# ヘイトスピーチとソーシャルメディア

水　谷　瑛嗣郎
民主主義の再生と「公共圏」研究班　研究員
関西大学　社会学部　准教授

ヘイトスピーチと呼ばれる憎悪・差別的表現による被害は、現代においては特にSNSをはじめとするソーシャルメディア上で広がりを見せている。ヘイトスピーチに対する規制は、表現の自由との兼ね合いから各国の対応に差があるものの、日本においては、国法レベルと条例レベルの二段階で規制が設けられている。国法レベルでは、いわゆるヘイトスピーチ解消法が、理念法として機能し、ヘイトスピーチを許容しないという政府の姿勢を示している。またいくつかの地域では、こうした国法を補う条例が制定されている。インターネット上のヘイトスピーチ対策との関連に絞れば、大阪市の条例には、プロバイダへの削除要請等を行う拡散防止措置の権限が市長に与えられている（5条）。同条例は、最判令和4年2月15日民集第76巻2号190頁で、制裁規定の不在などを理由に憲法（表現の自由）に反しないと結論付けられている。しかしながら、国境すらも超えるインターネットの特性上、こうした地方自治体の条例では対応に限界が生じる。

その一方、ソーシャルメディアを運営するプラットフォーム（PF）事業者は、私企業でありながら、自社のオンライン空間をデザインし、そこで流通するコンテンツを管理するためのルールを設定し、コンテンツを管理する情報環境形成力を有している。こうしたオンライン空間に対する事実上の権力をもって空間を統治するPF事業者の姿をして、日本の憲法学者の山本龍彦は、中世のカトリック教会との類似性を見出し、しばしば国家が「リヴァイアサン」と評されることに対応して、「ビヒモス」と称している。

こうしたPF事業者の統治能力とヘイトスピーチは決して無縁な関係ではない。例えば、YouTubeに対するある調査は、ユーザーが視聴して「後悔」した動画（その中にはヘイトスピーチも含まれている）の表示に、レコメンド（推薦）システムが関係していることを示唆している。その一方で、PF事業者は多くの場合、コミュニティ規定などによりヘイトスピーチを規制している。いわゆるコンテンツ・モデレーションと呼ばれる仕組みでは、規定違反の投稿を削除し、違反アカウントを凍結するといったものにとどまらない多彩な手段が用いられている。そして、そのモデレーションの対象範囲には、日本において違法とされないヘイトスピーチも

含まれている。ただこうした仕組みは国家による法規制とは異なる私企業による自主的な取り組みであり、基本的に国家に対する制限規範である憲法との間で、これまで問題視されることはなかった。しかし近年、PF事業者のこの仕組みを「プラットフォーム法」と位置づける動きが見受けられる。先に見た通り、PF事業者がオンライン空間で発揮する権力が単なる私企業というより国家権力に匹敵するものになりつつあり、そうした観点から「プラットフォーム法」の諸相を明らかにしたうえで、その憲法的な統御（ガバナンス）を検討する必要が生じているのである。本講演では、PF事業者のモデレーションの仕組み（ルールおよび執行プロセス、執行状況）を概観したうえで、「プラットフォーム法」を、①契約法、②規範的（実体的）コモン・ロー、③手続き的コモン・ロー、④技術法という4つの観点から整理する議論を紹介したうえで、ヘイトスピーチ対応という観点から、その手続き上の不透明性及び文脈の無視という二つの課題に対処するためのガバナンスの一つとして、人権的アプローチ及びシステミック・アプローチについて検討を行った。

KANSAI UNIVERSITY

関西大学　経済・政治研究所第256回産業セミナー
民主主義の再生と「公共圏」研究班：「公共圏」の構築における自由と参加

2023年6月30日（金）13:00～

# ヘイトスピーチとソーシャルメディア

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）

1

KANSAI UNIVER

## 水谷瑛嗣郎（みずたにえいじろう）の自己紹介

- 実家は、三重県松阪市で四代続く蜂蜜屋。でも生まれは大阪の豊中。
- 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。立正大学、神奈川大学、法政大学、国際基督教大学などで非常勤で教える。
- 2019年4月より関西大学社会学部に移籍。
- 博士（法学）。専門は憲法学・メディア法・情報法。
- 主として、**インターネットやAI技術の普及した世界におけるマスメディアの役割や民主政治の未来**について研究中

## KANSAI UNIVERSITY 水谷瑛嗣郎（みずたにえいじろう）の自己紹介


- 共著本『**AIと憲法**』（日本経済新聞出版社）のうち、「AIと民主主義」の章を担当【写真・左】。
- 共著本の『**憲法の現在地**』（日本評論社）では、「マスメディアの自由と特権」の章を担当【写真・左】。
- 共著本の『**Liberty2.0**』（弘文堂）では、「ポスト・トゥルース」の章を担当【写真・右下】。
- 編著本として『**リーディング　メディア法・情報法**』（法律文化社）【写真・右上】。

→若手研究者が集まり、インターネットやAIの普及を踏まえて、報道の自由や放送の自由、プロバイダ責任制限法、個人情報保護法などについて解説・検討したもの。


※画像（上）は、以下より引用。
https://www.hanmoto.com/bd/isbn/9784589042200
画像（下）は、以下より引用。
https://www.koubundou.co.jp/book/b618726.html

3

## KANSAI UNIVERSITY メディア掲載


4

KANSAI UNIVERSITY

## 突然ですが、皆さんはこのニュースをどのように受け止めましたか？

- 2022年10月に、スペースXやテスラで有名な世界的な大富豪のイーロン・マスク氏が、Twitter社を440億ドル（約5.6兆円）相当で買収しました。

- 彼はどのくらい金持ちなのか？

→Forbesの“Today’s Winners and Losers”


画像はイーロン・マスクのTwitterアカウントより
https://twitter.com/elonmusk/status/1041880316657336320

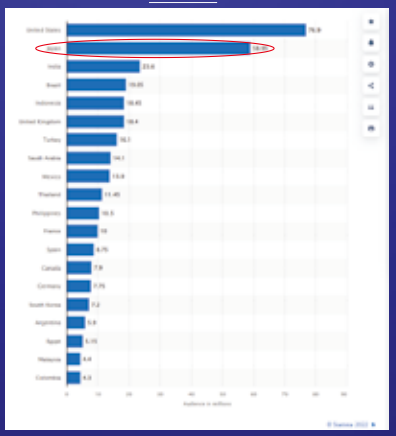
5

KANSAI UNIVERSITY

## 突然ですが、皆さんはこのニュースをどのように受け止めましたか？

- 圧倒的にTwitterに依存している日本・・・

Twitter



Facebook

画像（左）は、Statista, Leading countries based on number of Twitter users as of January 2022(in millions), Mar 22, 2022, at https://www.statista.com/statistics/242606/number-of-active-twitter-users-in-selected-countries/ より引用。画像（右）は、 Statista, Leading countries based on Facebook audience size as of January 2022(in millions), Mar 8, 2022, at https://www.statista.com/statistics/268136/top-15-countries-based-on-number-of-facebook-users/より引用。

6

KANSAI UNIVERSITY

## 突然ですが、皆さんはこのニュースをどのように受け止めましたか？

➢ 圧倒的にTwitterに依存している日本・・・

- 世界的に見れば、ソーシャルメディアのMAUs（月間アクティブユーザー数）は、圧倒的にFacebook

**→約29億**


画像は、Statista, Most popular social networks worldwide as of January 2023, ranked by number of monthly active users(in millions), Feb 14, 2023, at https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/ より引用。

7

KANSAI UNIVERSITY

## 突然ですが、皆さんはこのニュースをどのように受け止めましたか？

➢ 圧倒的にTwitterに依存している日本・・・

| | 全年代(N=1,500) | 10代(N=142) | 20代(N=213) | 30代(N=250) | 40代(N=326) | 50代(N=287) | 60代(N=282) | 男性(N=759) | 女性(N=741) |
|---|---|---|---|---|---|---|---|---|---|
| LINE | 90.3% | 93.7% | 97.7% | 95.6% | 96.6% | 85.4% | 76.2% | 88.0% | 92.7% |
| Twitter | 42.3% | 67.6% | 79.8% | 48.4% | 38.0% | 29.6% | 13.5% | 42.7% | 41.8% |
| Facebook | 31.9% | 19.0% | 33.8% | 48.0% | 39.0% | 26.8% | 19.9% | 32.4% | 31.4% |
| Instagram | 42.3% | 69.0% | 68.1% | 55.6% | 38.7% | 30.3% | 13.8% | 35.3% | 49.4% |
| mixi | 2.3% | 2.1% | 3.8% | 3.6% | 3.4% | 0.7% | 0.4% | 2.2% | 2.3% |
| GREE | 1.3% | 2.1% | 4.2% | 1.2% | 0.6% | 1.0% | 0.0% | 1.8% | 0.8% |
| Mobage | 2.7% | 4.9% | 6.6% | 2.4% | 0.9% | 2.4% | 1.4% | 3.8% | 1.6% |
| Snapchat | 1.5% | 4.9% | 5.6% | 0.4% | 0.3% | 0.3% | 0.4% | 1.1% | 2.0% |
| TikTok | 17.3% | 57.7% | 28.6% | 16.0% | 11.7% | 7.7% | 6.0% | 15.3% | 19.4% |
| YouTube | 85.2% | 96.5% | 97.2% | 94.0% | 92.0% | 81.2% | 58.9% | 87.9% | 82.5% |
| ニコニコ動画 | 14.5% | 26.8% | 28.2% | 14.8% | 12.0% | 7.7% | 7.8% | 17.9% | 11.1% |

画像は、総務省情報通信政策研究所「令和２年度情報通信メディアの利用時間と情報行動に関する調査 報告書」（令和3年8月）66頁の表5-1-1より引用。

8

## 突然ですが、皆さんはこのニュースをどのように受け止めましたか？

- 新しいCEOがTwitter上のルールやインターフェイスが大きく変化することで、Twitterを利用する多くの日本人ユーザーのユーザー体験、ひいては社会それ自体をも、変えてしまう。

例えば…

- トランプ元大統領のTwitterアカウント復活！
- **イーロンツイートが「For You」を埋め尽くす・・・！？**
- ヘイトスピーチやフェイクニュースなどに対するモデレーションが緩和されたり、アルゴリズムが変更されることで、不快なコンテンツがどんどん飛び込んでくるように・・・？

→報道によると、買収後にヘイトスピーチが急増しているとの調査結果もある*。

9

* Sheera Frenkel and Kate Conger, Hate Speech's Rise on Twitter Is Unprecedented, Researchers Find, The New York Times, Dec. 2, 2022, at https://www.nytimes.com/2022/12/02/technology/twitter-hate-speech.html

## 突然ですが、皆さんはこのニュースをどのように受け止めましたか？

- 2023年4月、**ドイツ連邦司法省が、ネットワーク執行法に基づき、ヘイトスピーチを含む違法コンテンツへの適切な対応を怠ったことを理由にした、Twitterに対して訴訟手続きに入った**との報道*。

- **独ネットワーク執行法（NetzDG）****

→3条において、プラットフォームに刑法上の規定に抵触する違法コンテンツの苦情対応・削除の手続きの設置を義務付けている。

➡基本的には、「明らかに違法なコンテンツ」に関して、苦情を受け付けてから24時間以内に削除またはアクセス遮断を行わなければならない。（例外有）

10

* Emma Woollacott「ドイツ当局がツイッターを提訴へ、「違法コンテンツ」放置で」Forbes（2023年4月6日）at https://forbesjapan.com/articles/detail/62228
** ネットワーク執行法については、實原隆志「ドイツの SNS 法 ――オーバーブロッキングの危険性について――」情報法制研究第 4 号（2018. 11）4 9 頁）、鈴木秀美「ドイツのSNS対策法と表現の自由」慶應義塾大学メディア・コミュニケーション研究所紀要68号（2018年）1頁以下を参照。

KANSAI UNIVERSITY

## 突然ですが、皆さんはこのニュースをどのように受け止めましたか？


- ネットワーク執行法は、違法コンテンツに対応する体制づくりを怠った場合に執行されるが、報道によると、そうした「システミックな失敗」を立証するために600件以上の事例が集められ、提出されたとのこと。

→ツイッター社が削除を拒否したコンテンツの中には、マスク氏による「恩赦」によって復活したユーザーの投稿もあったという*。

- 1件につき5000万ユーロの罰金×600件

＝**最高額で約300億ユーロ（約4兆7000万円）**

* See, Natasha Lomas, Twitter could be facing slew of fines in Germany over illegal hate speech, TechCrunch, April 6, 2023, at https://techcrunch.com/2023/04/06/elon-musk-twitter-germany-hate-speech-takedowns/

11

KANSAI UNIVERSITY

## 目次




12

KANSAI UNIVERSITY

# 1．ヘイトスピーチとインターネット

（1）日本におけるヘイトスピーチ規制の状況

**国法：ヘイトスピーチ解消法**

→いわゆる理念法。努力義務のみで、制裁規定無し*。

政府言論としての効果：「**国は、ヘイトスピーチ解消法を制定することにより、ヘイトスピーチは許されないとする強いメッセージを発しており**、また、同法は、国および地方公共団体に、教育や啓発活動など、さらにメッセージを発することを求めている」**。

**＋名誉毀損罪、侮辱罪、民事上の名誉毀損（不法行為）等**

→個人（法人）に向けられたものが対象、京都朝鮮学校判決**

**条例：大阪市ヘイトスピーチ対策条例、川崎市ヘイトスピーチ対策条例**

→ヘイトデモ等、リアル空間でのヘイトスピーチに対応。

＋ネット対応

13

* 同法について詳しくは、桧垣伸次「ヘイトスピーチ解消法と非規制的施策」桧垣・奈須編著『ヘイトスピーチ規制の最前線と法理の考察』（法律文化社、2021年）3-5頁。
** 前掲5－6頁。
** 京都地判平成25年10月7日判時2208号74頁

KANSAI UNIVERSITY

# 1．ヘイトスピーチとインターネット

（1）日本におけるヘイトスピーチ規制の状況

「現実空間のヘイトスピーチとインターネット上のヘイトスピーチは密接に関連している」との指摘*。

→ヘイト・デモなどの様子をYouTubeにアップロードすることで、ヘイトスピーチの被害が拡大する。

- 大阪市のヘイトスピーチ対策条例を例に…

**拡散防止措置**　→「インターネット上に書き込みがなされている場合は、プロバイダに削除要請を行うこと」を含む**

14

* 成原慧「インターネット上のヘイトスピーチとその規制」桧垣・奈須編著『ヘイトスピーチ規制の最前線と法理の考察』（法律文化社、2021年）63頁。
** 大阪市ウェブサイト「『大阪市ヘイトスピーチへの対処に関する条例』の解説及び審査の実例（5条）」（2022年5月12日）at https://www.city.osaka.lg.jp/shimin/page/0000438270.html

## 1．ヘイトスピーチとインターネット

**条例第5条** 市長は、次に掲げる表現活動がヘイトスピーチに該当すると認めるときは、事案の内容に即して当該表現活動に係る**表現の内容の拡散を防止するために必要な措置**をとるとともに、当該表現活動がヘイトスピーチに該当する旨、表現の内容の概要及びその拡散を防止するためにとった措置並びに当該表現活動を行ったものの氏名又は名称を公表するものとする。ただし、当該表現活動を行ったものの氏名又は名称については、これを公表することにより第1条の目的を阻害すると認められるとき、当該表現活動を行ったものの所在が判明しないときその他特別の理由があると認めるときは、公表しないことができる。

**(1) 本市の区域内で行われた表現活動**

**(2) 本市の区域外で行われた表現活動（本市の区域内で行われたかどうか明らかでない表現活動を含む。）で次のいずれかに該当するもの**

ア 表現の内容が市民等に関するものであると明らかに認められる表現活動

**イ アに掲げる表現活動以外の表現活動で本市の区域内で行われたヘイトスピーチの内容を本市の区域内に拡散するもの**

15

## 1．ヘイトスピーチとインターネット

（2）大阪市ヘイトスピーチ対策条例最高裁判決

➢**最判令和4年2月15日民集第76巻2号190頁***

「また，本件各規定により制限される表現活動の内容及び性質は，上記のような過激で悪質性の高い差別的言動を伴うものに限られる上，**その制限の態様及び程度においても，事後的に市長による拡散防止措置等の対象となるにとどまる**。そして，**拡散防止措置については，**市長は，看板，掲示物等の撤去要請や，**インターネット上の表現についての削除要請等を行うことができると解されるものの，当該要請等に応じないものに対する制裁はなく**，認識等公表についても，表現活動をしたものの氏名又は名称を特定するための法的強制力を伴う手段は存在しない。」

⇔拡散防止措置が「機能的に検閲・事前抑制に近い性質をもつ」ことに加え、氏名等公表に比べて手続保障が簡略化されている点をどう考えるか**

16

* https://www.courts.go.jp/app/hanrei_jp/detail2?id=90920
** 成原慧「インターネット上のヘイトスピーチとその規制」桧垣・奈須編著『ヘイトスピーチ規制の最前線と法理の考察』（法律文化社、2021年）69, 71頁。

## 1．ヘイトスピーチとインターネット

➢ **条例５条各号の趣旨***

**1号（区域内）**：「インターネット上の表現活動については、その表現内容が大阪市内において閲覧・視聴可能な状態であることをもって、大阪市内で行われた表現活動と認定されるわけではありません」

→1号に該当しない例として「インターネット上の短文投稿サイトにおいて投稿をしていた表現活動」が挙げられており、その理由として「**インターネット上の表現活動の実施場所を特定するためには、表現活動を行ったものに投稿した場所を問い合わせるだけでは足りず、表現活動を行ったものの回答内容が事実であるかどうかを確認するため、サイトの運営者や関係プロバイダからIPアドレス等の必要な情報を取得する必要がある。**この点、サイトの運営者や関係プロバイダから投稿が行われた場所を特定するために必要な情報が任意に提供される可能性は非常に低く、仮にサイトの運営者や関係プロバイダから情報が得られたとしても、**インターネット上のサイトへの投稿の多くが無線の通信端末機器により行われている現状に鑑みると、投稿が行われた場所を特定することは極めて困難であると考えられたため。**」

**2号イ（区域外または不明）**：「市域内で行われたヘイトスピーチ(これ自体は条例第5条第1項第1号に該当します。)の動画等を、インターネット上のサイトに掲載すること等によって大阪市の区域内に拡散させる行為などについては、**条例第5条第1項の規定による措置等の対象となります。**」

17

* 大阪市ウェブサイト「『大阪市ヘイトスピーチへの対処に関する条例』の解説及び審査の実例（5条）」（2022年5月12日）at https://www.city.osaka.lg.jp/shimin/page/0000438270.html

## 1．ヘイトスピーチとインターネット

（3）デジタル空間を取り巻く特質

➢デジタル空間のアルゴリズミックな性質

→情報流通は、ますますアーキテクトにより「**設計（デザイン）された場**」の上で行われる。

➢情報資本主義の到来*

→思想の市場と経済の市場の間の垣根が融解

➢アテンション・エコノミー（関心経済）

→より効率的に私たちの「**注目**」を奪えるよう、センセーショナルまたはサブリミナルな形態で情報の発信と流通が行われる。

18

* 拙稿「思想の自由『市場』と国家：表現の自由の『環境』構築を考える」法律時報92巻9号（2020年）30-37頁を参照。

KANSAI UNIVERSITY

# 1．ヘイトスピーチとインターネット

➢アテンション・エコノミー

19

KANSAI UNIVERSITY

# 1．ヘイトスピーチとインターネット

➢**現在のデジタル空間に通底しているビジネスモデルは、人々の「注目」をいかに獲得するかというもの**

「**注目**は異なるプラットフォーム間で共通のものであり、……つねに**稀少な資源**である」*。

- （常にではないが）主として広告収入を高めるため、いかにユーザーの「**注目**」を引き、「場」に取り込み、「**粘着性**」を高めるかが重要。

「**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。」**。

→いまコンテンツを生成するすべての人々が直面しているのは、「粘着性」をめぐって行われる「**スマートフォンの画面に浮かぶアプリケーション及びコンテンツ間の競争**」***。

20

* フィン・プラントン（生貝直人他訳）『スパム［spam］　インターネットのダークサイド』（河出書房新社、2015年）24頁。
** 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。
*** 拙稿「デジタル情報空間における放送と放送法制」ジュリスト8月号（2022年）41頁。

## 1．ヘイトスピーチとインターネット

➢アテンション・エコノミー

- 主として広告収入を高めるため、いかにユーザーの「**注目**」を引き、「場」に取り込み、「**粘着性**」を高めるかが重要。

21

画像（左）はいらすとやより　https://www.irasutoya.com/　画像（右上）はfreepikウェブサイトより　https://jp.freepik.com/premium-vector/kyiv-ukraine-march-30-2021-iphone-with-social-media-apps-set-instagram-facebook-twitter-you-tube-wechat-tik-tok-whatsapp-and-pinterest-ui-ux-white-user-interface_16313247.htm　画像（右下）はPRTIMEウェブサイトより　https://prtimes.jp/main/html/rd/p/000000082.000011276.html

## 2．　ソーシャルメディアのヘイトスピーチ対策

（1）デジタル・プラットフォーム事業者の台頭

➢「プラットフォーマー型ビジネスの台頭に対応したルール整備の基本原則」*。

**①「社会経済に不可欠な基盤を提供していること」**

**②「多数の消費者（個人）や事業者が参加する場そのものを、設計し運営・管理する存在であること」**

**③「そのような場は、本質的に操作性や技術的不透明性があること」**

※画像はいらすとやより
https://www.irasutoya.com/

- このうち、「ソーシャルメディア」プラットフォームについては、「**インターネットを利用して誰でも手軽に情報を発信し、相互のやりとりができる双方向のメディア**であり、代表的なものとして、**ブログ**、FacebookやTwitter等の**SNS（ソーシャルネットワーキングサービス）**、YouTubeやニコニコ動画等の**動画共有サイト**、LINE等の**メッセージングアプリ**がある」とされる*。

＝ここでは、主として、検索エンジン、ソーシャルメディア（Google、Facebook、Twitter、YouTube、etc…）を主眼に。

22

* 経済産業省・公正取引委員会・総務省「プラットフォーマー型ビジネスの台頭に対応したルール整備の基本原則」（2018年12月18日）
** 総務省『情報通信白書（平成27年版）』199頁。

## 2．　ソーシャルメディアのヘイトスピーチ対策

➢ソーシャルメディアを運営するデジタル・プラットフォーム事業者は、国家機関でないにもかかわらず、自社が形成し、管理する「場」において、いまや国家に匹敵する（か場合によってはそれ以上に）**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*。

＝「『手のひらの上』の自由」**

➡より多くのユーザーに「**心地よく**」場を利用してもらうための「**場の形成・管理・運営**」を行う機能。

　+こうした環境のデザインが、オンライン上のユーザー体験を形成している、という点にも注意を払う必要がある。

23

* 拙稿「米大統領のアカウント凍結からプラットフォームのあり方を探る」Journalism 371号（2021年）54-61頁を参照。
** 例えば水谷瑛嗣郎（聞き手：小嶋麻友美）「トランプ氏のアカウント凍結で気づいた「手のひらの上の自由」　水谷瑛嗣郎・関西大准教授」東京新聞2021年1月28日 at https://www.tokyo-np.co.jp/article/82699

## 2．　ソーシャルメディアのヘイトスピーチ対策

➢DPF事業者は、自社が提供する場の管理者として、国家に“類似”する統治システムを有しつつある。

→「**オンライン言論の新たな統治者（the New Governors）**」*。

- 領域主権（territorial sovereignty）から機能主権（functional sovereignty）へ

→DPF事業者はすでにルール形成、執行、紛争裁定といった機能を担い始めている。

＝ユーザーは「**専制君主の臣民（subjects of a despot）**」 **の地位へ。

- リヴァイアサンVSビヒモス***

「国家とDPFとの関係について考えるうえで有用だと思われるのが、中世ヨーロッパ封建制社会における国家的なるものと『中間集団』との関係である。……**特に重要なのはカトリック教会との類似性であろう**」。

24

* *See*, Kate Klonick, *The New Governors: The People, Rules, and Processes Governing Online Speech*, 131 Harv. L. Rev. 1598（2018）.
** *See*, Frank Pasquale, *From territorial to functional sovereignty: the case of Amazon*, Open Democracy, 5 Jan 2018, <https://www.opendemocracy.net/en/digitaliberties/from-territorial-to-functional-sovereignty-case-of-amazon/>.; See also, Frank Pasquale, *Digital Capitalism–How to Tame The Platform Juggernauts*, Friedrich Ebert Stiftung, Jun 2018, <https://library.fes.de/pdf-files/wiso/14444.pdf>.
*** 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。

## 2．　ソーシャルメディアのヘイトスピーチ対策

（2）私たちは何を「見せさせられている」か

➢問題の一端としての「アルゴリズム」

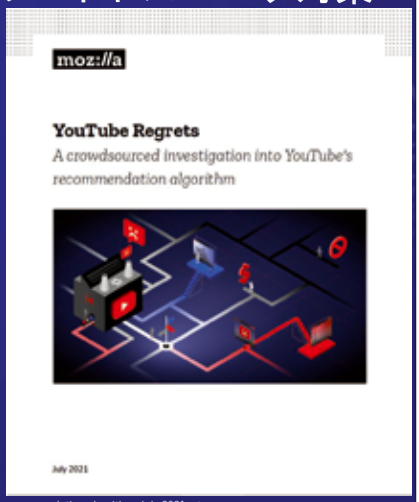
**Mozillaの研究レポート「YouTube Regrets」***

→YouTubeで視聴して「後悔（regret）」したコンテンツをボランティア・ユーザーに報告してもらったところ、以下の３点が判明。

①「後悔」コンテンツには様々なものが含まれる

**②レコメンドによって表示されたコンテンツが70%**

**③非英語圏での被害が多い**

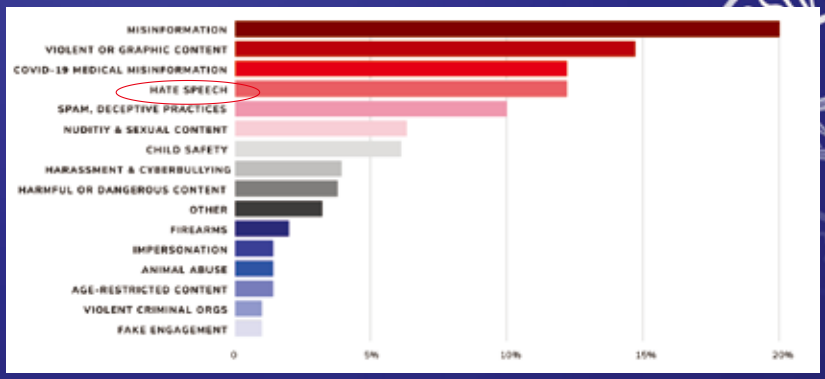



25

* See, Mozilla foundation, YouTube Regrets: A crowdsourced investigation into YouTube's recommendation algorithm, July 2021, at https://assets.mofoprod.net/network/documents/Mozilla_YouTube_Regrets_Report.pdf

## 2．　ソーシャルメディアのヘイトスピーチ対策

➢「後悔」動画の類別

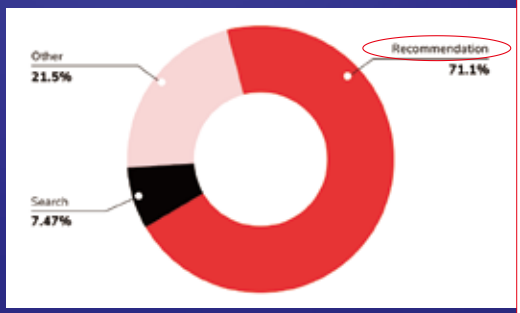


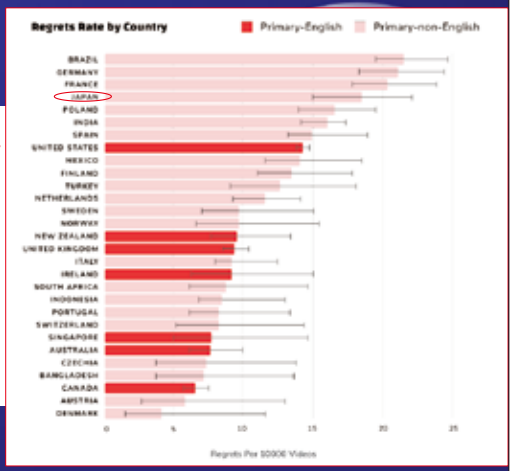

26

* Mozilla foundation, YouTube Regrets: A crowdsourced investigation into YouTube's recommendation algorithm, July 2021, p. 9, at https://assets.mofoprod.net/network/documents/Mozilla_YouTube_Regrets_Report.pdf

## 2．　ソーシャルメディアのヘイトスピーチ対策

（下）「後悔」動画をどのように受領したか
（右）「後悔」動画率の国別




27　* Mozilla foundation, YouTube Regrets: A crowdsourced investigation into YouTube's recommendation algorithm, July 2021, p.17, 22, at https://assets.mofoprod.net/network/documents/Mozilla_YouTube_Regrets_Report.pdf

## 2．　ソーシャルメディアのヘイトスピーチ対策

（2）私たちは何を「見せさせられている」か

**②レコメンドによって表示されたコンテンツが70%**

→重要なのは、こうした事態が、アルゴリズムの機械的誤作動によって生じているのではなく、**むしろ機械的には正常に動作した結果として引き起された「エラー」という可能性がある**という点*。

**③非英語圏の被害が多い**

→「ポリシー違反の検出や動画の推薦に使用されるアルゴリズムが、言語固有の機械学習モデルに依存しているため」ではないか？**

- エンゲージメントの強化策がネガティブ・コンテンツを引き寄せているのではないか？

⇔PF上のアルゴリズムを改築することでヘイトスピーチに対処可能になるのでは？

28　* Mozilla foundation, YouTube Regrets: A crowdsourced investigation into YouTube's recommendation algorithm, July 2021, p. 29, at https://assets.mofoprod.net/network/documents/Mozilla_YouTube_Regrets_Report.pdf
** Id, p. 20.

## 2．　ソーシャルメディアのヘイトスピーチ対策

（3）ヘイトスピーチ対策と「モデレーション」

➢**検閲の現代的形態―「検閲官」はどこにいる？**

→現代において政府（行政官）による「検閲」というものが大っぴらに行われることはほとんどない。

But…私たちの前から「検閲的なもの」は消えていない。

✓ミルトンの時代　⇒　行政の検閲官

✓原題　⇒　**企業が形成するオンライン・プラットフォームに潜むアルゴリズム（あるいはコード）や、ポリシーに基づいたコンテンツ・モデレーションの存在**

例：Google八分（Censorship by Google）

- アメリカにおける「**プラットフォームによる検閲**」論争*

* See, Will Oremus, Tech giants banned Trump. But did they censor him?, The Washington Post（Jan. 7, 2022）, at https://www.washingtonpost.com/technology/2022/01/07/trump-facebook-ban-censorship/

29

## 2．　ソーシャルメディアのヘイトスピーチ対策

（3）ヘイトスピーチ対策と「モデレーション」

➢私たちの日々のコミュニケーションを支えている「ソーシャルメディア」（SNS、動画共有サービス、検索エンジンetc…）では、日々、投稿されるユーザー生成コンテンツ（UGC）に対する「**モデレーション（監視及び適正化）**」が行われている。

→もちろん、膨大な数のヘイトスピーチにも日々対処している。

- **モデレーション＝削除やアカウント凍結“だけ”ではない**

→警告、表示ランクの操作、シャドーバン、ブロックetc…

30

KANSAI UNIVERSITY

# 3．「プラットフォーム法」の諸相

（1）コンテンツ・モデレーションとはなにか？

➢**コンテンツ・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示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」*

➢国連人権理事会特別報告者によるレポート**

「企業は、……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**」。

→国際人権法の観点から提言を行っている。

31

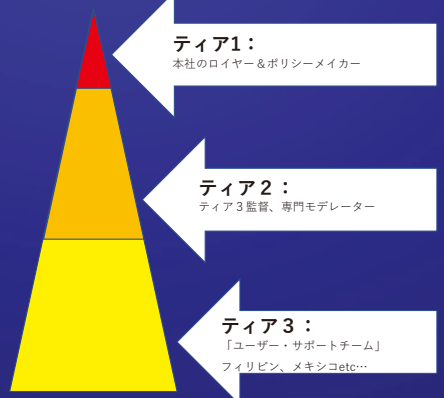
* Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, UN Doc A/HRC/38/35, 18 June-6 July 2018, pp. 3 footnote 2.
** Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, UN Doc A/HRC/38/35, 18 June-6 July 2018, pp. 3.

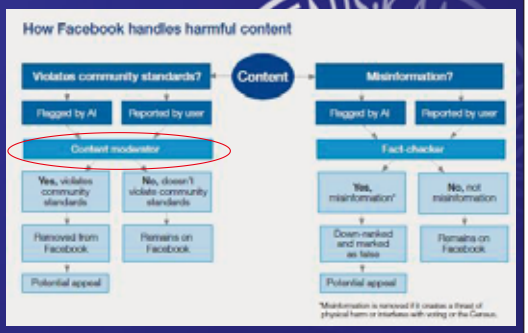
KANSAI UNIVERSITY

# 3．「プラットフォーム法」の諸相

（2）モデレーション・プロセス

(1) Facebookの例




- 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.

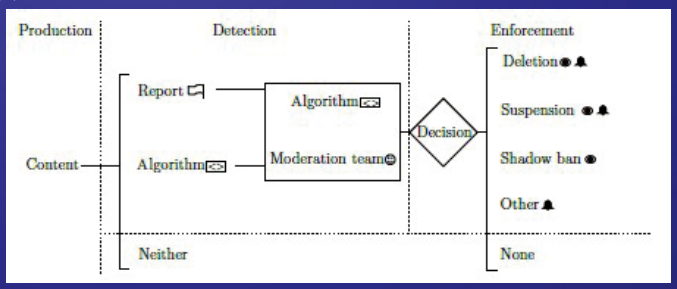
32

KANSAI UNIVERSITY

## 3．「プラットフォーム法」の諸相

（2）モデレーション・プロセス

(2) Twitterの例




- Twitterの対応措置*をみるとわかるが、ツイートやアカウントの削除以外にも、停止措置、シャドーバン（+ラベリング）など、モデレーションにも多様な手段があることが分かる。

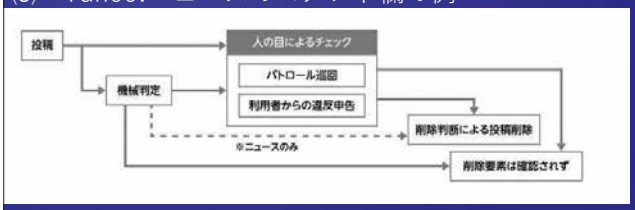
33

※図は、Rafael Jiménez Durán, The Economics of Content Moderation: Theory and Experimental Evidence from Hate Speech on Twitter (Nov 2021), The University of Chicago booth school of business, the George J. Stigler Center for the Study of the Economy and the State, New Working Paper Series No. #324, pp.14, <https://www.chicagobooth.edu/-/media/research/stigler/pdfs/workingpapers/324jmpjimenezduran003.pdf>より引用.
* Twitter, Our range of enforcement options, 2021. <https://help.twitter.com/en/rules-and-policies/enforcement-options>.

KANSAI UNIVERSITY

## 3．「プラットフォーム法」の諸相

(3) Yahoo!ニュースのコメント欄の例*




- 人間のモデレーション・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つのモデルを採用している。

34 * 以下は、Yahoo！JAPAN「透明性レポート」（https://about.yahoo.co.jp/common/transparencyreport/）を参照。画像も同様。

## 3．「プラットフォーム法」の諸相

（3）「プラットフォーム法」とヘイトスピーチ

➢**Facebook コミュニティ規定**

→**真正性、安全、プライバシー、尊厳**の４つの価値とのバランシング

**ヘイトスピーチ**：「概念や制度ではなく、人種、民族、国籍、障がい、宗教、社会階級、性的指向、性別、ジェンダーアイデンティティ、重度の病気など、保護特性と呼ばれるものを理由に人々を直接攻撃すること」*。

レベル１：記述または視覚的な形での暴力的な発言や支援、昆虫・動物・汚物・ウイルス等による「(文章または視覚による)比喩、安易な一般化、不適切な見解による主張の形での人間性を否定するような発言や画像」、ヘイトクライム被害者への嘲笑

レベル２：身体的・精神的・道徳的欠陥に対して「劣等感を与える安易な一般化」、役立たず等能力不足・特性の優劣・変人等の普通でないことを意味する表現で「劣等感を与える発言」、侮辱表現、「人を見下げる表現」、嫌悪表現、暴言

レベル３：行動の呼びかけ等による差別・排除、「人を標的にして中傷するコンテンツ」

35

* https://transparency.fb.com/ja-jp/policies/community-standards/hate-speech/

## 3．「プラットフォーム法」の諸相

➢**YouTubeコミュニティ・ガイドライン**

**①スパムと欺瞞行為、②デリケートなコンテンツ、③暴力的または危険なコンテンツ、④規制品、⑤誤った情報***

③→　**ヘイトスピーチに関するポリシー****

- 「**次のいずれかの特性に基づいて個人や集団に対する暴力や差別を助長するコンテンツは削除されます**」。

→年齢、カースト、障がい、民族、性同一性や性表現、国籍、人種、在留資格、宗教、性別 / ジェンダー、性的指向、深刻な暴力的出来事の被害者とその親族、**従軍経験**

- 「上記の特性に基づいて個人や集団に対する暴力を助長する」行為、「上記の特性に基づいて個人や集団に対する憎悪を扇動する」行為は、禁止される。
- その他のコンテンツ：人間以外の存在に例える、暴力称賛・美化、人種・宗教等の固定観念等を使用した憎悪の扇動・助長、性的関心への攻撃、「**歌詞、メタデータ、画像の中で差別的な至上主義を奨励するミュージック ビデオ**」

36

* https://support.google.com/youtube/answer/9288567?sjid=8633766659481419180-AP
** https://support.google.com/youtube/answer/2801939?sjid=8633766659481419180-AP

KANSAI UNIVERSITY

## 3．「プラットフォーム法」の諸相

➢Tiwtter ルール

**①安全、②プライバシー、③信頼性**

(1)暴力的な発言、(2)暴力行為やヘイト行為の主体、(3)児童の性的搾取、(4)攻撃的な行為/嫌がらせ、**(5)ヘイト行為**、(6)暴力行為の加害者、(7)自殺、(8)センシティブなメディア、(9)違法または特定の規制対象商品・サービス*

**「人種、民族、出身地、社会的地位、性的指向、性別、性同一性、信仰している宗教、年齢、障碍、深刻な疾患を理由とした他者への直接的な攻撃行為」****

- **ヘイト行為への言及:**「Twitterは、個人または集団を、一部の国や地域で規定されている保護対象のカテゴリーの人々が主な標的または犠牲者となった暴力や暴力事件をほのめかし、攻撃する意図を持ったコンテンツの標的にすることを禁止しています」。
- **煽動:**「Twitterは、法的または社会的に守られるべき特定のカテゴリーに属する個人または集団を標的とする行為の扇動を禁止しています」。
- **中傷や差別的揶揄:**「Twitterは、他者を繰り返し中傷すること、差別的揶揄の対象にすること、そして法的または社会的に守られるべき特定のカテゴリーの人々を貶めたり、そのような人々に対する否定的または有害な偏見を助長したりすることを目的とした、その他のコンテンツの標的にすることを禁止しています」。
- **人間性の抹消:**「また、宗教、社会的地位、年齢、障碍、深刻な疾病、出身地、人種、民族、性別、性同一性、性的指向を理由に特定の集団を非人間的に扱う行為を禁止しています」。
- **ヘイト表現を伴う画像:**「人種、宗教、障碍、性的指向、性同一性、民族/出身国を理由に他者に対して敵意や悪意を増幅させることを目的とするロゴ、象徴、画像は、ヘイト表現を伴う画像とみなします」。
- **ヘイト表現を伴うプロフィール:**「ヘイト表現を伴う画像や象徴をプロフィールの画像またはプロフィールのヘッダー画像に使用してはなりません。また、ユーザー名、表示名、プロフィールを利用して、攻撃的な行為を取ることを禁じます」。

37

* https://help.twitter.com/ja/rules-and-policies/twitter-rules
** https://help.twitter.com/ja/rules-and-policies/hateful-conduct-policy

KANSAI UNIVERSITY

## 3．「プラットフォーム法」の諸相

➢ **Yahooニュース　コメントポリシー***

差別的発言、ヘイトスピーチ

- ***特定の地域や家柄、性別、性的指向、性自認、病気、障がい、職業、宗教、信仰などへの差別的な内容を含む投稿***
- ***特定の人種や民族、国や地域に対する差別やヘイトスピーチにあたる投稿***

投稿例

- *人種、民族を理由に、優劣をつけたり、知能が劣る、野蛮であるなどとしたりする投稿は禁止しています。*

*「○○人は知能が低い」「土人」*

- *特定の国の出身であることやその子孫であることを理由に、合理的な理由なく社会から追い出そうとする投稿や、危害を加えようとする投稿は禁止しています。*

*「○○人は日本から出ていけ。」「○○人は強制送還したほうがいい。」*

- *特定の国や地域の出身である人を、著しく見下し、昆虫や動物に例える投稿は禁止しています。*

38

* イタリック部分は、以下より引用　https://news.yahoo.co.jp/info/comment-policy

## 3．「プラットフォーム法」の諸相

（4）プラットフォーム法の執行状況

**➢Metaコミュニティ規定の執行***

• 2023年1月～3月期

ヘイトスピーチ表示頻度：0.2-0.1%

ヘイトスピーチ対応数：**約1070万件**


* 画像も含め、以下のウェブサイトから抜粋。 https://transparency.fb.com/ja-jp/policies/community-standards/hate-speech/

39

## 3．「プラットフォーム法」の諸相

（4）プラットフォーム法の執行状況

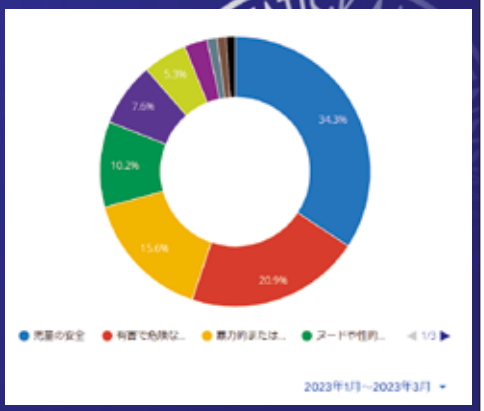
**➢YouTubeコミュニティ・ガイドラインの執行***

• 2023年1月～3月期

動画削除件数全体：6,487,896件

→1位はインド（1,906,071）、2位はアメリカ（654,968）、3位はロシア（491,933）、**日本は24位（36,531）**

そのうちヘイトスピーチ（悪意ある表現等）関連：**177,921件(2.7%)　※右の円グラフ赤紫部分**




* 画像も含め、以下のウェブサイトから抜粋。 https://transparencyreport.google.com/youtube-policy/removals?hl=ja

40

KANSAI UNIVERSITY

## 3．「プラットフォーム法」の諸相

（4）プラットフォーム法の執行状況

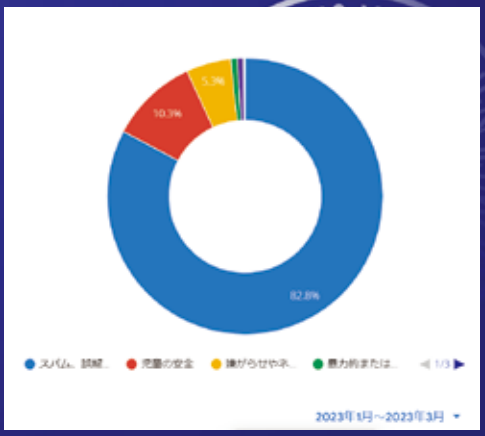
**➢YouTubeコミュニティ・ガイドラインの執行***

- 2023年1月～3月期

→コメント削除件数全体：853,275,342件

そのうちヘイトスピーチ関連：

**6,004,883件（0.7%） ※右の円グラフ紫部分**




* 画像も含め、以下のウェブサイトから抜粋。 https://transparencyreport.google.com/youtube-policy/removals?hl=ja



KANSAI UNIVERSITY

## 3．「プラットフォーム法」の諸相

**➢Twitterルールの執行**

- 2021年7月～12月期*

削除コンテンツ全体：5,103,156件

→そのうちヘイトスピーチ行為：

**1,293,178件**

**※なお2022年1～6月期については透明性レポートが出ていないが、数値は公表されている**。**

削除コンテンツ全体：**6,586,109件**

→そのうちヘイトスピーチ行為：

**1,527,442件**


* 画像も含め、以下のウェブサイトから抜粋。 https://transparency.twitter.com/ja/reports/rules-enforcement.html#2021-jul-dec
** https://blog.twitter.com/en_us/topics/company/2023/an-update-on-twitter-transparency-reporting

KANSAI UNIVERSITY

# 3．「プラットフォーム法」の諸相

（4）プラットフォーム法の執行状況

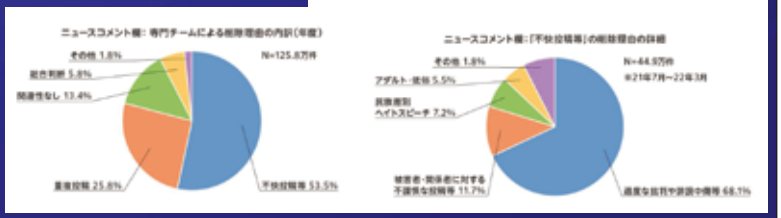
➢**ヤフーニュース・コメントポリシーの執行**

• 2021年度

→コメント削除件数全体：513.1万件

専門チームによる削除理由「不快投稿等」のうち、

ヘイトスピーチ関連：**7．2％**


43

*画像も含め、以下のウェブサイトから抜粋。 https://about.yahoo.co.jp/common/transparencyreport/

KANSAI UNIVERSITY

# 3．「プラットフォーム法」の諸相

| Element（要素） | Description（説明） | Examples（例示） |
|---|---|---|
| Contract Law<br>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 |
| Common Law (Norms)<br>コモン・ロー（規範） | 予測される規範、手順、制裁に関するコミュニケーション | コミュニティの基準や規則（ポリシーの根拠や例を含む）<br>ポリシーとアクションの説明（投稿、声明）<br>ユーザーに送られた警告、通知、アクションの説明<br>透明性に関する報告 |
| | 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<br>ユーザーに制裁を加えるかについての決定<br>不服申し立てに対する決定<br>特定のコンテンツに関する専門機関の決定 |
| | 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 | 内部方針と文書<br>モデレーターの研修資料<br>会議の記録 |
| Common Law (Procedures)<br>コモン・ロー（手続き） | 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 | ユーザーによるフラグ立て<br>アルゴリズムによるフラグ立て<br>モデレーター審査<br>コミュニティモデレーター<br>ユーザーモデレーター (reddit karma, 4chan janitors)<br>エスカレーションの手順と基準<br>制裁の手順と基準<br>不服申し立てプロセス |
| Technical Law<br>技術法 | 技術上およびシステム設計上の選択 | どのような投稿が許可されているか（リンク、テキスト、写真）<br>どの程度のコンテンツを許可するか<br>実名制<br>コンテンツモデレーターのアルゴリズム<br>アルゴリズムによる嗜好（「いいね！」）のメカニズム<br>コンテンツを共有するためのツール（リツイート、シェア）<br>ユーザーがコンテンツをコントロールできるツール |

44

* Molly K. Land, The Problem of Platform Law: Pluralistic Legal Ordering on Social Media (Sept 15, 2019), at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3454222, pp.6-7.

KANSAI UNIVERSITY

## 3．「プラットフォーム法」の諸相

（5）なぜ「プラットフォーム法」が構築されたのか？

- 「粘着性」（＝より長く、繰り返しPFを利用してもらう）が重視されるアテンション・エコノミーのもとでは、より多くのユーザーに「**安心・安全**」な空間を提供する必要がある（**トラスト＆セーフティ**）。
- 広告主に対するブランド維持（**ブランドセーフティ**）。

**→経済的必要性**

➢モデレーションはDPF事業者にとって商品であるコンテンツの「品質管理」であり、ポリシーはそのための「品質決定指針」として機能（≠職人的　＝生産ライン的）。

- 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*

➢DPF事業者＝「新たな統治者」≠政府

➡違法情報対策に関しては、「公共の福祉」の主たる担い手である政府による一定程度の関与も必要。

45

* 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。

KANSAI UNIVERSITY

## 4．システミック・アプローチに向けて

（1）権利ベース・アプローチの限界

表現の自由の「権利」に対する「規制」

国家権力（警察など）による違法な表現の取り締まり

裁判所による規制立法の違憲審査＆規制された個人の救済

判決によっては、法律改正につながったり、事後、同じような表現を警察等が取り締まらないよう影響することも。

民主的な議会による表現規制立法の制定

※画像はいらすとやより
https://www.irasutoya.com/

46

KANSAI UNIVERSITY

# 4．システミック・アプローチに向けて

（1）権利ベース・アプローチの限界

被害者

誹謗中傷？？

発信者

名誉権やプライバシー権等の「権利」を主張

表現の自由の「権利」を主張

裁判所による「権利」と「権利」の衝突に対する民事上の裁定（不法行為訴訟）

※画像はいらすとやより
https://www.irasutoya.com/

47

KANSAI UNIVERSITY

# 4．システミック・アプローチに向けて

（1）権利ベース・アプローチの限界

➢アルゴリズミック・コンテンツ・モデレーション*

- 例えば、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。

➢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

- 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****。
- またYahooニュースコメント欄で、**約7割**をAIが自動処理で削除している*****。

→必然的に生じるエラー（過少or過剰執行）。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。**

* *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*, 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/>.
***** Yahoo！JAPAN「メディア透明性レポート（2021年度版）」<https://about.yahoo.co.jp/common/transparencyreport/>を参照.

48

## 4．システミック・アプローチに向けて

（1）権利ベース・アプローチの限界

➢こうした**AIの積極的な導入**により、コンテンツ・モデレーションにおける「**プラットフォームの権利裁定を “保険数理的（actuarial）”なものにしている**」*。

- この「自動化」はおそらく完全ではないが、今後も推進されていく可能性が高い。

「アルゴリズムの従業員は、人間の従業員よりもさらにコストが低くなる。彼らは家族を持たず、コーヒーブレイクもせず、人間の従業員ができる発見や選択の作業の一部（ただし、すべてではない）を行うことができる」**。

**➡機械学習の限界（過剰削除、文脈無視）**

49

* See, Evelyn Douek, *Governing Online Speech: From 'Posts-As-Trumps' to Proportionality and Probability*, 121 COLUM. L. REV. 759, 797 (2021).
** Jack M. Balkin, *Free Speech Is a Triangle*, 118 Colum. L. Rev. 2011, 2024（2018）.

## 4．システミック・アプローチに向けて

➢人間によるモデレーターを増やせばよいか？

- Yahoo！JAPANで約70名*、Facebookで約15000人**、Twitterで約2000人***ともいわれており、その多くがアウトソーシングされている。

＋「**ソーシャルメディア掃除人**」の人道的観点****

「弁護士にとっては、個人主義的で、事後的なエラー訂正の枠組みは、言論紛争についての議論を行う上でお馴染みのフォーマットである」が、「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*****。

50

* Yahoo！JAPAN「メディア透明性レポート（2020年版）」<https://about.yahoo.co.jp/common/transparencyreport_2020/>を参照。
** Meta「Metaのコンテンツ、安全性、透明性へのアプローチ」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第3回（2023年3月3日）配布資料4 <https://www.soumu.go.jp/main_content/000866254.pdf >を参照。
*** Twitter社のYoel Roth氏は、2022年11月5日のツイートで、前線で審査業務に従事しているモデレーターの数を「2000人以上」と呟いている（See, https://twitter.com/yoyoel/status/1588657229628321792）
**** 動画は以下のウェブサイトから抜粋。 BBC News Japan「「処刑映像を何百回も見た」　SNSの掃除人が抱える苦悩」YouTube（2018年10月17日）at https://www.youtube.com/watch?v=zQZYKh7t4x8
***** Evelyn Douek, Content Moderation as Systems Thinking, 136 Harv. L. Rev. 526, 606 (2022).

## 4．システミック・アプローチに向けて

（2）システミック・アプローチへ

➢インターネット・ガバナンスは、権利の時代（The Rights Era）から**公衆衛生の時代（The Public Health Era）**に移行しつつあるとの指摘がある*。

→「個人の権利よりもリスクと利益を重視」

- コンテンツ・モデレーションの世界は、個人の権利ベースの世界ではなく、**確率論（probabilistic）とシステミックな世界****

→例えば、単純に特定の「ワード」をシステム上禁止した場合、どうしても過剰規制や過少規制の可能性（＝エラー）が生じてくる。

➡システミックなアプローチでは、権利ベースのアプローチと異なり、**必然的に生じるエラーを許容したうえで、それをガバナンス設計に組み込むことが重要**となる***。

51

* See, Jonathan Zittrain, *Three Eras of Digital Governance* (Sept. 23, 2019), at https://www.ssrn.com/abstract=3458435
** See, Evelyn Douek, *Governing Online Speech: From 'Posts-As-Trumps' to Proportionality and Probability*. 121 COLUM. L. REV. 759 (2021)
*** Id, at 791.

## 4．システミック・アプローチに向けて

（3）示唆

➢条例レベルでインターネット上の削除要請には、該当コンテンツの規模からいっても、限界があるのでは？

→国レベルでの対応が不可避。むしろ自治体は、デモや集会の規制に注力すべき、か？

- 国レベルでの対応の場合にも、オーバーブロッキングや濫用可能性を考慮し、政府とPFの間の「適切な距離」を維持する必要がある。

→DPF事業者の「**検閲代理人**」化を防ぐため。

➡必要なのは、政府とDPF事業者の**協調的ガバナンス（あるいは共同統治）モデル**

52

## 4．システミック・アプローチに向けて

（おまけ）政府からの削除リクエスト

「**Google では常に、政府からのリクエストの正当性と完全性を評価します。**リクエストが Google によって適切に評価されるためには、リクエストを書面で行うこと、削除するコンテンツをできる限り具体的に示すこと、そのコンテンツが違法である理由を明確に説明することが重要です。不適切な方法によるリクエストは受理されません。リクエストを口頭で受けた場合は、書面で提出していただくようお願いしています。」*

**日本政府からのリクエストの内容例****

- リクエスト

県警のある巡査長から、業務時間の休憩中に女性を盗撮したかどで逮捕され罰金刑を受けたことを取り上げた URL を除外するよう求められました。

- 結果

Google では、当該 URL を除外しませんでした。

53

* https://transparencyreport.google.com/government-removals/overview
** https://transparencyreport.google.com/government-removals/government-requests/JP

## 4．システミック・アプローチに向けて

（おまけ２）日本の法務省、Googleに「公認」される

「法務省人権擁護局は、Google LCC のサービス『YouTube』について、同社が提供する『YouTube 公認報告者プログラム』※に参加することとなりました」*。

→公認報告者は、YouTubeからガイドライン違反のコンテンツを報告するための様々なツールを使えるようになる。

＋「**公認報告者から報告されたコンテンツは、自動的に削除されたり、異なるポリシーによって対処されたりするわけではありません。他のユーザーから受け取った報告と同じ基準が適用されます。**ただし、その高い正確性により、YouTube 公認報告者からの報告は優先的に審査されます。」**

54

* http://www.moj.go.jp/JINKEN/jinken03_00084.html
** https://support.google.com/youtube/answer/7554338?hl=ja

## 4．システミック・アプローチに向けて

（3）示唆

「ソーシャルメディアにおける強固なヘイトスピーチ規制の出現は、異常なことではなく、より広範な文化的シフトの産物である。プラットフォーム法は、憲法修正第1条の法理よりも、ユーザーの社会規範や、ヘイトスピーチに対する米国と欧州のアプローチの橋渡しをしたいという企業の願望によって推進されてきた」*。

➢プラットフォーム法によるヘイトスピーチ規制の問題**

①不透明さ、②文脈の無視

➡「**手続上の公正さ、透明性、情報開示を優先し、人権監査及びAIシステムにおける潜在的差別の定期的審査へのコミットメントを意味する、AIへの人権的アプローチ**」***

55

* See, Richard Ashby Wilson & Molly K. Land, Hate Speech on Social Media: Content Moderation in Context, 52 CONN. L. REV. 1029, 1055 (2021).
** Id, at 1056-1060.
*** Id, at 1071.

## 4．システミック・アプローチに向けて

（3）示唆

❶アカウンタビリティ担保のための透明性確保

→禁止表現のローカライズ・リスト、自動対応率

❷事前のモデレーション年次計画策定、継続的な透明性報告書、外部監査

→DPF事業者のコンテンツ・モデレーションのPDCAに対するガバナンス

＝ヘイトスピーチをどれだけ取り締まっているか、ではなく、**個々のDPF事業者がどれだけきちんと計画を立てているか、目標値を達成できているかという視点へ**のシフト

+公法上の義務としての削除請求制度など、さらに踏み込んだ仕組みを取り入れるならば、**政府自身も（PF対応に関する）透明性を確保しなければならない**。

56


KANSAI UNIVERSITY
ご清聴ありがとうございました。
57