資料４－３

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4回）

EUデジタルサービス法と偽・誤情報対策

2023年12月15日

生貝直人　博士（社会情報学）
一橋大学大学院法学研究科教授

デジタルサービス法（2022年発効）

- EUのプロバイダ責任を規定してきた電子商取引（2000年）を元に、違法・有害情報に対するプラットフォームの責任・責務や透明性のあり方を全面的にアップデート
- 第III章では、媒介サービス事業者一般やプラットフォーム事業者一般に適用されるコンテンツモデレーション透明性・救済規律の他、EU域内で月間アクティブ利用者4,500万人以上を有する「**超大規模オンラインプラットフォーム（very large online platform、VLOP）**」＋「**超大規模オンライン検索エンジン（very large online search engine、VLOSE）**」事業者に、偽・誤情報対策を含む追加的な義務を課す
  - 2023年4月25日に17のVLOPと2のVLOSEが指定
  - VLOP・VLOSE条項は2023年8月から適用開始、2024年2月全面適用開始
- デジタルサービス法の要点
  - ①コンテンツモデレーション：透明性と救済
  - ②データ保護：プロファイリング規制
  - ③VLOP/VLOSE：偽・誤情報を含むシステミックリスクの評価と軽減

デジタルサービス法の対象事業者区分と主な規律

仲介サービス（IS）：導管、キャッシング、ホスティングの3種類
免責ルール（2章）、連絡先・代理人・利用規約規制等（3章１節）

ホスティングサービス（HS）：利用者提供情報ホスティング全般
違法コンテンツ通知と措置、理由の説明、透明性レポート等（3章2節）

オンラインプラットフォーム（OP）：利用者提供情報の公衆配布
零細・中小義務除外、内部苦情処理、信頼できる騎手、反復侵害者対応、広告規制・透明性、
レコメンデーション透明性、未成年者保護（3章３節）

超大規模OP（VLOP）：EU域内利用者4,500万人以上のOP
システミックリスクの評価と軽減、危機対応メカニズム、独立監査、当局・
研究者へのデータ提供、コンプライアンス体制整備等（3章5節）
欧州委員会による監督・調査・執行・モニタリング（4章4節)

取引OP：消費者と事業者の間の契約締結を可能とするOP
事業者トレーサビリティ等（3章４節）

超大規模オンライン検索エンジン（VLOSE）：EU域内利用者4,500万人以上の検索エンジン
VLOPに課される3章5節の義務とほぼ同様

①コンテンツモデレーション：透明性と救済

3条(t)：「コンテンツモデレーション」とは、自動的か否かに関わらず、媒介サービス提供者が行う、特にサービス受領者が提供する違法コンテンツ又はその利用規約に適合しない情報の検出、識別、対処を目的とした活動をいい、降格、収益不能化、アクセス不能化、削除など、違法コンテンツ又はその利用規約違反情報の利用可能性、可視性、アクセス性に影響を与える措置、サービス受領者のアカウントの終了又は停止など、サービス受領者の情報提供能力に影響を与える措置を含む。

- 利用規約へのコンテンツモデレーションポリシー明記（IS、14条）
  - 利用者提供情報に関する制限の情報（アルゴリズムによる意思決定と人間によるレビューを含むコンテンツモデレーションのあらゆる方針・手順・手段・ツール、内部苦情処理システム手続に関する情報を含む）
  - 制限の実施における表現の自由やメディアの自由・多元性、その他基本権等の利益への配慮義務
  - VLOP/VLOSEは全サービス提供加盟国の言語で当該情報を提供
- 透明性レポート（IS~VLOP段階、15条他）→VLOPの第一次レポート
  - 当局命令・対応、違法・規約違反別の通知・対応件数と対応時間、コンテンツモデレーション担当者訓練内容、自動処理のエラー率指標とセーフガード措置等（※VLOPは加盟国の公用語ごとに整理）
- 理由の説明（IS、17条）　→欧州委による集約データベース
  - コンテンツ削除・降格やアカウント停止等を受けた利用者への明確かつ具体的な理由説明
- 内部苦情処理システムの整備（OP、20条）
  - 削除やアカウント停止等の判断が誤っていた場合の回復等
- 裁判外紛争処理の利用（OP、21条）
  - 紛争処理機関に対する当局の認定等

②データ保護：プロファイリング規制

- PF上の**ターゲティング広告**のパラメータ等の明示（OP~VLOP段階、26条他）
- **レコメンダーシステム**のパラメータ明示とユーザーによる修正可能性（VLOPはプロファイリングに基づかない選択肢の提供を含む）（OP～VLOP、27条他）
- GDPR特別カテゴリー個人データのプロファイリング広告利用禁止（OP、26条3項）
- 青少年保護と未成年個人データのプロファイリング広告利用禁止（OP、28条2項）

- ※ダークパターンの禁止（OP、25条）：「サービス受領者を欺いたり操作したりするような方法で、又はその他の方法でサービス受領者が自由かつ情報に基づく決定を行う能力を実質的に歪めたり損なったりする方法で、オンライン・インターフェースを設計、組織、運用しないこと」

③VLOP/VLOSE：偽・誤情報を含むシステミックリスクの評価と軽減

- VLOP/VLOSEは、自らのサービスがもたらしうる違法コンテンツ流布、基本権（特に人間の尊厳、プライバシー、個人データ保護、表現・情報の自由、非差別、児童の権利、消費者保護）、**市民言説と選挙**、ジェンダー暴力・公衆衛生・青少年保護等への影響**等の「システミックリスク」を自ら特定・分析・評価し（34条）、合理的・比例的・効果的な軽減措置を採る義務（35条）**と、公共の安全・公衆衛生への重大な脅威における危機対応メカニズムにおいて出される欧州委員会の要請決定の対象となる（36条）
- 欧州委員会が奨励・推進・招請して策定する、行動規範（codes of conduct）（45条）や危機プロトコル（48条）による共同規制メカニズム
  - 行動規範や危機を通じた協力はリスク軽減措置の一つ（35条1項(h)）
  - デジタルサービス法採択以前から偽情報行動規範が策定、複数PFにより署名済み
- 34条・35条の義務及び、行動規範・危機対応プロトコルの遵守について、年1回以上の独立監査を受ける義務（37条）
  - 評価・緩和措置検証のための外部研究者データアクセス提供義務（40条）

デジタルサービス法
行動規範と自主・共同規制

(103) 欧州委員会およびEU理事会は、本規則の適用に資するため、自発的な行動規範の作成と、それらの規範の規定の実施を奨励すべきである。欧州委員会およびEU理事会は、行動規範が、取り組む公益目的の本質を明確に定義し、その目的の達成を独立的に評価する仕組みを含み、関係当局の役割が明確に定義されていることを目指すべきである。(…)

(104) 本規則は、そのような行動規範のために考慮すべき特定の分野を特定することが適切である。**特に、特定の種類の違法コンテンツに関するリスク軽減措置は、自主規制および共同規制の合意を通じて検討されるべきである。**また、**偽情報や操作的な悪用行為、未成年者への悪影響など、システミックリスクが社会と民主主義に及ぼしうる負の影響も考慮すべき領域である。**これには、意図的に不正確な、あるいは誤解を招くような情報を、時には経済的利益を得る目的で作成するためにボットや偽アカウントを使用するなど、偽情報を含む情報の増幅を目的とした協調的な操作が含まれ、これらは特に未成年者などサービスの受け手である弱者にとって有害である。このような分野に関連して、**VLOPやVLOSEによる所定の行動規範の遵守と順守は、適切なリスク軽減措置として考えられる。オンラインプラットフォームまたはオンライン検索エンジンのプロバイダーが、そのような行動規範の適用への参加に対する欧州委員会の招請を適切な説明なしに拒否した場合、当該オンラインプラットフォームまたはオンライン検索エンジンが本規則の定める義務を侵害したか否かを判断する際に、関連性があれば考慮されうる。**ある行動規範に参加し、それを実施しているという事実だけで、それ自体が本規則を遵守していると推定すべきではない。　（強調報告者）

デジタルサービス法
行動規範と自主・共同規制

違法ではないが有害なコンテンツに効果的に対処するには？

違法でない限り、有害なコンテンツを違法コンテンツと同様に扱うべきではない。新規則は、表現の自由を完全に尊重した上で、違法コンテンツの削除または削除を促す措置のみを課す。

同時に、DSAは、偽情報、デマ、パンデミック時の操作、社会的弱者への危害、その他の新たな社会的危害のようなシステミックな問題に関して、VLOPとVLOSEの責務（※responsibilities）を規制する。欧州委員会により、4,500万人のユーザーを抱えるVLOPおよびVLOSEに指定された後、これらの企業は、毎年リスク評価を行い、サービスの設計および利用に起因する対応するリスク軽減措置を講じる必要がある。そのような措置は、表現の自由の制限と慎重にバランスをとる必要がある。また、独立した監査を受ける必要もある。

さらに、本提案では、サービスプロバイダーが行動規範のもとで、違法コンテンツの拡散や、児童や未成年者などサービスの弱者にとって特に有害な操作的・虐待的行為に関する悪影響に対処できるような、共同規制的枠組（co-regulatory framework）を定めている。

DSAは、ディスインフォメーション（偽情報）に関する行動規範の改訂版のような行動規範や、危機管理プロトコルを含む、オンライン上の危害に関する共同規制的枠組を促進する。

欧州委員会 Questions and Answers: Digital Services Act（2023年4月25日付）より仮訳

危機対応メカニズム／プロトコル（crisis response mechanism / protocol）

前文(91) **危機に際しては、VLOPプロバイダーが、**本規則に基づく他の義務を考慮して講じる措置に加えて、**特定の措置を緊急に講じる必要が生じる可能性がある。**この点で、危機は、欧州連合またはその重要な部分において、公共の安全または公衆衛生に対する重大な脅威につながり得る異常な状況が発生した場合に発生すると考えられるべきである。このような危機は、**武力紛争やテロ行為（新たな紛争やテロ行為を含む）、地震やハリケーンなどの自然災害、パンデミックや公衆衛生に対する国境を越えたその他の深刻な脅威**から生じる可能性がある。欧州委員会は、欧州デジタルサービス会議（「会議」）の勧告に基づき、VLOPプロバイダーおよびVLOSEプロバイダーに対し、**緊急の問題として危機対応を開始するよう求めることができる**べきである。これらのプロバイダーが特定し、適用を検討しうる措置には、**例えば、コンテンツモデレーションプロセスの適合およびコンテンツモデレーションに専念するリソースの増加、利用規約、関連するアルゴリズムシステムおよび広告システムの適合、信頼できるフラグ作成者との協力のさらなる強化、啓発措置の実施、信頼できる情報の促進、オンラインインターフェースのデザインの適合などが含まれる。**このような措置が極めて短期間で講じられることを確保するために必要な要件が規定されるべきであり、また、危機対応メカニズムは、厳密に必要な場合に、必要な範囲でのみ使用され、このメカニズムの下で講じられる措置は、すべての関係者の権利と正当な利益を十分に考慮した上で、効果的かつ比例的なものでなければならない。（…）

前文(108) VLOPおよびVLOSEのための危機対応メカニズムに加えて、**欧州委員会は、**オンライン環境における迅速かつ集団的で国境を越えた対応を調整するために、**ボランタリーな危機対応プロトコルの作成を開始することができる。**例えば、オンラインプラットフォームが違法コンテンツや偽情報の迅速な拡散に悪用された場合や、信頼できる情報を迅速に拡散する必要が生じた場合などがこれにあたる。（…）（強調報告者）

AI規制論の焦点（私見）：

- ビフォー生成AI（情報処理）
  - EU以外はソフトロー重視
  - 主な焦点リスクは**製品安全＋プロファイリング**
- アフター生成AI（情報生成）
  - EU以外もハードローを意識
  - 主な焦点リスクは**偽・誤情報＋情報環境全般への影響**
  - さらに、国家安全保障、競争政策の前面化

AI法とデジタルサービス法：偽情報行動規範（2022年6月改訂）

AIシステムの透明性義務

コミットメント15：AIシステムを開発・運用し、AIが生成・操作（※manipulate）したコンテンツ（例：ディープフェイク）をサービスを通じて流布する関連署名者は、透明性義務と、AI法に関する提案で禁止されている操作行為のリストを考慮することを約束する：

- **対策15.1. 関連署名者は、コンテンツを生成・操作するAIシステムに対し、利用者への警告やそのようなコンテンツの積極的な検知など、禁止操作行為に対抗するための方針を策定または確認する。**
- QRE 15.1.1： EU法および国内法に従い、関連する署名機関は、コンテンツを生成・操作するAIシステムに対し、禁止操作行為に対抗するための実施方針を報告する。

措置15.2. 関連署名者は、EUおよび加盟国の法令に従い、サービス上の許されない行為やコンテンツの検知、モデレーション、制裁に使用されるアルゴリズムが信頼でき、エンドユーザの権利を尊重し、その行動を不当に歪める禁止された操作行為とならないことを保証するためのポリシーを確立または確認する。

- QRE 15.2.1：関連署名者は、そのサービスで許容されない行為やコンテンツの検知、調停、制裁に使用されるアルゴリズムが信頼に足るものであり、エンドユーザの権利を尊重し、EUおよび加盟国の法規に沿った禁止された操作行為に該当しないことを保証する方針と行動について報告する。

AI法とデジタルサービス法：
汎用AIとシステミックリスクの評価と軽減

- 2023年12月9日合意では、特に影響力の大きい汎用AI（general purpose AI）にき、透明性要件等の他、システミックリスクの評価と軽減の義務を課す内容が含まれる見通し。

●汎用AIモデルはどのように規制されているのか？

大規模な生成AIモデルを含む汎用AIモデルは、様々なタスクに使用することができる。個々のモデルは、多数のAIシステムに統合される可能性がある。

汎用AIモデルに基づいて構築しようとする提供者は、そのシステムが安全でAI法に準拠していることを確認するために必要なすべての情報を持っていることが重要である。

そのため、AI法は、このようなモデルの提供者に対して、下流のシステム提供者に対して一定の情報を開示することを義務付けている。このような透明性は、これらのモデルに対する理解を深めることを可能にする。

モデル提供者はさらに、モデルを訓練する際に著作権法を尊重することを保証するポリシーを持つ必要がある。

加えて、これらのモデルの中には、非常に高性能であったり、広く使用されているため、システミックリスクを引き起こす可能性があるものもある。

現時点では、**10^25FLOPs以上の計算能力を用いて訓練された汎用AIモデルは、システミックリスクがあると考えられている。**AIオフィス（欧州委員会内に設置）は、技術の進歩に照らしてこの閾値を更新することができ、さらに特定の場合には、さらなる基準（ユーザー数やモデルの自律性の程度など）に基づいて、他のモデルをそのように指定することができる。

したがって、システミックリスクを持つモデルのプロバイダーは、リスクの評価と軽減、重大インシデントの報告、最先端のテストとモデル評価の実施、サイバーセキュリティの確保、モデルのエネルギー消費に関する情報の提供を義務付けられている。

このため、他の専門家と協力し、**ルールを詳細に規定する中心的なツールとして、行動規範を作成するために、欧州AIオフィスと協力することが求められている。**科学委員会は、汎用AIモデルの監督において中心的な役割を果たす。

欧州委員会 Artificial Intelligence – Q&As （2023年12月12日時点）（強調報告者）

いくつかの論点

- コンテンツモデレーション：透明性と救済
  - 判断への意義申立と情報空間に関する開かれた議論の前提
- データ保護：プロファイリング規制
  - 大規模なレコメンダーやターゲティング広告への対応
- VLOP/VLOSE：偽・誤情報を含むシステミックリスクの評価と軽減
  - 行動規範を通じた共同規制メカニズムと危機対応のあり方
- 汎用・生成AI：システミックリスクの評価と軽減
  - ディープフェイクを含む偽・誤情報の「生成」と「流通」の両面
- 非公式な要請か、法に基づく要請か