

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 第四回

2023年12月15日（金）10:00～12:00

於・オンライン

**デジタル空間の健全化とプラットフォーム・ガバナンス
—アメリカの状況を参考に—**

関西大学社会学部メディア専攻 准教授
水谷瑛嗣郎（ミズタニ エイジロウ）
博士（法学）

目次

I. DPF事業者の社会的機能

II. DPF事業者規制をめぐるアメリカの動向

III. プラットフォーム・ガバナンスと透明性

今後の課題

参考資料



I. DPF事業者の社会的機能

▶ デジタル・プラットフォーム（DPF）事業者は、デジタルメディア環境において、私たちユーザーの発信、閲覧、受領において非常に大きな力（**情報環境形成機能**）を有している*。

(1) アーキテクチャ

→例：Twitterの字数制限、スロットマシーンのデザイン

(2) アルゴリズム

→例：レコメンドシステム（パーソナライゼーション）、エンゲージメント強化システム（トレンドの仕組み）

(3) モデレーション

→「インターネット企業が、ユーザー生成コンテンツが利用規約やその他の規則で明確に規定された基準を満たしているかどうかを判断するプロセス」**

→ただの情報媒介者にとどまらない、デジタルメディア環境の「新たな統治者」*または「ビヒモス」****（=国家そのものではないが、国家に匹敵するアクター）**

* 拙稿「ソーシャルメディア・プラットフォームに対する憲法的ガバナンスの展望」法とコンピュータ41巻（2023年）33頁以下を参照

** 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, pp. 3 footnote 2.

*** See, Kate Klonick, *The New Governors: The People, Rules, and Processes Governing Online Speech*, 131 Harv. L. Rev. 1598 (2018).

**** 山本龍彦「近代主権国家とデジタル・プラットフォーム——リヴァイアサン対ビヒモス——」山元一編『講座 立憲主義と憲法学〔第1巻〕』（信山社、2022年）164頁。



I. DPF事業者の社会的機能

- **コンテンツ・モデレーションの現在地**（参考資料①・②）
 - アルゴリズムック・コンテンツ・モデレーション*
 - 個人主義的な権利ベースの世界ではなく **公衆衛生的世界****
 - 膨大な量のエラー（偽陽性／偽陰性）
 - コンテンツ・モデレーションは、「**品質管理**」であり、ポリシーは「品質決定指針」として機能。
 - 公的機関の介入とユーザーの大量離脱が脅威となるため、その「**防御策**」としてコンテンツ管理を行う***
 - 背景にある **市場（ユーザー、広告主）、法、上位DPF（GoogleとApple）の存在******
 - 2018年以降には「**ビッグデック改革**」の「**黄金時代**」が到来
 ⇔ 昨今の経済不況による自主的取組みの後退*****
- ➔ **ユーザー生成コンテンツの管理を行うAIと人のハイブリッドシステム*******



* See, Robert Gorwa, Reuben Binns & Christian Katzenbach, *Algorithmic Content Moderation: Technical and Political Challenges in the Automation of Platform Governance*, Big Data & Soc'y, Jan.–June 2020.

** See, Jonathan Zittrain, *Three Eras of Digital Governance* (Sept 23, 2019) < <https://ssrn.com/abstract=3458435> >.

*** ショシャナ・ズボフ（野中香方子訳）『監視資本主義』（東洋経済新報社、2021年）参照。

**** See, Yoel Roth, *I Was the Head of Trust and Safety at Twitter. This Is What Could Become of It.*, The New York Times, Nov. 18, 2022, <<https://www.nytimes.com/2022/11/18/opinion/twitter-joel-roth-elon-musk.html>>.

***** Kate Klonick, *The End of the Golden Age of Tech Accountability*, The Klionicles, Mar. 4, 2023, <<https://klonick.substack.com/p/the-end-of-the-golden-age-of-tech>>.

***** See, Tim Wu, *Will Artificial Intelligence Eat the Law? The Rise of Hybrid Social-Ordering Systems*, 119 COLUM. L. REV. 2001, 2021 (2019).

II. DPF事業者規制をめぐるアメリカの動向

➤ アメリカにおけるデジタル空間に関する規制

→ 基本的には、**自由放任（レッセフェール）**

➔ 憲法修正一条の強力な保護があるために、偽情報の発信者に制裁を加える法規制もさることながら、DPF事業者に対する法規制を行うことも「ほぼ」不可能（生成系AIについては参考資料③）。

- COVID-19禍における虚偽医療情報流布に関するカリフォルニア州の規制*については、連邦裁判所(McDonald v. Lawson)にて係争中だったが10月1日に廃止**。

➤ 特にアメリカ憲法は、EUと異なり、**憲法はあくまで国家権力の統制に用いられるもの**であり、私的アクターであるDPF事業者に対しては「**ステイト・アクション**」の法理を充たさなければ修正一条は適用されない***。

+ ①表現の自由の過剰な保護と偽情報対策を阻む強力な先例の存在****、②通信品位法230条の存在

→ もちろんこうした傾向には、修正一条の「**武器化（Weaponizing）**」であるといった批判も寄せられる*****。



* CA AB-2098 Physicians and surgeons: unprofessional conduct (2021-2022) <https://leginfo.ca.gov/faces/billTextClient.xhtml?bill_id=202120220AB2098>. See also, Allison M. Whelan, *How Should State Licensing and Credentialing Boards Respond When Government Clinicians Spread False or Misleading Health Information?*, AMA J Ethics. 2023;25(3):E210-218.; See also, Sage William M. Sage & Tony Yang, *Reducing “COVID-19 Misinformation” While Preserving Free Speech*, JAMA. 2022;327(15):1443-1444. 加えて、高良幸哉「インターネット上の医療フェイク情報規制について」情報ネットワーク法学会第22回研究大会個別報告C-4（2022年12月4日）も参照。

** Liberty Justice Center Website, *Liberty Justice Center Wins Battle for Doctors’ First Amendment Rights as California Repeals Physician Censorship Law*, Oct 2, 2023, <<https://libertyjusticecenter.org/pressrelease/liberty-justice-center-wins-battle-for-doctors-first-amendment-rights-as-california-repeals-physician-censorship-law/>>.

*** See, *Manhattan Community Access Corp. v. Halleck*, 139 S. Ct. 1921 (2019).

**** *New York Times v. Sullivan*, 376 U.S. 254 (1964), *United States v. Alvarez*, 567 U.S. 709 (2012).

***** *Janus v. Am. Fed’n of State, Cnty. & Mun. Emps., Council 31*, 138 S. Ct. 2448, 2501 (2018) (Kagan, J., dissenting).; See, Seth F. Kreimer, *The ‘Weaponized’ First Amendment at the Marble Palace and the Firing Line: Reaction and Progressive Advocacy Before the Roberts Court and Lower Federal Courts*, 72 Emory L. J. 1143 (2023).; See also, Jeremy K. Kessler, *The Early Years of First Amendment Lochnerism*, 116 Colum. L. Rev. 1915 (2016).;

II. DPF事業者規制をめぐるアメリカの動向

▶ 表現の自由（合衆国憲法修正一条）の「**ステロイド強化版**」*としての**通信品位法230条**（日本でいうプロ責法3条に相当）

→ 「双方向型コンピューターサービスのプロバイダ又はユーザー」に対する非常に広範な免責を与えている。

→ 初期のオンライン情報媒介者に対する法的責任に関し、判例によって生じた課題を解消するため、1996年に制定された。

- この規定の存在は、**事業者の自主的なコンテンツ・モデレーションを促進することも目的**としていた**。
- ただし近年では、免責範囲及び対象を限定する判例の動向や、FOSTA-SESTAによる性的人身売買（Sex Trafficking）に関する例外が設けられている。
- テロコンテンツのレコメンドに対するCDA230条免責をめぐる裁判が係争中であったが、連邦最高裁はCDA230条の解釈を回避した***。

But…特にトランプ政権前後から、主に保守派によるDPF事業者のコンテンツ・モデレーションへの反発も相まって、**事業者に対するコモンキャリア的規制（＝コンテンツ・モデレーションの制限）を行う試み**が、いくつかの州で実施されている。

* ジェフ・コセフ『ネット企業はなぜ免責されるのか—言論の自由と通信品位法230条』（みすず書房、2021年）280頁。

** 前掲・318頁。

*** Twitter, Inc. v. Taamneh, 598 U.S. 471 (2023).; See also, Gonzalez v. Google LLC, 598 U.S. 617 (2023).

II. DPF事業者規制をめぐるアメリカの動向

➤ コンテンツ・モデレーションを制限することを試みた二つの州法

① フロリダ州法 SB7072*

ソーシャルメディア・プラットフォームによる候補者のアカウントに対する凍結措置や、候補者の投稿等に対する事後的な優先順位付け・シャドウバンを禁止、ジャーナリズム企業への検閲等の禁止etc…

+ 透明性やユーザーへの告知の義務付け

② テキサス州法 HB20**

ソーシャルメディア・プラットフォームによるユーザーの「見解」に基づいた「検閲」の禁止

+ 透明性、情報開示の規定

➔ いずれも現在は、修正一条違反を理由にした州法の予備的差止め命令をめぐる係争中***

➤ クラレンス・トーマス判事****

➔ ソーシャルメディアのコモンキャリアまたはパブリックアコモデーションとの類似性を指摘。

* See, <<https://www.flsenate.gov/Session/Bill/2021/7072>>

** See, <<https://capitol.texas.gov/billlookup/History.aspx?LegSess=872&Bill=HB20>>

*** See also, *NetChoice, LLC v. AG, Fla.*, 34 F.4th 1196 (May 23, 2022).; *NetChoice, L.L.C. v. Paxton*, 49 F.4th 439 (Sep 16, 2022).

**** *Biden v. Knight First Amendment Institute at Columbia Univ.*, 141 S. Ct. 1220 (2021) (THOMAS, J., concurring).

Ⅲ. プラットフォーム・ガバナンスと透明性

➤ DPF事業者の不透明性をめぐる背景

✓ アメリカにおける自主規制促進を背景にDPF事業者は「**プラットフォーム法**」システムを構築

- DPF事業者は「依然として得体のしれない規制当局であり、明確性、一貫性、説明責任及び救済策が捉えどころのない一種の『プラットフォーム法』を制定している」*。

→ デジタル空間の秩序形成の主導権が、公的機関から私的アクターへの移行（=**法の支配の後退**）

✓ 「**コンテンツ・カルテル（content cartels）**」の形成**

- DPF事業者間またはDPF事業者と政府の間の「協調（collaboration）」は、場当たりの、特定の危機対応のみを目的としており、そのプレッシャーから「**適切な監督なく、コンテンツやアクターをサービスから排除するためにプラットフォーム間が一致協力する取り決め**」を形成してしまう。

→ 特に政府関係者との「協力関係」がもたらすDPF事業者の「**検閲代理人化（censorship by proxy）**」のリスク***



* 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, pp. 3.

** Evelyn Douek, *The Rise of Content Cartels*, 20-04 Knight First Amend. Inst. (Feb. 11, 2020), pp.6, <<https://knightcolumbia.org/content/the-rise-of-content-cartels>>.

*** Andrew M. Grossman and Kristin A. Shapiro, *Shining a Light on Censorship: How Transparency Can Curtail Government Social Media Censorship and More*, Cato Institute Briefing Paper No. 168, Oct 3, 2023, <<https://www.cato.org/briefing-paper/shining-light-censorship-how-transparency-can-curtail-government-social-media>>.

III. プラットフォーム・ガバナンスと透明性

➤ アメリカ版DSA？—プラットフォーム説明責任及び透明化法案 (The Platform Accountability and Transparency Act, PATA)

*

➔超党派で連邦上院に提出（2023年12月現在）。全10セクションで構成（第1：ショートターム、第2：定義、第3：適格研究計画、適格研究者、適格データ及び情報、第4：プラットフォームの義務及び免責、第5：適格研究者の義務及び免責、第6：報告、第7：執行、第8：ソーシャルメディアプラットフォームにおける研究のためのセーフハーバーの確立、第9：規則制定権限、第10：予算承認）

➤PATAは、3つの柱から構成されている**。

(1) 研究者による特定のデータアクセス

(2) 自動データ収集に関する限界付けられた法的セーフハーバー

(3) 開示を通じた強化された透明性

①バイラル・コンテンツ、②広告ライブラリ、③アルゴリズムック・デザイン、④コンテンツ・モデレーション

➔データ提供に従わない事業者へのCDA230条免責のはく奪条項は見送りに***



* Platform Accountability and Transparency Act, S.1876 — 118th Congress (2023-2024).; *See also*, Senator Chris Coons Website, *Senator Coons, colleagues introduce legislation to increase transparency around social media platforms*, Press Releases, June. 8, 2023, <<https://www.coons.senate.gov/news/press-releases/senator-coons-colleagues-introduce-legislation-to-increase-transparency-around-social-media-platforms>>.

** <<https://www.coons.senate.gov/download/pata-one-pager-118th-congress-june-2023>>.

*** John Perrino, Platform Accountability and Transparency Act Reintroduced in Senate, Tech Policy Press, Jun. 8, 2023, <<https://techpolicy.press/platform-accountability-and-transparency-act-reintroduced-in-senate/>>.

III. プラットフォーム・ガバナンスと透明性

▶ バイデン政権が行ってきたCOVID-19禍の偽情報対策等の一環でDPF事業者に対して行われてきた「**口先介入 (jawboning)**」に関して、現在、修正一条違反を理由にした差止訴訟が係争中*

「**口先介入**」：DPF上の合法的な言論の取り締まりを行うために、政府関係者が事業者に対して非公式な圧力をかけること

- 政府による事業者への適切な「**説得 (persuade)**」は必要な国家行為であり、憲法上も許容される**。
- 他方で「**強制 (coerce)**」は許されないと考えられるが、その間の線引きは？***

▶ 口先介入は修正一条に基づく訴訟や禁止立法では解決し得ない。

→ 政府職員によるDPF事業者に対する表現の自由に関連する「**接触**」を「**透明化**」することで、検閲代理人化に対処するという方向性***。

→ “**The Social-Media Censorship Transparency Act**” という提案



* Missouri v. Biden, 2023 U.S. App. LEXIS 23965.; Murthy v. Missouri, 601 U. S. ____ (2023).

** See, Ugarte, Ramiro Álvarez, *Ruling by Bullying?: On the Difficult Balance between Illicit Threats and Effective Government*, VerfBlog, Oct 12, 2023, <<https://verfassungsblog.de/ruling-by-bullying/>>.

*** See, Kennedy v. Warren, 66 F.4th 1199, 1207 (9th Cir. 2023).

*** Andrew M. Grossman and Kristin A. Shapiro, *Shining a Light on Censorship: How Transparency Can Curtail Government Social Media Censorship and More*, Cato Institute Briefing Paper No. 168, Oct 3, 2023, <<https://www.cato.org/briefing-paper/shining-light-censorship-how-transparency-can-curtail-government-social-media>>.

**** See, Will Duffield, *Toward a Jawboning Transparency Act*, Knight First Amend. Inst. (Oct. 19, 2023), <<https://knightcolumbia.org/blog/toward-a-jawboning-transparency-act>>.

III. プラットフォーム・ガバナンスと透明性

➤プラットフォーム・ガバナンスにおける「透明性」の意味

➔透明性はそれ自体が目的ではなく、手段。

①「正統性 (legitimacy)」の担保*

→透明性レポートは、DPF事業者が、「ユーザー及び市民社会組織と正統性を醸成する(cultivate)」役割を果たす**。

②不誠実な対応への抑止効果

→”Sunlight is said to be the best of disinfectants; electric light the most efficient policeman”***。

③規制の（正負の）効果観測のための継続的モニタリング

→削除の迅速化などの規制はしばしばエラーの増加とトレードオフの可能性があるため****

④AI等の技術利用がもたらすリスクに対する継続的モニタリング

→フェイスブック・ファイルの事例*****



* Edoardo Celeste, Nicola Palladino, Dennis Redeker, Kinfe Yilma, The Content Governance Dilemma: Digital Constitutionalism, Social Media and the Search for a Global Standard 119 (2023).

** Nicolas Suzor, Tess Van Geelen, Sarah Myers West, *Evaluating the legitimacy of platform governance: A review of research and a shared research agenda*, the International Communication Gazette Vol. 80(4), 393 (2018).

*** Louis D. Brandeis, Other People's Money, and How the Bankers Use It, Chapter V: What Publicity Can Do, < <https://louisville.edu/law/library/special-collections/the-louis-d.-brandeis-collection/other-peoples-money-chapter-v> > .

**** See, Evelyn Douek, Content Moderation as Systems Thinking, 136 Harv. L. Rev. 526, 549 (2022).

***** Georgia Wells, Jeff Horwitz, Deepa Seetharaman, *Facebook Knows Instagram Is Toxic for Teen Girls, Company Documents Show*, The Wall Street Journal, Sept. 14, 2021 < <https://www.wsj.com/articles/facebook-knows-instagram-is-toxic-for-teen-girls-company-documents-show-11631620739> > .

今後の検討課題



- デジタルメディア環境の設計に関する基本理念を設定する必要性：
デジタル立憲主義—デジタル空間の立憲化へ*
- 日本の文脈においては、情報受領者（国民）の「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」**、特に信頼性の高い情報へのアクセス保障が重要。
- 個別の権利救済のための議論に加え、**デジタルメディア「環境」におけるDPF事業者らの社会的責務に応じた責任配分の議論**へとシフト
- 透明性確保についても、事業者に対する市場等の圧力が不足するようであれば、**要請ベースから公法上の義務化**へのシフトも検討すべき。
- 継続的な透明性レポートに対する**外部からのモニタリング**も必要
- DPF事業者らの「検閲代理人化」を防止するため、**事業者へのコンテンツ削除要請などを行う政府機関の透明性確保も非常に重要。**

+ 情報受領者に対するリテラシー教育も、「メディア・インフラ・リテラシー」***のようにデジタルメディア環境の「構造」に応じたアップデート****を行う必要性

* 詳細は、山本健人「EUのAI規則案とデジタル立憲主義」東京大学未来ビジョン研究センター・ワーキングペーパーNo.13（2023年2月）、山本健人「デジタル立憲主義と憲法学」情報法制研究13号（2023年）56-71頁、豊秀一・荒ちひろ「AIと私たち『デジタル立憲主義』とは 山本健人さん、エドアルド・セレステさん」朝日新聞2023年12月14日 <<https://digital.asahi.com/articles/DA3S15816128.html>>を参照。

** 最大判昭和58年6月22日民集37巻5号793頁。

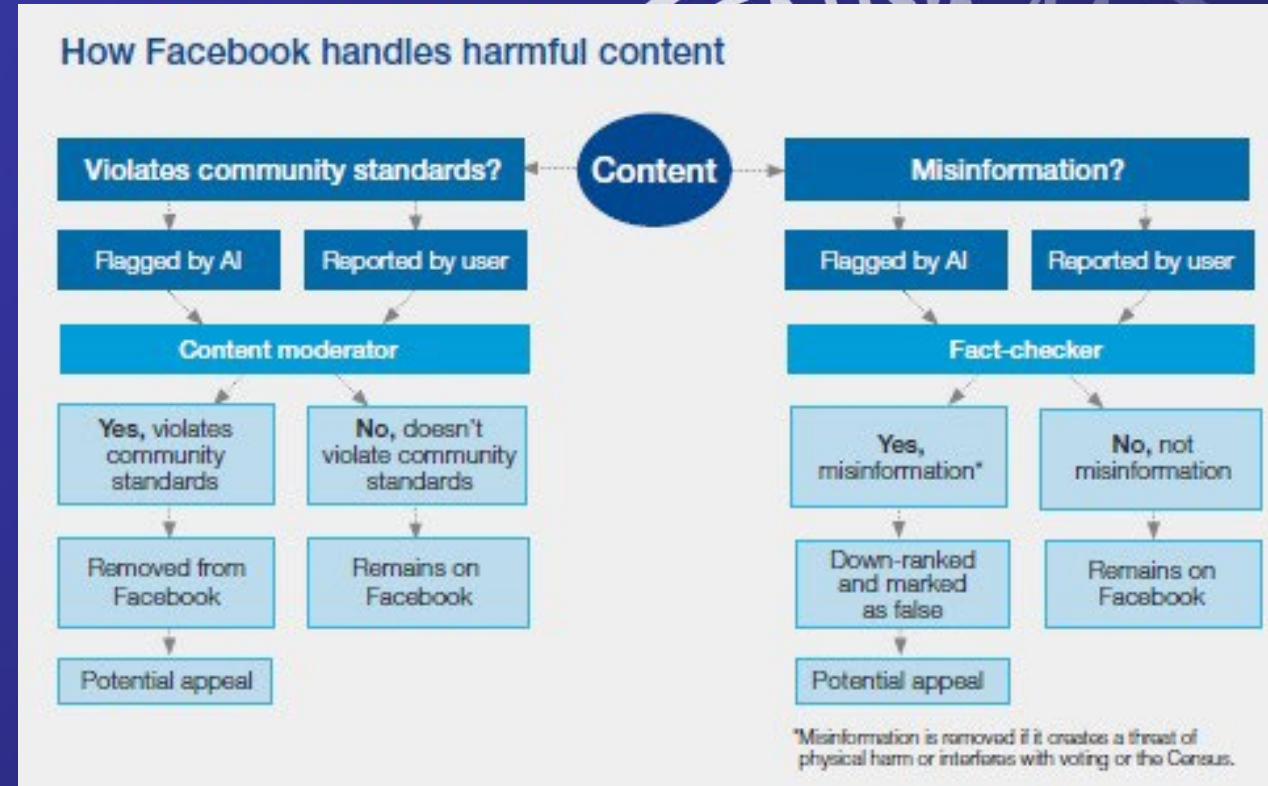
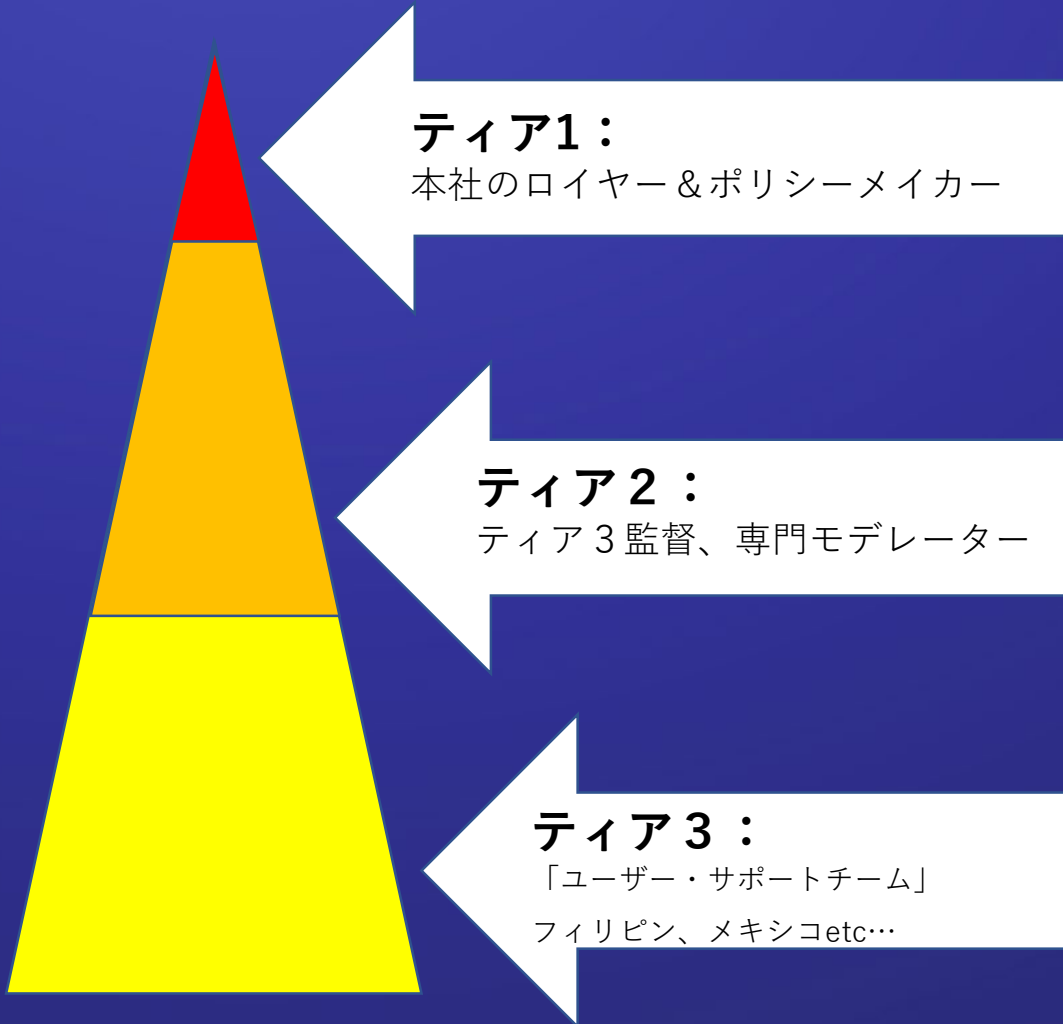
*** 水越伸・宇田川敦史・勝野正博・神谷説子「メディア・インフラのリテラシー—その理論構築と学習プログラムの開発」情報学研究：東京大学大学院情報学環紀要No.98（2020年）1-30頁。

**** デビッド・バッキンガム（水越伸監訳、時津啓・砂川誠司訳）『メディア教育宣言—デジタル社会をどう生きるか』（世界思想社、2023年）も参照。

ご清聴ありがとうございました。



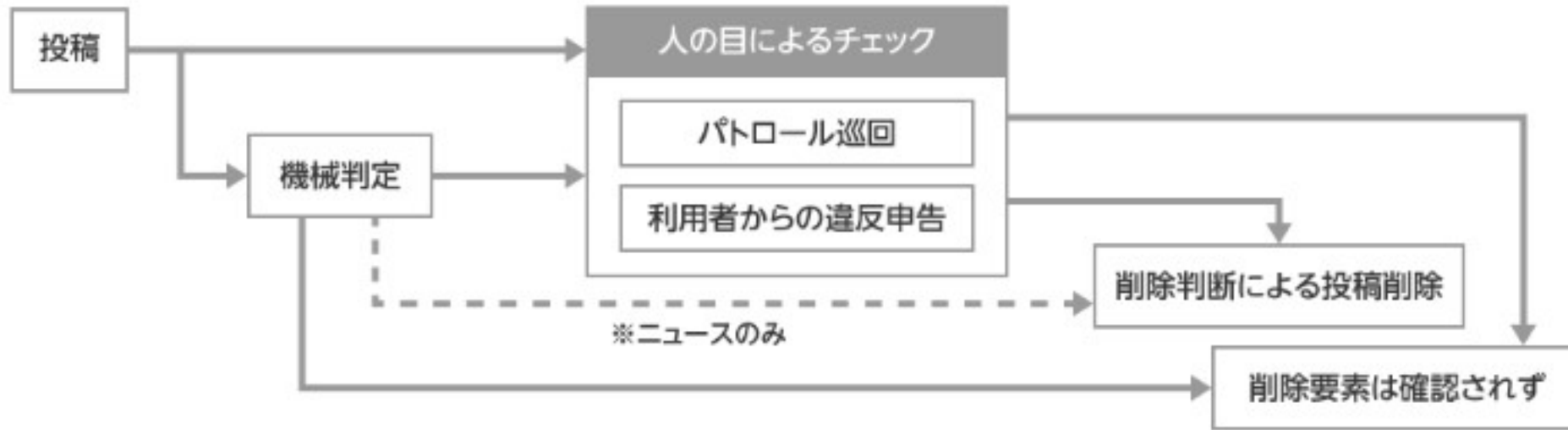
(1) Facebookの事例



- なおモデレーションに事前的な措置と事後的な措置があるとされている。

※図（左）は、Kate Klonick, *The New Governors: The People, Rules, and Processes Governing Online Speech*, 131 Harv. L. Rev. 1598, 1639-41 (2018) を参考に筆者が作成。
 ※図（右）は、PAUL M. BARRETT, *Who Moderates the Social Media Giants? A Call to End Outsourcing*, NYU Stern Center for Business and Human Rights (June. 8, 2020), pp.1, <<https://www.stern.nyu.edu/experience-stern/faculty-research/who-moderates-social-media-giants-call-end-outsourcing>>より引用。

(2) Yahoo!ニュースのコメント欄の事例*



- 約70名のモデレーション・チームとAIの組み合わせでモデレーションが行われており、AI・機械学習については、「建設的」、「関連度」、「不適切投稿判定」という3つのモデル（最近はさらにコメント多様化モデル**）を採用している。

* 以下は、Yahoo! JAPAN「メディア透明性レポート（2020年度版）」

（https://about.yahoo.co.jp/common/transparencyreport_2020/）を参照。画像も同様。

** newsHack「コメント欄の『おすすめ順』で多様な意見が上位に表示されやすくなる機能の導入について」（2023年4月18日）<https://news.yahoo.co.jp/newshack/information/comment_20230418.html>.

➤ 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*。

→モデレーションの量が膨大で、人力だけでは限界がある。

- 例えばテロリズムコンテンツのフラグ立てについて、Twitterはアカウント凍結の**93%**がスパム対策ツールによるもの**であり、YouTubeに至っては削除動画の**98%**が機械学習アルゴリズムによるもの***。

➤今ではモデレーションにおいて自動化は不可欠なツール（右表****）

→必然的に生じるエラー（偽陽性／偽陰性）。そして、**Metaの例でいえばわずか1%のエラーですら、一日100万件になる。**

表：Facebookのプロアクティブ検知のシェア

	2018年第3 クォータ	2019年第3 クォータ	2020年第3 クォータ	2021年第3 クォータ	2022年第3 クォータ
いじめと ハラスメ ント	14.8%	16.2%	31.0%	59.4%	67.8%
自殺と自 傷	N/A	96.8%	95.7%	99.0%	98.6%
児童性的 搾取	N/A	N/A	N/A	99.1%	99.5%
ヘイトス ピーチ	52.9%	80.6%	94.8%	96.5%	90.2%
ドラッグ	N/A	93.8%	91.7%	94.1%	94.8%
スパム	99.7%	99.9%	99.9%	99.6%	98.5%

* See, Over Sight Board, PAO-2021-02, < <https://oversightboard.com/decision/PAO-NR730OFI/>>.

** See, Isobel Asher Hamilton, Facebook, YouTube, and Twitter could face fines if they fail to take down terrorist content within minutes, BussinessInsider, Aug 20, 2018, <<https://www.businessinsider.com/eu-could-fine-tech-firms-if-they-fail-to-remove-terror-content-2018-8>>.

*** See, Susan Wojcicki, Expanding our work against abuse of our platform, YouTube Official Blog, Dec 05, 2017, <<https://blog.youtube/news-and-events/expanding-our-work-against-abuse-of-our/>>.

**** 右表は、Edoardo Celeste, Nicola Palladino, Dennis Redeker, Kinfe Yilma, The Content Governance Dilemma: Digital Constitutionalism, Social Media and the Search for a Global Standard 117 Table 5-10 (2023).の一部を抜粋し、水谷が作成

参考資料③：米・AI規制に関する大統領令の概要

▶ 安全性、セキュリティ及び信頼性のある人工知能の開発及び利用に関する大統領令*

→NISTによるAIに関する厳格な安全性基準の設定、AIによる差別への対応、AIによる労働市場への害悪の軽減措置といった各種取り組みが掲げられた、法的拘束力を有するAI規制（ただし、議会の制定する法律ではない）

<ファクトシート>**

①AIの安全性及びセキュリティに関する新基準

→AI生成コンテンツの検出及び公的コンテンツの認証のための規準、電子透かしの開発、アメリカ人が政府から受領する情報の真正性を確保

②アメリカ人のプライバシー保護

③公平性及び公民権の増進

④消費者、医療患者及び学生の福利向上

⑤労働者への支援

⑥イノベーション及び競争促進

⑦海外でのアメリカのリーダーシップ増進

⑧責任ある、効果的な政府によるAI利用の確保

* Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence (Oct. 30, 2023), <<https://www.whitehouse.gov/briefing-room/presidential-actions/2023/10/30/executive-order-on-the-safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence/>>

** White House Website, FACT SHEET: President Biden Issues Executive Order on Safe, Secure, and Trustworthy Artificial Intelligence, Oct. 30, 2023, <<https://www.whitehouse.gov/briefing-room/statements-releases/2023/10/30/fact-sheet-president-biden-issues-executive-order-on-safe-secure-and-trustworthy-artificial-intelligence/>>.