止措置件数（9号）
・AIによる役務提供（アカウント）停止措置の件数（10号）
・送信防止措置に対する不服申立て件数（11号）
・前号のうちAIによる送信防止措置に対する不服申立て件数（12号）
・不服申し立て件数のうち、送信防止措置を撤回した件数（13号）
・前号のうちAIによる送信防止措置を撤回した件数（14号）
・役務提供（アカウント）停止措置に対する不服申立て件数（15号）
・前号のうち、AIによる役務提供（アカウント）停止措置に対する不服申立て件数（16号）
・役務提供（アカウント）停止の件数のうち、不服申立てを受けて撤回した件数（17号）
・前号のうちAIによる役務提供（アカウント）停止措置を撤回した件数（18号）
・専門員の専門性及び訓練内容（19号）
・モデレーターのうち日本語理解者の数及び訓練内容（20号）
・コンテンツモデレーションの人的体制及び技術的措置に関する説明（21号）
・基準変更により送信防止措置の対象となることが明らかになった情報の種類（22号）
・事例集の公表方法（23号）

参考資料：情プラ法の規定

5号項目（6項）	6号項目（7項）
➢ 自己評価にあたっては、以下の事項について「評価の基準を定めて行うものとする」。 ・申出窓口の方法（1号） ・25条の申出に基づく「侵害情報送信防止措置の実施状況」（2号） ・侵害情報送信防止措置を行わなかった場合の理由、25条2項3号の「やむを得ない理由」、阪神者に対する通知が「個別的かつ具体的なものであるか」（3号） ・「日本の利用者に関する送信防止措置を講ずるための人的体制及び技術的措置の整備」（4号） ・「送信防止措置の実施に関する基準の内容」（5号） ・25条の申出の場合を除く、「日本の利用者に関する送信防止措置の実施状況」（6号） ・権利侵害情報、法令違反情報、事業者の基準に基づき送信防止措置の対象となる情報の流通状況（7号）	・自己評価の基準