

インターネット上の違法・有害情報対策 に関する諸外国動向

MRI 三菱総合研究所

2026年6月15日

モビリティ・通信政策本部

デジタルコンテンツ・データ戦略グループ

目次

1. EUの動向	2
1.1 EU・プラットフォーム事業者等の取組状況	
1.2 EU・DSAに基づくプラットフォーム事業者への主な対応	
2. ドイツ・英国の動向	8
2.1 ドイツ・デジタル暴力対策法案の構想	
2.2 英国・Ofcomによる調査	
3. 韓国の動向	11
3.1 韓国・情報通信網法改正の概要	
3.2 韓国・情報通信網法改正に対する反応・評価	
4. 米国の動向	14
4.1 米国・TAKE IT DOWN Actの概要	
4.2 米国・アルゴリズム説明責任法案の概要	

1. EUの動向

- 1.1 プラットフォーム事業者等の取組状況
- 1.2 DSAに基づくプラットフォーム事業者への主な対応

(1) 各事業者の取組状況

- デジタルサービス法(DSA)では、違法コンテンツへの対応を含むプラットフォームのコンテンツモデレーション活動について、透明性の確保を通じて説明責任を強化するため、透明性レポートの公表が義務付けられている。
- 超大規模オンライン・プラットフォーム(VLOP)および超大規模オンライン検索エンジン(VLOSE)の提供者については、DSA第15条、第24条、第42条に基づき、透明性レポートの公表が義務付けられている。

公表項目*1	DSAの根拠条項	対象事業者
1. 加盟国当局からの命令件数(削除、情報提供)	第15条(1)(a)	仲介サービス提供者*2
2. 違法コンテンツに関する通知件数(信頼された旗手からの通知を含む)	第15条(1)(b)	
3. 自発的なコンテンツモデレーションの実施状況(実施した措置の内容を含む)	第15条(1)(c)	
4. 苦情処理メカニズム	第15条(1)(d)	
5. 自動化手段の利用状況	第15条(1)(e)	
6. 裁判外紛争解決手続きの状況、不正利用に対する措置	第24条	オンラインプラットフォーム提供者
7. コンテンツ・モデレーション人員	第42条(2)(a),(b)	VLOP/VLOSEの提供者
8. 平均月間アクティブユーザ数	第42条(3)	

※ 以下では、DSAの規制対象となる違法情報及び利用規約違反情報に対するプラットフォーム事業者の取組状況を端的に把握する上で有用と考えられる上記3. 及び5.について、主要事業者の透明性レポート(2025年8月公表版)の内容を整理して示す。

*1 <https://digital-strategy.ec.europa.eu/en/library/implementing-regulation-laying-down-templates-concerning-transparency-reporting-obligations>

*2 ネットワークインフラを提供する仲介サービス。インターネットアクセスプロバイダ、ドメイン名レジストラの他、ホスティングサービス(クラウドやウェブホスティングなどのホスティングサービス、オンライン・プラットフォーム、等)も含む。

(2) 自発的なコンテンツモデレーション(抜粋)

公表内容: 自発的なコンテンツモデレーションについて(定性情報)

区分	Meta	Google	TikTok	X
自発的に実施したコンテンツ・モデレーションの概要	<ul style="list-style-type: none"> グローバルに適用されるコミュニティ規定を中核に、広告、コマース、収益化のそれぞれについてポリシー・標準規定を設け、許容されるコンテンツや行為の基準を定めている。 これらのポリシーに基づき、<u>自社の判断によりコンテンツやアカウントの検出、制限、削除等の措置を実施する。</u> 	<ul style="list-style-type: none"> 各サービスにおいてコンテンツ・ポリシーを定め、許容されるコンテンツや収益化の基準、開発者によるデータの取扱い等のルールを明確化している。 これらのポリシーに基づき、<u>安全で信頼性の高い利用環境の確保を目的として、コンテンツの作成・共有・表示に対するモデレーションおよび執行措置を実施している。</u> 措置は、<u>アルゴリズムや従業員等による検知に基づく能動的対応と、ユーザ通報や法的苦情に基づく受動的対応の双方を含む。</u> 	<ul style="list-style-type: none"> 利用規約、コミュニティガイドライン、広告ポリシー、TikTok Shopポリシー等に基づき、違反コンテンツやアカウントに対して削除、制限、年齢制限等の措置を自発的に実施している。 モデレーションは、①ポリシー違反コンテンツの削除または違法コンテンツの制限、②成熟コンテンツの年齢制限、③推薦アルゴリズム(For Youフィード)における適格性基準の維持、④ユーザーへの情報・ツール提供、の4つの原則に基づいて運用される。 	<ul style="list-style-type: none"> 「<u>公共的会話</u>」を支えることを目的とし、<u>表現の自由を不必要に制限することなく、体系的リスクを軽減するよう設計されたコンテンツ・モデレーション体制を採用している。</u> 措置は、個別の投稿、DM等のコンテンツ、アカウント単位で講じられる。
自発的に実施したコンテンツ・モデレーションに関する有意義で理解しやすい情報	<ul style="list-style-type: none"> <u>自動化と人のレビューを組み合わせて実施する。</u> <u>自動化技術は、違反性の高いコンテンツを検知し、場合によっては自動で削除・制限し、必要に応じて人手審査に回付する。</u> レビュー対象は、単に違反の有無だけでなく、重大性、拡散性、違反可能性に基づき優先順位付けされる。 	<ul style="list-style-type: none"> 動画、URL、画像、リスティング、アカウント等に対する削除、非表示化、アクセス制限など、可用性、可視性、アクセス性に影響を与える措置を実施している。 サービスごとに特性が異なるため、指標の粒度や集計方法が異なる。 	<ul style="list-style-type: none"> <u>違反コンテンツの大半を事前・自動的に検知する「プロアクティブ検知」に重点を置いている。</u>自動化技術により検知されたコンテンツは、削除・制限されるか、又は人手レビューに回付される。 外部専門機関(ファクトチェック機関、Tech Against Terrorism等)との連携により、検知精度の向上と迅速な対応を図っている。 	<ul style="list-style-type: none"> <u>ヒューリスティクスと機械学習アルゴリズムを併用して、ルール違反の疑いがあるコンテンツを検知している。</u> 成人向けコンテンツ、暴力的コンテンツ、児童性的搾取、暴力的・憎悪的団体等に関するポリシー違反は、優先して対応している。 悪意のある主体による回避行動や協調的操作を把握するため、大規模な人手調査も補完的に行う。

(3) 自動化手段とその精度(抜粋)

公表内容: 自動化手段に関する情報(定性情報)のポイント(i)

区分	Meta	Google	TikTok	X
自動化手段について	<ul style="list-style-type: none"> アカウント、投稿、コメント、写真その他のアップロードされたコンテンツに自動化手段を適用している。 主な手段は、レート制限(ボットの挙動の防止目的)、マッチング技術(違反コンテンツの再出現防止)、AIを用いている。 	<ul style="list-style-type: none"> スパムや明確なポリシー違反の検知・処理に適用する一方、文脈判断が必要な場合には人手によって判断する。複雑な事案は専門チームにエスカレーションする。 人手で判断された結果は、機械学習モデルの訓練データとして活用され、継続的な精度向上が図られている。 YouTubeでは、違反コンテンツの大半を自動で検出し、必要に応じて人手で確認する。 	<ul style="list-style-type: none"> 自動化技術を用いて、投稿前および投稿後のコンテンツやアカウント行動を分析し、ポリシー違反の可能性を評価する。 主な手段には、画像認識、音声分析、自動言語処理、類似性検知、行動分析、LLMおよびマルチモーダルLLMが含まれる。 	<ul style="list-style-type: none"> 機械学習とヒューリスティクスを用いて自動検知を行なっている。 機械学習には、自然言語処理モデル、画像処理モデル、その他の高度な手法を用いており、投稿文、添付画像、その他の特徴量を入力として、ルール違反の可能性を判定する。 ヒューリスティクスは、新しい違反形態への迅速な対応を可能にするため、特定の違反類型に典型的な行動・文章・キーワード等のパターンを用いてコンテンツを抽出する仕組みとなっている。
自動化手段の適用目的	<ul style="list-style-type: none"> 違反コンテンツの迅速な検知・制限・削除 優先順位付けによる人手レビューの負担軽減 広告資産のポリシー適合性確保(不適切配信の防止) 	<ul style="list-style-type: none"> 大規模環境での迅速な検知とユーザ保護等のため YouTubeでは拡散防止のため 	<ul style="list-style-type: none"> 迅速な検知・削除による有害コンテンツの露出軽減 レコメンダー・システムの質的維持 ファクトチェック支援など文脈依存コンテンツの評価効率化 	<ul style="list-style-type: none"> システミック・リスクの軽減と、テロ、プラットフォーム操作、その他高リスク行為への早期対応 表現の自由を不必要に制限しない適正な執行

(3) 自動化手段とその精度(抜粋)

公表内容:自動化手段に関する情報(定性情報)のポイント(ii)

区分	Meta	Google	TikTok	X
精度の指標および潜在的誤差率に関する説明	<ul style="list-style-type: none"> 担当者による手動の違反ラベル付けを、アルゴリズムやマッチング技術の継続的な改善に活用している。 新しい技術は即時導入ではなく、まずは動作記録と人手レビューによる検証を行う。導入後も措置件数や異議申立率を監視し、異常時はエンジニアが調査する。 レート制限はボット対策の有効性を再評価し、マッチング技術は参照データの妥当性を確認する。AIはサンプルの再ラベル付けで精度を検証し、必要なら廃止する。 無作為抽出調査による品質管理に加え、全体の指標として「自動化措置の取消率」を用いている。 	<ul style="list-style-type: none"> 人手レビューを基準に精度を評価し、誤検知は是正のうえ学習データへフィードバックすることでモデルを継続的に改善する。 ローカルレビューは自動削除の0.26%が異議申立により取り消された。 広告は自動措置の0.17%が人手審査を経て取り消された。 YouTubeは、自動検知の有効性指標として、削除前にユーザに視聴された割合を示す「違反コンテンツ視聴率」を用いている。 	<ul style="list-style-type: none"> 精度指標として「措置判断が維持された割合」を用い、誤差指標として「措置が覆された割合」を用いている。 自動化技術は人手判断の結果を学習データとして継続的に改善されるほか、異議申立の結果等も制度評価に反映される。 	<ul style="list-style-type: none"> 機械学習モデルやヒューリスティックスは、訓練済みモデルデータによる数千件のラベル付けデータで訓練・検証を行う。導入直前には統計的に有意なサンプルを抽出し、人手による項目別のレビューで検知精度を確定させる。 処理量、苦情数、取消率の変動を常時監視している。異常検知時は、エンジニアリング・データサイエンスチームが他部門と連携して原因を診断し、自動化ロジックの最適化を図る。 自動・人手双方の誤差を図る指標として、コンテンツ削除に対する異議申立数と執行判断が覆った件数を集計している。
適用される保護措置	<ul style="list-style-type: none"> 人手レビューの併用を基本とし、導入前検証から導入後監視、異常時の再評価まで全段階に人間が関与する。 手動ラベル付けによるフィードバックループと、<u>無作為監査による品質管理を徹底している。</u> 	<ul style="list-style-type: none"> <u>文脈(教育・報道など)に応じた人手判断を重視している。</u> YouTubeでは<u>再アップロード防止技術と人手による最終判断を組み合わせている。</u> 	<ul style="list-style-type: none"> <u>違反が明白な場合のみ自動削除を行い、疑わしい事案は人手レビューへ回付する。</u> <u>外部事実チェック機関との連携や異議申立による是正プロセスを通じて正確性と公平性を担保する。</u> 	<ul style="list-style-type: none"> 統計的サンプルに基づく導入前検証を必須とし、十分な精度確認後に技術を本番導入する。 導入後も苦情数や取消率の変動をエンジニアが監視し、<u>異議申立制度によって表現の自由とのバランスを図る。</u>

欧州委員会による対応

- 2025年12月5日、欧州委員会はDSAの義務規定の違反を理由に、Xに対し1億2,000万ユーロの制裁金を賦課する決定を下した。^{*1} 同決定は、DSAの施行後初めての不遵守決定であり、同法に基づく制裁金が実際に科された初の事例である。具体的には、「青いチェックマーク」の欺瞞的設計、広告リポジトリの透明性欠如、研究者のデータアクセス不備という複数の違反が認定された。
- また、2026年1月には、Xに統合された生成AI機能「Grok」に関し、違法コンテンツの拡散リスク等を理由としてDSAに基づく正式調査を開始したことが公表された。^{*2}

欧州委員会による手続きの経緯

- 2023年12月18日：上記の違反疑いに関する正式手続きの開始。
- 2024年7月12日：上記の違反を認定する予備調査結果の通達。
- 2025年12月5日：DSAの義務規定への非遵守を正式に決定し、制裁金の賦課を執行。
- 2026年1月26日：生成AI機能「Grok」についても、DSA義務違反の疑いで新たな正式手続きを開始する（Xについて実施中のレコメンダー・システムに関するリスク評価・軽減義務違反疑いの調査範囲を拡張する）ことを公表。

Xに対するDSA違反の認定のポイント

- **認証アカウントの欺瞞的デザイン**：Xでは認証済みアカウントに「青いチェックマーク」を付与しているが、実質的な本人確認は行っていない。その結果、ユーザに認証済みとの誤解を生じさせ、詐欺やなりすましのリスクを高めている。
- **広告リポジトリの不備**：Xの広告リポジトリは、検索性の低さや広告内容、広告主情報の不足、過度なアクセス遅延などにより、透明性やアクセシビリティの要件を満たしていない。
- **研究者アクセスの不備**：Xは、研究者による公開データへのアクセスを制限しており、利用規約においてスクレイピングを禁止しているほか、アクセス権申請のプロセスにおいて不必要な障壁を設けている。これにより、システムック・リスクに関する研究が実質的に妨げられている。

Grokに関する調査のポイント

- **システムック・リスクの評価、軽減措置義務**：違法コンテンツや性的暴力、健康被害等への影響を適切に評価し、軽減する措置を講じているか。(DSA第34条、第35条)
- **提供開始前の臨時評価実施義務**：Grokの機能の提供を開始する前に、適切なリスク評価を実施し、欧州委員会が求めた情報を提供していたか。(DSA第34条、第42条)
- **レコメンダー・システムによるリスク増幅**：Grokをベースとしたレコメンダー・システムの導入による影響を含め、レコメンダー・システムに関連するリスクを適切に評価し軽減しているか。(DSA第34条、第35条)

*1 https://ec.europa.eu/commission/presscorner/detail/en/ip_25_2934

*2 https://ec.europa.eu/commission/presscorner/detail/en/ip_26_203

2. ドイツ・英国の動向

- 2.1 ドイツ・デジタル暴力対策法案の構想
- 2.2 英国・Ofcomによる調査

2.1 ドイツ・デジタル暴力対策法案の構想

デジタル暴力対策法案の構想

- ドイツ連邦司法・消費者保護省(BMJV)は、ディープフェイク、サイバーストーキング、リベンジポルノ等の「デジタル暴力」が急増していることを踏まえ、2026年4月17日に「デジタル暴力対策法案(Gesetz gegen digitale Gewalt; GgdG)」を公表した。
- 同法案をめぐっては、2023年4月に連邦司法省が初めて具体的な法案構想を公表し、2024年末には草案も示されていたが、これまで連邦議会への正式な提出には至っていなかった。

構想当初の問題意識(2023年)

- 2023年4月に法案構想を公表した際のプレスリリース*1において、ブッシュマン連邦司法大臣(当時)は、法案の背景と目的を次のように説明した。
 - ✓ デジタル空間では、**侮辱、脅迫、誹謗中傷等の「デジタル暴力」**による被害が急増しており、被害者は精神的苦痛だけでなく、経済的にも深刻な影響を受ける恐れがある。
 - ✓ 現行法では、多くの場合において**権利侵害情報の発信者を特定することが難しく、加えて法的手続に負担や費用も大きい**ため、十分な被害救済につながっていない。
 - ✓ このため、法案は、デジタル空間における権利侵害に対抗するための法的な選択肢を充実させることを目的とする。

草案の公表(2026年)

- 上記構想に基づき2023年に意見募集が実施され、連邦司法省(BMJ, 当時)はその結果も踏まえて法案を検討し、年内には正式提案する計画を発表したが、結果として議会に法案は提出されなかった。(2024年には草案も公表されたが、法案提出には至らなかった。)
 - 野党(ドイツのための選択肢; AfD)の議員団は、デジタル暴力を新たに定義するメリットが十分に説明されていないことや、プロバイダの責任や裁判所によるアカウント停止命令の対象は重大な犯罪に限定すべきであること等を指摘し、法案作成・審議を注視するよう求める声明を公表した。*2
- 2026年4月、連邦司法・消費者保護省(BMJV)が「デジタル暴力対策法案」の草案を公表*3し、意見募集を実施。
- この最新の草案は、**「デジタル暴力」として新たな犯罪類型を創設する刑法改正、発信者情報開示等の民事上の救済手段の強化、オンライン事業者の責任強化等**を含む内容となっている。

*1 https://www.bmjbv.de/SharedDocs/Pressemitteilungen/DE/2023/0412_Digitale_Gewalt.html

*2 <https://www.bundestag.de/presse/hib/kurzmeldungen-1002970>

*3 https://www.bmjbv.de/SharedDocs/Gesetzgebungsverfahren/DE/2026_Gesetz_gegen_digitale_Gewalt.html

Ofcomによる調査

- 英国のOfcomは、Grokを用いた同意に基づかない性的画像の生成・共有が報告されていること等を受け、OSAに基づきXに対する正式な調査を開始した。

Ofcomによる調査の開始

- 2026年1月12日、OfcomはXに対しOSA義務違反の疑いで正式な調査を開始することを公表した。
- 同月15日には、Xが対策を講じ、Grokの機能を制限したと報告されたが、同社への調査は継続されることになった。

調査の論点

- 違法コンテンツのリスク評価義務(OSA第9条)
- 優先違法コンテンツ対策義務(OSA第10条)
- 違法コンテンツの迅速な削除(OSA第10条)
- プライバシー侵害への配慮(OSA第22条) 等

今後の展開

- Ofcomは、今後、証拠収集・分析を行い、違反が認められる場合には暫定判断通知をXに発令する。その際、Xは反論の機会が与えられる。Xへの意見照会の際、OSA違反の該否の最終判断が行われる。
- OSA義務違反が認められる場合、最大1,800万ポンドまたは全世界売上高の10%のいずれか高いほうの額の罰金が科され得る。また、Ofcomは具体的改善措置の実施命令や英国内からのアクセス遮断等の措置を命じる可能性がある。

<https://www.ofcom.org.uk/online-safety/illegal-and-harmful-content/ofcom-launches-investigation-into-x-over-grok-sexualised-imagery>

3. 韓国の動向

- 3.1 韓国・情報通信網法改正の概要
- 3.2 韓国・情報通信網法改正に対する反応・評価

3.1 韓国・情報通信網法改正の概要

情報通信網法の改正(虚偽・操作情報根絶法の制定)

- 2025年12月に可決された改正情報通信網法(虚偽・操作情報根絶法)は、悪意ある偽情報の流通防止を目的に「虚偽・操作情報」を新たに違法化し、大規模プラットフォーム事業者に対し通報窓口の設置や透明性レポートの公表を課している。虚偽・操作情報の流通に関し故意または過失が認められた場合、実損害の最大5倍の懲罰的損害賠償が課され得る。
- 同改正は与党・共に民主党がEUのDSAを参考にして草案提示したものが元になっている。国内外の報道機関や野党等からは、表現の自由や過剰規制等を理由に反発する声が挙げられている。

改正の背景・経緯

- 韓国政府及び与党(共に民主党)はSNSや動画共有サービス等を通じて拡散する「フェイクニュース」やAIを悪用したディープフェイク等を、社会的・政治的リスクとして強く問題視しており、2024年以降、偽情報対策を目的とする複数の改正案を国会に提出していた。
- 2025年9月には、与党系のメディア改革特別委員会がEUのデジタルサービス法(DSA)を参考にした情報通信網法改正案(大規模事業者への規制強化を含む)を提示した。^{*1}
- 2025年12月10日、国会の科学技術放送通信委員会において与党が主導する改正法案が承認され、同12月24日、違法情報の類型に「虚偽・操作情報」を追加し、プロバイダの責任を強化する情報通信網法の改正(通称「**虚偽・操作情報根絶法**」)が本会議で可決された。改正法は2026年7月7日に施行される。

虚偽・操作情報根絶法の規制内容

- **「虚偽・操作情報」の違法化**:「虚偽・操作情報」を違法化。定義は、以下の3つのすべてに該当する情報とされている。
 - ① 全部または一部が虚偽である情報または事実と誤認されるよう加工された情報であり、
 - ② その情報が虚偽または操作された情報であることを認識しながら、他人に損害を与える意図または不当な利益を得る意図で流通させられ、
 - ③ ①、②により、他人の人格権、財産権、公共の利益を侵害する場合
 - ※ 風刺やパロディは明示的に除外される。
- **「大規模情報通信サービス提供者」の新設**:規制対象として、従前の「情報通信サービス提供者」に加えて「大規模情報通信サービス提供者」の区分を設け、該当事業に対し追加的な義務を課す。
 - ⇒ 大規模情報通信サービス提供者の定義は、一定の利用者数や売上高等の規模の条件を満たすものとされ、具体的な要件は大統領令によって別途定められる。2026年5月時点で、大統領令は採択されていない。
- **「大規模情報通信サービス提供者」の義務等**:虚偽・操作情報に関して、「通報受付・対応措置義務」、「異議申立・紛争解決」、「自主規制運営方針の策定義務」、「透明性報告義務」の義務が規定された。また「懲罰的損害賠償」、「ファクトチェック機関との連携(任意)」が規定された。

*1 <https://ccianet.org/wp-content/uploads/2025/10/CCIA-Concern-over-proposed-Korean-Content-Control-legislation.pdf>

改正法に対する反応・評価及び憲法訴訟

- 可決直後から、韓国内外で表現の自由や過剰規制の観点から懸念が示されている。
- 2026年4月には、表現の自由の侵害や事前検閲に該当するとして市民団体やインターネットメディア関係者等が集団で憲法訴願を提起した。

改正法に対する反応・評価

報道関係団体(放送記者連合、全国言論労働組合、韓国記者協会など)	同改正が権力者による批判報道の萎縮手段に転化するおそれがあるとして反発した。 ^{*1}
野党・国民の力党	法律の概念が曖昧で恣意的検閲や過剰・二重処罰につながり得ると批判し、のちに関連条項の削除を目指す対案提出にも動いている。 ^{*2}
米国・国務省 ロジャース次官	同改正が米国発のオンライン・プラットフォームの事業に悪影響を及ぼし、表現の自由を損なう懸念があることから、韓国がデジタルサービスに「不必要な障壁」を課すべきではないと述べている。 ^{*3}
国際報道団体 International Press Institute (IPI)	核心概念の曖昧さや、メディアの自己検閲を促す危険性、当局に裁量を与える点を問題として指摘し、施行停止と人権影響評価の実施を求めている。 ^{*4}

憲法訴訟での請求人の主な主張

- **臨時措置の拡張(第44条の2)**: 情報通信サービス提供者は、利用者から権利を侵害する情報の削除要請を受けた際、権利侵害の有無を判断することが困難であるか、当事者間で争いが予想される場合には、当該情報へのアクセスを一時的に遮断する臨時措置を最長30日間講じることができる。(※従来から存在し、主に名誉毀損等の権利侵害情報に適用されてきたが、改正により虚偽・操作情報にも適用可能となった。)
 - ⇒ 「利用者の通報のみを根拠に最大30日間、投稿が遮断され得る仕組みは、**事実上の事前検閲にあたり、公職者や公共的事項に対する批判を無力化する手段として悪用される恐れがある**」と指摘。
- **懲罰的損害賠償制度(第44条の10)**: 虚偽・操作情報の流通により他者の権利が侵害された場合、裁判所は発信者に対し実害の最大5倍まで損害賠償を命じることができる。
- **課徴金(第44条の24)**: インターネットを通じて事実・意見を不特定多数に伝達することを業とする者が、既に裁判所によって違法情報または虚偽・操作情報と認定され、有罪判決、損害賠償命令等の判決が確定した情報を、インターネット上で2回以上流通させた場合、最大10億ウォンの課徴金が科され得る。
 - ⇒ **過激な制裁規定が言論機関や市民の自由な表現活動を萎縮させると主張。**

*1 <https://jp.yna.co.kr/view/AJP20251224002600882>

*2 <https://www.news1.kr/politics/assembly/6040708>

*3 https://www.chosunonline.com/site/data/html_dir/2026/01/01/2026010180014.html

*4 <https://www.hani.co.kr/arti/society/media/1241097.html>

*5 <https://www.korea.kr/news/policyNewsView.do?newsId=148960037>

*6 <https://www.fntoday.co.kr/news/articleView.html?idxno=382559>

4. 米国の動向

- 4.1 米国・TAKE IT DOWN Actの概要
- 4.2 米国・アルゴリズム説明責任法案の概要

4.1 米国・TAKE IT DOWN Actの概要

TAKE IT DOWN Actの概要

- 2025年5月、ソーシャルメディア等の事業者に対し、AI生成によるディープフェイクを含む合意のない性的画像の削除を義務付ける「TAKE IT DOWN Act(Tools to Address Known Exploitation by Immobilizing Technological Deepfakes on Websites and Networks Act)」が成立した。

TAKE IT DOWN Act
の提案と成立

- **TAKE IT DOWN Act**(ウェブサイトおよびネットワーク上の技術的なディープフェイクを無効化することで、既知の搾取に対処するための手段法)は、オンライン上の合意のない性的搾取画像(ディープフェイク画像を含む)の拡散に対処することを目的としている。
- 2025年1月に、クルーズ上院議員(共和党)、クロブチャー上院議員(民主党)ら超党派によって上院及び下院に提出された。
- 同年2月に上院、4月に下院で可決され、5月にトランプ大統領が署名して成立した。

法律の概要

合意のない性的画像の違法化	識別可能な個人に関する性的な視覚的描写について、①同意なき親密な画像(性的画像)および②AI等により生成・改変した性的なデジタル偽造のオンラインでの公開を、 <u>連邦法上の犯罪行為として規定</u> する。 違反した場合、被害者が成人の場合は最長2年、未成年の場合は最長3年の拘禁刑が科される。
合意のない性的画像の通報・削除要求プロセスの確立	ユーザ生成コンテンツを扱うプラットフォーム事業者は、法律の成立日から1年以内に、ユーザが対象プラットフォーム上の性的画像について当該ユーザの同意なく公開されていることを通報し、削除を要求できるプロセスを確立しなければならない。
合意のない性的画像の削除	ユーザ生成コンテンツを扱うプラットフォーム事業者は、被害者からの通報を受けてから48時間以内に当該コンテンツを削除しなければならない。不履行の場合、連邦取引委員会法違反として罰則が科される可能性がある。 また、当該画像のコピーを合理的に特定し削除するために合理的な努力措置を講じなければならない。
執行権限	FTCによる強制執行の権限を定める。

アルゴリズム説明責任法案の背景・経緯及び審議状況

- 2025年11月に超党派で提出された「アルゴリズム説明責任法(Algorithm Accountability Act)」案は、現在のSNSが抱えるアルゴリズム主導の課題に対し、プラットフォーム企業へ生命・身体等に対する危険への合理的注意義務を新たに課すことを定めている。
- 同法案は、エンゲージメント最大化を優先するアルゴリズムが暴力や自傷などの有害情報を増幅させている現状を危惧し、通信品位法230条による広範な免責を制限することで、企業の設計責任を明確化することを目指す。
- 類似の名称で「アルゴリズム的説明責任法(Algorithmic Accountability Act; AAA)」案がワイデン上院議員(民主党)らにより2019年から複数回にわたり提出されてきたが、いずれも採決に至らず、成立していなかった。

法案提出の背景・経緯

- 2025年11月、ケリー上院議員(民主党)およびカーティス上院議員(共和党)、ケネディ下院議員(共和党)およびディレイニー下院議員(民主党)がそれぞれ、両院議会に「**アルゴリズム説明責任法(Algorithm Accountability Act)**」案を提出した。^{*1*2}
- **提案の背景:**提案者らは、通信品位法230条は、約30年前のインターネット環境を前提として設計されたものであり、現在のアルゴリズム主導型のSNSの実態に適合していないと指摘している。^{*3}
 - 特に、エンゲージメント最大化と広告収益増加を目的とするレコメンダー・アルゴリズムが、暴力的コンテンツ、自傷・自殺関連情報、過激思想、違法薬物関連情報等を増幅し、未成年者を含む利用者に深刻な影響を及ぼしているとの問題意識が共有されている。
 - また、ケネディ議員は、医師としての経験にも言及し、アルゴリズム設計が精神的・身体的健康に現実の影響を与えていると指摘している。

法案審議の状況

- アルゴリズム説明責任法案は上下両院に提出され、各委員会で審議が進行中とみられるが、2026年5月時点で採決には至っていない。
- 州政府関係者、州司法長官、児童保護団体、被害者家族団体などから法案を支持する声明が出されており^{*4}、関係者の関心は高い。
- なお、類似の名称で「アルゴリズム的説明責任法(Algorithmic Accountability Act; AAA)」案がワイデン上院議員(民主党)らにより2019年から複数回にわたり提出されてきたが、いずれも採決に至らず、成立していなかった。^{*5*6}

*1 <https://www.congress.gov/bill/119th-congress/senate-bill/3193>

*2 <https://www.congress.gov/bill/119th-congress/house-bill/6266/text>

*3 <https://mikekenedy.house.gov/media/press-releases/utah-rep-mike-kennedy-introduces-algorithm-accountability-act-protect-families>

*4 <https://www.kelly.senate.gov/newsroom/press-releases/kelly-curtis-introduce-algorithm-accountability-act/>

*5 <https://www.congress.gov/bill/116th-congress/senate-bill/1108>

*6 <https://www.congress.gov/bill/119th-congress/house-bill/5511>

未来を問い続け、変革を先駆ける

MRI 三菱総合研究所