総務省
デジタル空間における情報流通に係る制度ワーキンググループ（第３回）

未定稿

資料３－３

違法・有害情報に関する諸外国の対応状況

株式会社野村総合研究所

コンサルティング事業本部
ICT・コンテンツ産業コンサルティング部

2025年3月26日

NRI

Envision the value,
Empower the change

目次

1．リスク評価・軽減措置

1．各国制度の概要（DSA・OSA）

項目	DSA（Digital Services Act）	OSA（Online Safety Act）
リスク評価	• 大規模PF事業者（大規模オンラインプラットフォーム事業者（VLOP）及び大規模オンライン検索エンジン事業（VLOSE））に対して、リスク評価の実施を義務付け（34条）。 なお、リスク評価の具体的な実施方法については、法律等では定められていなく、事業者に委ねている。 ＜リスク評価を行うリスクの例＞ ・サービスを通じた違法コンテンツの拡散リスク ・基本的権利への影響 ・公共の安全への悪影響	• ユーザー間サービスや検索サービスを提供する事業者に対して、リスク評価の実施を義務付け（９条）。 なお、リスク評価の具体的な実施方法については、法律に基づきOfcomが策定するガイダンスにおいて明記（98条）。 ＜リスク評価を行うリスクの例＞ ・優先違法コンテンツへの遭遇リスク ・個人への危害のリスク ・利用の多様性と利用者に対するリスク
軽減措置	• 大規模PF事業者に対して、リスク評価を踏まえた軽減措置を行うことを義務付け（35条）。 なお、軽減措置の内容については、法律で例示はあるが具体的な内容について定められていなく、事業者に委ねている。また、行動規範に参加することで軽減措置の手段とみなされることもある（※行動規範に参加することで、DSAの義務を満たしていることになるわけではない。）(前文104項) ＜軽減措置の例＞ ・コンテンツモデレーションプロセスの適応 ・アルゴリズムシステムの適合 ・サービス受領者向けの啓発	• ユーザー間サービスや検索サービスを提供する事業者に対して、リスク評価を踏まえた軽減措置を行うことを義務付け（10条）。 なお、軽減措置の具体的な内容については、法律等では定められていないが、法律に基づきOfcomが策定する行動規範において明記（41条）。 ＜軽減措置の例＞ ・優先的犯罪の実行または支援に使用されるリスクを効果的に軽減・管理（2項(b)） ・個人への害のリスクを効果的に軽減・管理
義務履行の確認方法	• 大規模PF事業者に対して、第三者の監査主体による監査を受けることを義務付けとともに、監査レポート、監査取組レポート、システミックリスク評価レポートを欧州委員会・DSCに提出することを義務付け（37条）。 • 欧州委員会は、DSA違反の疑念がある場合に、大規模PF事業者に対して情報提供命令やインタビュー等を行い、違反を認定（66条等）。	• ユーザー間サービスや検索サービスを提供する事業者に対して、サービスに関する透明性報告書を作成し、Ofcomに提出することを義務付け（77条）。 • Ofcomは、OSA違反の疑義がある場合には、情報要求（100条）、インタビューの要求（106条）等を行い、違反仮通知やサービス提供者の意見陳述を経たうえで、違反を決定（132条）。

1－1．EUにおけるDSA （Digital Services Act）の リスク評価・軽減措置

DSAの各事業者カテゴリにかかる規定

規律	該当条文		仲介サービス	ホスティングサービス	オンラインプラットフォームサービス	VLOP・VLOSE
違法コンテンツに関する措置命令・情報提供の命令	第二章	第9条・第10条	●	●	●	●
連絡先（対DSC、対欧州委員会、対閣僚理事会）、サービス提供者の窓口、法定代理人		第11条・第12条・第13条	●	●	●	●
利用規約の要件		第14条	●	●	●	●
仲介サービス提供者に対する透明性報告義務		第15条	●	●	●	●
利用者への通知・行動の仕組み、情報提供・理由の記載義務		第16条・第17条		●	●	●
刑事犯罪の疑いに関する通知		第18条		●	●	●
内部苦情処理体制・救済の仕組みと法廷外紛争解決		第20条・第21条			●	●
信頼された旗手		第22条			●	●
悪用に対する措置と保護		第23条			●	●
オンライン・プラットフォームのプロバイダーに対する透明性報告義務		第24条			●	●
オンラインインターフェースのデザインと構成		第25条			●	●
オンラインプラットフォームでの広告		第26条			●	●
レコメンダシステムの透明性		第27条			●	●
未成年者のオンラインでの保護	第三章	第28条			●	●
超大規模オンライン検索エンジン		第33条				●
リスク評価、リスク軽減		第34条・第35条				●
危機対応メカニズム		第36条				●
独立監査（外部リスク監査と公的説明責任）		第37条				●
レコメンダシステム		第38条				●
オンライン広告の透明性向上		第39条				●
データへのアクセスと精査（当局・研究者）		第40条				●
コンプライアンス機能		第41条				●
透明性報告義務		第42条				●
監督手数料		第43条				●
標準		第44条		●	●	●
行動規範、オンライン広告・アクセシビリティの行動規範		第45条・第46条・第47条		●	●	●
危機対応への協力		第48条		●	●	●

出所）DSA https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32022R2065#art_22

DSAにおける有害コンテンツの位置づけ

違法コンテンツについてリスク評価（34条）・軽減措置（35条）を義務付け。また、違法でないが有害なコンテンツ（偽情報等）についても、リスク評価・軽減措置の対象(前文84項、104項)。

項目	内容	条項
有害コンテンツ（harmful content）（偽情報） **※DSA本文で規定なし**	本則では直接的な言及はされていないが、DSA前文や欧州委員会が公表しているDSAに関するQ&Aにおいて、違法ではないが有害なコンテンツ（サービスの設計や意図的な操作（ボットや偽アカウントの使用）によって大量に拡散された偽情報、脆弱な層（未成年など）に特に害を及ぼす誤情報等）についても、対応が必要であると示されている。 ＜リスク評価に関連する前文＞ 「本規則で特定された**システム上のリスクを評価する際**には、・・・、**違法ではないが、本規則で特定されたシステム上のリスクに寄与する情報にも焦点を当てるべき**である。したがって、そのような提供者は、**誤解を招く**ような、または**欺瞞的なコンテンツ（偽情報など）**を拡散または増幅するために自社のサービスがどのように利用されているかに**特に注意を払うべき**である。」（前文84項） ＜軽減措置に関連する前文＞ 「・・・検討すべきもう一つの領域は、**偽情報や操作および濫用行為、未成年者への悪影響など、社会および民主主義に対するシステミックリスクの潜在的な負の影響**である。・・・**このような分野**に関連して、超大規模オンラインプラットフォームまたは超大規模オンライン検索エンジンが**特定の行動規範を遵守**し、それに従うことは、**適切なリスク軽減措置であると考えられる**。・・・特定の行動規範への参加および実施という事実だけでは、それ自体では本規則の遵守を推定すべきではない」（104項）	● 前文84項 ● 前文104項

出所）DSA https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32022R2065

リスク評価・軽減措置に関する制度 | 第34条「リスク評価」

■ **VLOP/VLOSEに対して**「リスク評価」を**義務付けており**、サービスおよびアルゴリズムシステムを含む関連システムの設計・機能、サービスの利用に起因する欧州域内のシステミックリスクを真摯に特定・分析・評価をしなければならない（34条）。

● **リスク評価は、少なくとも1年に1回、又は特定されたリスクに重大な影響を及ぼす可能性のある機能を展開する前に、実施しなければならない。**（34条1項）

システミックリスク（34条1項）

- 違法なコンテンツの拡散に関するリスク
- 基本的権利（人間の尊厳、私生活および家族生活の尊重、個人情報の保護、メディアの自由と多元性を含む表現と情報の自由、非差別、児童の権利の尊重、高水準の消費者保護）の行使に関するリスク
- 市民的言論や選挙プロセス、治安に及ぼす実際の、あるいは予測可能な悪影響に関するリスク
- ジェンダーに基づく暴力、公衆衛生および未成年者の保護、人の身体的・精神的福利に対する深刻な悪影響に関連する、実際または予見可能な悪影響に関するリスク

システミックリスクの評価（34条2項）

- リスク評価において、特に以下の要因について考慮する
  (a)推奨システムおよびその他の関連するアルゴリズムシステムの設計
  (b)提供者のコンテンツ修正システム
  (c)適用約款及びその執行
  (d)広告の選択・提示システム
  (e)提供者のデータ関連の実務
- サービスの意図的な操作（不正使用、自動的利用）、違法コンテンツ・利用規約に違反数r情報の増幅及び迅速かつ広範な普及の可能性の影響の有無・形態についても分析・評価
- 特定の地域的または言語的側面も考慮して評価（特定の加盟国固有の場合）

● さらに、リスク評価は、**違法ではないが、システム上のリスクに寄与する情報（有害コンテンツ）にも重点を置く**必要があり、「**誤解を招くコンテンツや意図的に誤った情報（誤情報・偽情報）がどのように拡散・増幅されるのかについても分析する必要がある**」（前文84項）とされている。

出所）DSA https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32022R2065

（参考）34条　リスク評価（1/2）

■ 第34条 リスク評価

1　**超大規模オンラインプラットフォームおよび超大規模オンライン検索エンジンの提供者は、そのサービスおよびアルゴリズムシステムを含む関連システムの設計もしくは機能、またはそのサービスの利用に起因する、域内のシステミックリスクを真摯に特定、分析および評価しなければならない。**リスク評価は、第33条第6項第2号で言及されている適用日までに、また、その後**少なくとも1年に1回**、さらに、いかなる場合においても、本条に従って特定されたリスクに重大な影響を及ぼす可能性のある機能を展開する前に、実施しなければならない。この**リスク評価は、**そのサービスに特有であり、システミックリスクに比例し、その重大性と蓋然性を考慮したものでなければならず、**以下のシステミックリスクを含むものとする：**

(a)その**サービスを通じて違法なコンテンツが広まること**

(b)**基本的権利**、特にEU基本権憲章第1条に規定される人間の尊厳、第7条に規定される私生活および家族生活の尊重、第8条に規定される個人情報の保護に関する基本的権利の行使に対する、現実または予見可能な悪影響、第11条に謳われるメディアの自由と多元性を含む表現と情報の自由、第21条に謳われる非差別、第24条に謳われる児童の権利の尊重、第38条に謳われる高水準の消費者保護

(c)**市民的言論**や**選挙プロセス、治安に及ぼす**実際の、あるいは予測可能な**悪影響**

(d)**ジェンダーに基づく暴力、公衆衛生**および**未成年者の保護、人の身体的・精神的福利に対する深刻な悪影響**に関連する、実際または予見可能な悪影響

2　**超大規模オンラインプラットフォームおよび超大規模オンライン検索エンジンのプロバイダーは、リスクアセスメントを実施する際、**特に、**以下の要因が**第1項にいう**システミックリスクのいずれかに影響を及ぼすか**どうか、および**どのように影響を及ぼすか**を**考慮しなければならない：**

(a)**レコメンダシステム**およびその他の**関連するアルゴリズムシステムの設計**

(b)**コンテンツモデレーションシステム**

(c)**適用される条件**および**その実施**

(d)**広告の選択**および**表示システム**

(e)**提供者のデータに関する慣行**

また評価は、第1項による**リスクがサービスの意図的な操作**（真正でない利用や自動化された利用を含む）、**違法なコンテンツや利用規約と相容れない情報の増幅や潜在的な迅速かつ広範な拡散によって影響を受けているか**どうか、また**どのように影響を受けているか**を**分析する**ものとする。評価は、加盟国に特有の場合を含め、特定の地域的または言語的側面を考慮するものとする。

3　超大規模オンラインプラットフォームおよび超大規模オンライン検索エンジンのプロバイダーは、**リスク評価の実施後少なくとも3年間、リスク評価の裏付けとなる文書を保存**し、**要請があれば、欧州委員会および設置国のDSCに伝達**しなければならない。

出所）DSA　https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32022R2065

（参考）34条　リスク評価（2/2）

■ 前文

(84)　・・・本規則で特定されたシステム上のリスクを評価する際には、これらの提供者は、**違法ではないが、本規則で特定されたシステム上のリスクに寄与する情報にも焦点を当てるべきである。**したがって、そのような提供者は、**誤解を招くような、または欺瞞的なコンテンツ（偽情報など）を拡散または増幅するために自社のサービスがどのように利用されているかに特に注意を払うべき**である。アルゴリズムによる情報の増幅がシステム上のリスクに寄与している場合には、これらの提供者は、リスク評価にこれを適切に反映させるべきである。リスクが局所的である場合や言語の違いがある場合、それらの提供者はリスク評価においてその点を考慮すべきである。特に、超大規模オンライン・プラットフォームや超大規模オンライン検索エンジンの提供者は、自社のサービスの設計や機能、意図的で、しばしば協調的な操作や利用、あるいは利用規約の組織的な侵害が、そのようなリスクにどのように寄与するかを評価すべきである。このようなリスクは、例えば、偽のアカウントの作成、ボットの使用、またはサービスの不正使用、およびその他の自動化または部分的に自動化された行動など、サービスの不正利用を通じて生じる可能性がある。これらは、違法なコンテンツや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利用規約に適合しない情報を一般大衆に急速かつ広範囲に拡散させ、偽情報キャンペーンに寄与する可能性がある。

出所）DSA　https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32022R2065

リスク評価・軽減措置に関する制度 | 第35条「リスク軽減」

■ **VLOP/VLOSEに対して**「軽減措置」を**義務付けており、**34条に従って特定されたシステミックリスクに合わせた、合理的、比例的かつ効果的な緩和措置を、基本的権利に与える影響に配慮して講じなければならない（35条）。

- 当該措置については、以下が例示されている。

措置内容	条文
サービスデザイン等の適合	35条1項(a)
利用規約の変更	35条1項(b)
特定の違法コンテンツに関連する通知の処理速度の向上等コンテンツモデレーションプロセスの適応	35条1項(c)
レコメンダシステム等アルゴリズムシステムの適合	35条1項(d)
広告システムの適合	35条1項(e)
システミックリスクの検知に関する内部プロセスや監督の強化	35条1項(f)
Trusted Flaggerとの協力	35条1項(g)
行動規範（45条）・危機プロトコル（48条）を通じた、オンラインプラットフォーム等との協力・調節	35条1項(h)
サービス受領者向けの啓発	35条1項(i)
児童の権利を保護するための支援ツールなどの措置	35条1項(j)
人物や事象に著しく類似し、誤認させるような画像、音声、動画がオンラインインターフェースに表示される際に目立つマークによって区別できる機能等の提供	35条1項(k)

- さらに、前文において、「偽情報や操作および濫用行為、未成年者への悪影響など、社会および民主主義に対するシステミックリスクの潜在的な負の影響である。・・・このような分野に関連して、超大規模オンラインプラットフォームまたは超大規模オンライン検索エンジンが特定の行動規範を遵守し、それに従うことは、適切なリスク軽減措置であると考えられる。」（前文104項）と規定。

出所）DSA https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32022R2065

（参考）35条　リスク軽減（1/2）

■ 第35条 リスク軽減（抜粋）

1　**超大規模オンラインプラットフォームおよび超大規模オンライン検索エンジンの提供者は**、第34条に従って**特定された特定のシステミックリスクに合わせた**、**合理的**、**比例的**かつ**効果的**な**緩和措置を**、当該措置が**基本的権利に与える影響に特に配慮して講じなければならない**。当該措置には、該当する場合、**以下が含まれる**：

(a)**オンラインインタフェース**を含む、**サービスのデザイン、特徴または機能を適合**させること

(b)**利用規約及びその実施方法を変更**すること

(c)**特に違法なヘイトスピーチやサイバー暴力に関して**、特定の種類の違法コンテンツに関連する**通知の処理速度**や**質**、適切な場合には、**通知されたコンテンツの迅速な削除やアクセス無効化**、また、関連する**意思決定プロセスやコンテンツ調整のための専用リソースの適応**を含む、**コンテンツモデレーションプロセスの適応**

(d)**レコメンダシステム**を含む**アルゴリズムシステムをテスト**し、**適応**させること

(e)**広告システムを適合**させ、提供するサービスに**関連する広告の提示を制限または調整することを目的とした的を絞った措置を採用**すること

(f)特に**システミックリスクの検知に関して**、その活動の**内部プロセス**、**リソース**、**テスト**、**文書化**、または**監督**を**強化する**こと

(g)第22条に従った**Trusted Flaggerとの協力**、および第21条に従った**裁判外の紛争解決機関の決定の実施**を開始または調整すること

(h)第45条および第48条にそれぞれ言及される**行動規範**および**危機プロトコルを通じて**、**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他のプロバイダーとの協力を開始**または**調整する**こと

(i)サービスの受領者により多くの情報を提供するために、**啓発措置**を講じ、**オンラインインターフェースを適合**させること

(j)児童の権利を保護するために、**年齢認証やペアレンタルコントロールツール**、**未成年者が虐待を通報したり支援を受けたりするのを支援するためのツール**など、**的を絞った措置**を適宜講じること

(k)**既存の人物、物、場所、その他の実体や事象に著しく類似し、真正または真実であるかのように人に誤認させるような情報**の項目が、生成または操作された画像、音声または動画であるかどうかにかかわらず、**オンラインインターフェースに表示される際に目立つマークによって区別できるようにし**、さらに、**サービスの受信者がそのような情報を示すことができる使いやすい機能を提供する**こと

出所）DSA　https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32022R2065

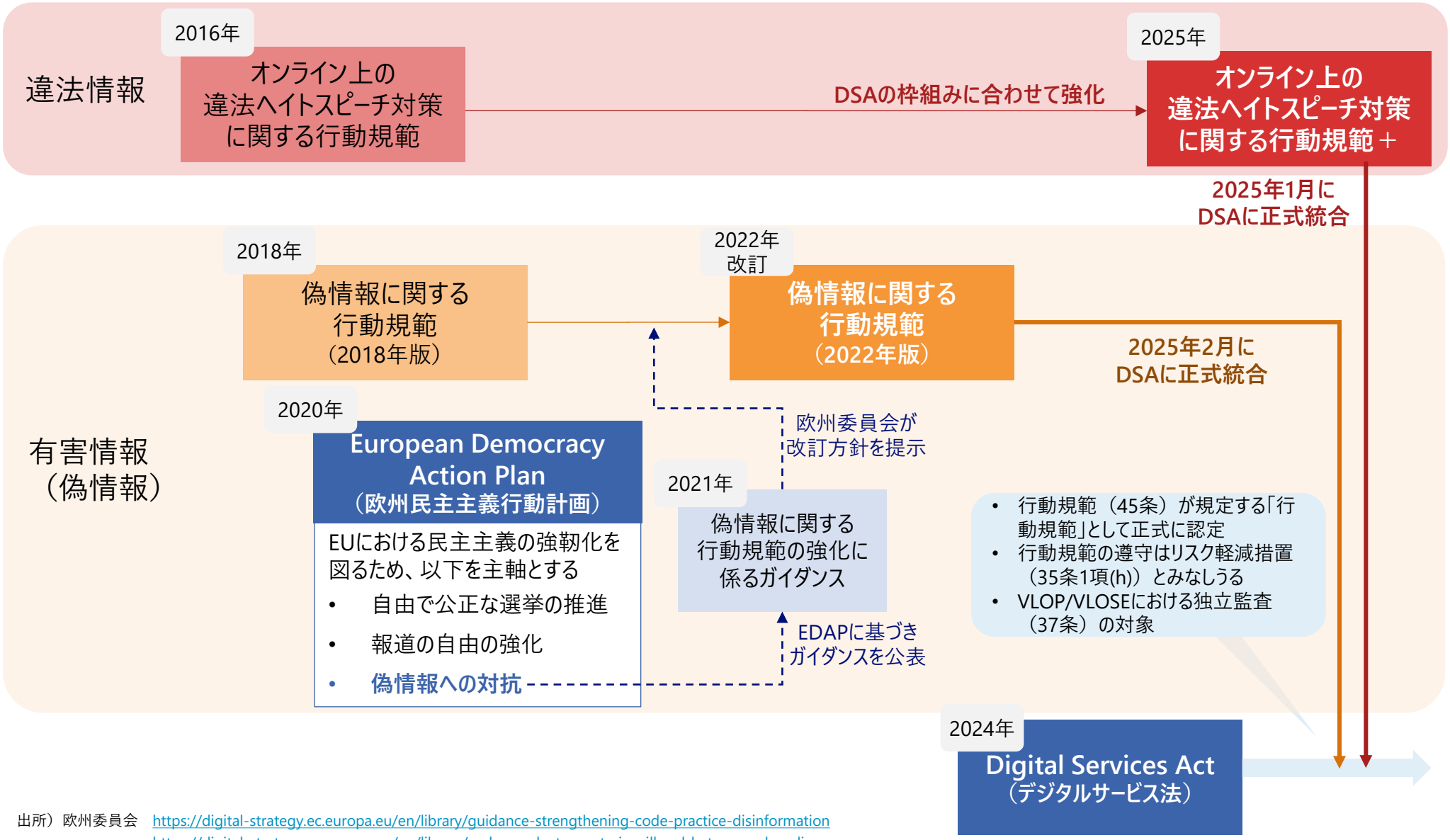
（参考）35条　リスク軽減（2/2）

■ 前文

(104)　・・・検討すべきもう一つの領域は、**偽情報や操作および濫用行為、未成年者への悪影響**など、社会および民主主義に対するシステミックリスクの潜在的な負の影響である。これには、**ボットや偽アカウントを使用して故意に不正確または誤解を招く情報を流すなど、偽情報を含め、情報の増幅を目的とした協調的な活動が含まれる**。このような活動は、経済的利益を得ることを目的としている場合もあり、特に未成年者など、サービスの受け手として弱い立場にある人々にとって有害である。このような分野に関連して、超大規模オンラインプラットフォームまたは超大規模オンライン検索エンジンが特定の行動規範を遵守し、それに従うことは、適切なリスク軽減措置であると考えられる。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提供者が、当該行動規範の適用への参加を求める委員会の要請を適切な説明なく拒否した場合は、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で定められた義務に違反しているかどうかを判断する際に、関連性がある場合には考慮される可能性がある。特定の行動規範への参加および実施という事実だけでは、それ自体では本規則の遵守を推定すべきではない。

出所）DSA　https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32022R2065

EUの違法情報・有害情報に関する政策の全体像



出所）欧州委員会　https://digital-strategy.ec.europa.eu/en/library/guidance-strengthening-code-practice-disinformation
https://digital-strategy.ec.europa.eu/en/library/code-conduct-countering-illegal-hate-speech-online

（参考）DSAと行動規範の関係

- DSA35条において、**軽減措置の**１つとして行動規範を通じた、オンラインプラットフォーム等との協力・調節（35条1項(h)）を義務付け。
- 行動規範については、DSAでは、欧州デジタルサービス会議（※）は、DSAの適切な適用を促進するために、EUレベルでの自主的な行動規範の策定を奨励し、支援することとなっている（45条1項）ところであるが、DSAに先駆け、以下の行動規範が策定されており、当該行動規範については、2025年にDSAに基づく行動規範（45条）として認定。
- なお、VLOP/VLOSEが署名団体である場合、**行動規範の遵守状況はDSAの独立監査（37条）の対象**。

	特定の違法情報	有害情報
名称	オンライン上の 違法ヘイトスピーチ対策に関する行動規範	偽情報に関する行動規範
策定の背景	EU法および加盟国法で違法となるヘイトスピーチと見なされるオンラインコンテンツへの対処強化を目的に2016年に策定	民間事業者において、偽情報への対抗を目的に2018年に策定（その後、2022年に改訂）
DSA枠組みへの統合	2025年1月に認定	2025年2月に認定
署名事業者	プラットフォーム事業者12者（うちVLOP7者*を含む） *Facebook, Instagram, LinkedIn, Snapchat, TikTok, X, YouTube	プラットフォーム事業者、広告団体など40者以上（VLOP/VLOSEを含む*） *Facebook, Google Search, Instagram, Bing, LinkedIn, TikTok, YouTube　※Xは2023年に脱退

※各国のDSCの代表と欧州委員会で構成される独立した諮問会議

出所）欧州委員会 https://digital-strategy.ec.europa.eu/en/policies/dsa-codes-conduct
https://digital-strategy.ec.europa.eu/en/news/european-commission-launches-workshops-explore-voluntary-codes-conduct-online-advertising

（参考）偽情報に関する行動規範―概要

プラットフォーム事業者や広告業界等を巻き込んだ自主規制として策定され、欧州委員会の関与を強める形で改訂された。偽情報の削減や拡散防止のための広範な措置が求められる

項目	内容
策定の背景	• 2018年10月にオンラインプラットフォーム、大手テック企業、広告団体が自主規制の枠組みとして策定し、2019年に署名事業者・団体による自己評価報告書が公表された。 • 2021年5月に欧州委員会が「偽情報に関する行動規範の強化に係るガイダンス」を公表し、2022年6月に欧州委員会の関与が強化される形で「2022年偽情報に関する行動規範」として改訂された。 • **2025年2月に正式にDSA45条における「自主的な行動規範」として認定された。**
署名事業者・団体	• オンラインプラットフォーム：Google、Meta、Microsoft、TikTok、Twitchなど　※Xは脱退済み • そのほか、広告関連団体、ファクトチェック団体、市民団体が参加
対象情報	• 偽情報（disinformation）が対象。 偽情報：人を欺いたり、経済的・政治的利益を確保したりする意図で流布される虚偽または誤解を招く内容であり、公衆に害を及ぼす可能性がある情報
コミットメントの概要	• 偽情報の拡散者に対する金銭的インセンティブの削減 • 政治広告のラベル表示等の政治広告の透明性確保 • 偽アカウントによる偽情報の拡散、ボットによる増幅、悪意のあるディープフェイク等の削減 • ユーザーが偽情報を識別できるツール、研究者のデータへの広範なアクセス、ファクトチェックの実施

出所）欧州委員会 https://ec.europa.eu/commission/presscorner/detail/en/ip_25_505
https://digital-strategy.ec.europa.eu/en/policies/code-practice-disinformation

（参考）オンライン上の違法ヘイトスピーチ対策に関する行動規範―概要

テロ事件を契機としてヘイトスピーチ対策強化のため策定され、DSAに合わせて改訂された。違法ヘイトスピーチコンテンツに関する通知の迅速なレビューと対処を求めている

項目	内容
策定の背景	• 2016年3月のブリュッセルでのテロ攻撃後、臨時司法・内務理事会の共同声明において、「**テロリストによるプロパガンダに対抗**し、2016年6月までにオンライン上のヘイトスピーチに対する行動規範を策定するため、IT企業との協力を強化する」と言及されたことを受け、欧州委員会と事業者の合意のもと、2016年5月に策定。 • 違法ヘイトスピーチコンテンツにおけるDSAの遵守と効果的な執行を促進する枠組みとして、改訂され、**2025年1月に「オンライン上の違法ヘイトスピーチ対策に関する行動規範＋」として正式にDSA45条における「自主的な行動規範」として認定された。**
署名事業者	• 策定時点：Facebook（現Meta）、Microsoft、Twitter（現X）、YouTube • 2018年以降に参加：Dailymotion、Snapchat、Jeuxvideo.com、TikTok、Rakuten Viber、Twitch
対象情報	• オンライン上の**違法ヘイトスピーチ**が対象。具体的には加盟国法により定義されるが、EUの理事会枠組決定で以下の定義が示されている。 違法ヘイトスピーチ：人種、肌の色、宗教、家系、あるいは国籍や民族的な出自を参照して定義された個人またはグループに対して、暴力や憎悪を公然と扇動する行為
コミットメントの概要	• 署名事業者は、DSAに従って違法ヘイトスピーチの禁止を利用規約に明示（DSA14条）し、通知と行動の仕組みを提供（16条）し、通知が利用規約に違反する場合にコンテンツの削除またはアクセス無効化を迅速に行うこと。 • モニタリング報告者（専門性を有する非営利団体や公的団体を任命）による違法ヘイトスピーチコンテンツの通知の50%以上を24時間以内にレビューすること。 • 署名事業者は通知のレビュー状況をモニタリングし、欧州委員会への情報提供やユーザー啓発に努めること。

出所）欧州委員会 https://digital-strategy.ec.europa.eu/en/library/code-conduct-countering-illegal-hate-speech-online
https://commission.europa.eu/strategy-and-policy/policies/justice-and-fundamental-rights/combatting-discrimination/racism-and-xenophobia/eu-code-conduct-countering-illegal-hate-speech-online_en
欧州連合 “Council Framework Decision 2008/913/JHA of 28 November 2008 on combating certain forms and expressions of racism and xenophobia by means of criminal law”
https://eur-lex.europa.eu/eli/dec_framw/2008/913/oj/eng

DSAと行動規範の関係性

（参考）45条　行動規範

■ 第45条 行動規範

1　欧州委員会および欧州デジタルサービス会議は、特に競争法および個人情報の保護に関するEU法に従い、さまざまな種類の違法コンテンツおよびシステミックリスクへの取り組みという特定の課題を考慮しつつ、本規則の適切な適用に貢献するため、**EUレベルでの自主的な行動規範の作成を奨励し、促進する**ものとする。

2　34条1項の意味における重大なシステミックリスクが出現し、複数の超大規模オンラインプラットフォームまたは超大規模オンライン検索エンジンに関係する場合、欧州委員会は、関係する超大規模オンラインプラットフォームのプロバイダーまたは超大規模オンライン検索エンジンのプロバイダー、および他の超大規模オンラインプラットフォームのプロバイダー、超大規模オンライン検索エンジンのプロバイダーを招待することができる、適切な場合には、オンラインプラットフォームおよびその他の仲介サービスのプロバイダー、ならびに関連する管轄当局、市民社会組織およびその他の関連する利害関係者に対し、特定のリスク軽減措置を講じることを約束すること、および講じられた措置とその結果に関する定期的な報告枠組みを定めることを含め、**行動規範の策定に参加するよう求める**ことができる。

3　1項および2項を適用するにあたり、欧州委員会およびEBDSおよび関連するその他の機関は、行動規範がそれぞれの特定の目的を明確に定め、それらの目的の達成度を測る主要業績評価指標を含み、また、特にEUレベルにおけるすべての利害関係者、とりわけ市民のニーズおよび利益を十分に考慮することを確保することを目的とする。また、欧州委員会および理事会は、参加者が、定められた主要業績評価指標に照らして、実施した措置およびその結果について、欧州委員会およびそれぞれのDSCに定期的に報告することを確保することを目的とする。主要業績評価指標および報告義務は、参加者の規模および能力の相違を考慮するものとする。

4　欧州委員会およびEBDSは、**行動規範が1項および3項に規定された目的を満たしているかどうかを評価**し、行動規範に含まれる主要業績評価指標を考慮しながら、その目的の達成状況を定期的に監視および評価するものとする。両者は、その結論を公表しなければならない。欧州委員会およびEBDSはまた、**行動規範の定期的な見直しと適応を奨励し、促進**するものとする。行動規範の遵守に組織的な不履行があった場合、欧州委員会およびEBDSは、**行動規範の署名団体に対し、必要な措置を講じるよう求める**ことができる。

出所）DSA　https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32022R2065

欧州委員会は、リスク評価、軽減措置を含むVLOP/VLOSEの義務について事前調査を行うことができ、法的手続を経て、違反を決定する

■ 緊急性が高い場合には欧州委からの暫定措置命令*のほか、VLOP/VLOSEから欧州委に対してコミットメント*の提案が可能。

*命令・コミットメントの例として、レコメンダシステムの変更、特定のキーワードやハッシュタグの監視の強化、または申し立てられた違反事項の終了または是正がある。

事前調査開始

- 遵守のモニタリング（第72条）や信頼できる情報源（例：協力する市民団体等）から得た情報を評価し、**DSA違反について欧州委員会が疑念を抱いた場合、欧州委員会は調査**を行う。欧州委員会は情報提供の要請（第67条）やインタビューの実施（第68条）、立入検査（第69条）を行うことができる。

法的手続の開始

- **事前調査後もDSA違反の疑念が残る場合**、欧州委員会は違反の決定（第73条）や制裁金を課すこと（第74条）を採択する可能性を視野に、**法的手続を開始。**欧州委員会は、欧州デジタルサービス会議やDSCと連携しながら、調査権限を行使（第66条）。
- 欧州委員会は、対象のVLOP/VLOSEに対し、欧州委員会の出した暫定的な調査結果について意見を述べる機会を与えなければならない（第79条）。

暫定措置・予備的見解

予備的見解1件、コミットメント1件

- 法的手続で**違反認定がなされる可能性があり**、かつ**サービス利用者に重大な損害が及ぶ緊急の可能性がある場合**、欧州委員会は明白な違反の証拠に基づき、VLOP/VLOSEに対し相応かつ**一時的な暫定措置を命じることを決定**できる（第70条）。
- 法的手続中に**VLOP/VLOSEが**遵守確保のための**コミットメントを申し出た場合**、欧州委員会は**コミットメントに拘束力を持たせた上で、法的手続の中止を宣言**できる。（第71条）
- 欧州委員会は対象の**VLOP/VLOSEに対し、欧州委員会の出した予備的見解（第73条）について意見を述べる機会を与えなければならない。**（第79条）。

違反の決定

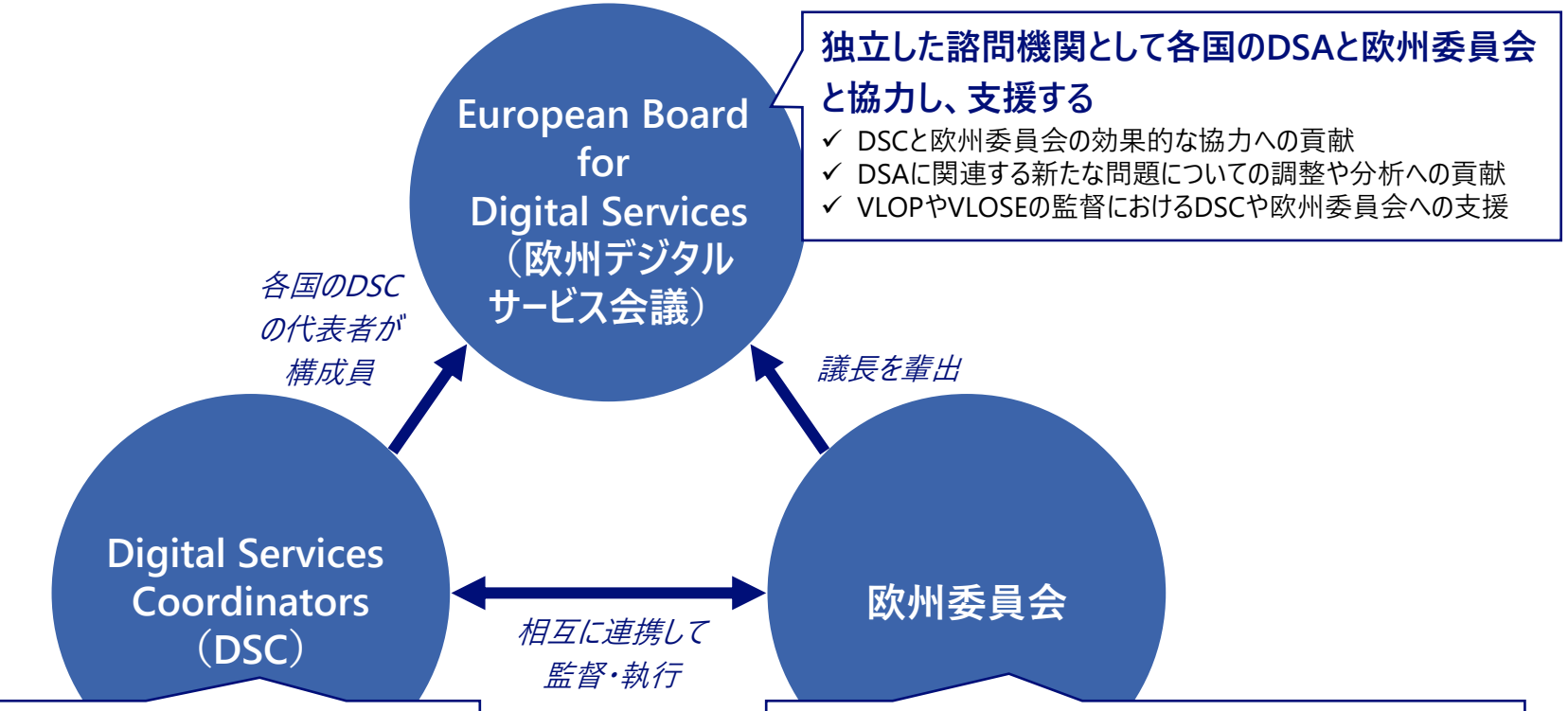
これまで0件

- VLOP/VLOSEが**DSAの関連規定への違反**や、**暫定措置やコミットメントに適合していない**と**欧州委員会が判断した場合、違反の決定**を行う（第73条）。
- 欧州委員会はVLOP/VLOSEに対し、遵守を確保するために必要な措置を指定期間内に講じること、遵守のために講じる措置の予定について情報提供することを命じる（第73条）。
- 違反が決定された場合、前会計年度における全世界年間売上高の6%を超えない制裁金を課すことができる（第74条）。

出所）DSAの各条文、" The enforcement framework under the Digital Services Act"を基にNRI作成
https://www.eu-digital-services-act.com/Digital_Services_Act_Articles.html
https://digital-strategy.ec.europa.eu/en/policies/dsa-enforcement

（参考）DSAにおいて、VLOP/VLOSEへの監督・執行は欧州委員会、その他事業者への監督・執行はDSCが担う

■ DSCの代表と欧州委員会で構成される「欧州デジタルサービス会議」がDSCと欧州委員会を支援する。



独立した諮問機関として各国のDSAと欧州委員会と協力し、支援する

- ✓ DSCと欧州委員会の効果的な協力への貢献
- ✓ DSAに関連する新たな問題についての調整や分析への貢献
- ✓ VLOPやVLOSEの監督におけるDSCや欧州委員会への支援

加盟国内におけるDSAの監督・執行の権限を有する

（ただし、VLOP・VLOSEに対する監督・執行は欧州委員会が担う）

- ✓ EU加盟国はDSAの適用および執行に責任を有する所管当局としてDSCを指定する
- ✓ DSCは独立した主体として、DSAの執行の権限と責任を国内において負うとともに、欧州委員会ならびに他の加盟国との連携を行う

VLOP・VLOSEの指定と監督・執行を担う

- ✓ 欧州委員会は、VLOP・VLOSEのみに課される追加義務について、独占的な監督および執行の権限を有する（追加義務以外についても、VLOP・VLOSEに対する監督・執行する権限を有する）
- ✓ 欧州委員会と各国のDSCは、DSAを一貫して効率的に適用するために、緊密に協力し、相互に援助しあう

出所）EU-Lex「Document 32022R2065」https://eur-lex.europa.eu/eli/reg/2022/2065/oj -
https://digital-strategy.ec.europa.eu/en/policies/digital-services-act-package

（参考）行政機関の制度の執行状況｜予備的見解の通知事例（X）

2023/12/18
法的手続き開始

- 欧州委員会は、XがDSAに違反している可能性があるかどうかを評価するための法的手続きを開始。
- 違法コンテンツの拡散に対するリスク評価・軽減措置（DSA34条、35条）、違法コンテンツの通報と適切な措置（16条）、情報操作に対抗するための措置の有効性（35条）、**ダークパターン（25条）、広告の透明性（39条）、研究者のためのデータアクセス（40条12項）**に関連する分野について調査が開始された。

2024/7/12
予備的見解の通知
*本通知はDSA初

- 欧州委員会は、XがDSAに違反しているという予備的見解をXに通知した。
- 欧州委員会の予備的見解が最終的に確認された場合、委員会は、X**がDSA25条、39条、および40条12項に違反**していると決定を採択することになる。

違反分野	内容
ダークパターン	• Xは、「青いチェックマーク」の付いた「認証済みアカウント」のインターフェースを、業界の慣例にそぐわない方法で設計・運用し、利用者を欺いている。 • 「認証済み」ステータスは誰でも取得でき、悪意ある主体がユーザーを欺くために「認証済みアカウント」を悪用しているエビデンスがある。
広告の透明性	• Xは、検索可能で信頼できる広告リポジトリを提供しておらず、その代わりに、リポジトリをユーザーに対する透明性の目的に適さない設計機能とアクセス障壁を設置しているため、広告に関する要求される透明性を遵守していない。
研究者のデータアクセス	• Xは、DSAに定められた条件に従って、研究者に公開データへのアクセスを提供していない。 • 特に、Xは、利用規約に記載されているように、資格を有する研究者がスクレイピング等によってXの公開データに独自にアクセスすることを禁止している。 • さらに、XのAPIへのアクセスを研究者に許可するプロセスは、研究者の研究プロジェクトの遂行を妨げるか、不相応に高い料金を支払う以外の選択肢を残していないように見える。

今後の動き

- Xは、欧州委員会の調査ファイル内の文書を閲覧し、書面で回答することにより抗弁権を行使できる。
- 並行して、欧州デジタルサービス委員会にも諮問が行われる。
- **予備的見解が最終的に確認された場合、XがDSA25条、29条、40条12項に違反していると認定される。**

出所）欧州委員会 “Commission opens formal proceedings against X under the Digital Services Act”, “Commission sends preliminary findings to X for breach of the Digital Services Act”
https://ec.europa.eu/commission/presscorner/detail/en/ip_24_3761
https://ec.europa.eu/commission/presscorner/detail/en/ip_23_6709

（参考）行政機関の制度の執行状況｜コミットメントの決定事例（TikTok）

- TikTok Liteは、TikTokアプリの別バージョンであり、TikTok Liteにおけるリワードプログラムは、ユーザーがTikTok Lite上で特定の「タスク」（動画の視聴、コンテンツへの「いいね！」、クリエイターのフォロー、友人のTikTokへの招待など）を実行する際にポイントを獲得できるというものである。

2024/4/22 法的手続開始

- TikTok Liteが2024年4月にスペイン・フランスでリリース。
- 欧州委員会は、利用者の身体的・精神的健康に影響を及ぼす可能性があるにも関わらず、**リリースにあたってのリスク評価レポートを提出していないことに関して**、TikTokがリスク評価・軽減措置（DSA34条、35条）に違反したかどうかを調査するために、法的手続を開始。同時にリワードプログラムを一時停止するように警告。

調査の焦点	内容
リスク評価の有無	• **システムリスクに重大な影響を及ぼす可能性のある機能（「Task and Reward Lite」プログラム）を展開する前に、リスク評価レポートを実施して提出するというDSAの義務を遵守したか否か。** • 特に、未成年者のメンタルヘルスを含むメンタルヘルスに悪影響が及ぶ可能性の評価有無。
軽減措置の有無	• リスクを軽減するためにTikTokは対策を講じたか否か。

2024/4/24 TikTokによるリワードプログラムの自主的停止

- 欧州委員会は、TikTokがTikTok Liteのリワードプログラムを自主的に停止したと発表。

2024/8/5 コミットメントの決定

- 欧州委員会は、TikTokのコミットメント（TikTokが4/22の法的手続をうけ提出したもの）に拘束力を持たせ、法的手続を終了した。これにより、コミットメントに違反した場合はDSA違反となる。
  *法的手続の終了、コミットメントの決定ともにDSA初
- コミットメントの内容は以下。
  - TikTok Liteリワードプログラムを EUから永久に撤退させること。
  - 撤退を回避するような他のプログラムを開始しないことを約束すること。

出所）欧州委員会 「TikTok commits to permanently withdraw TikTok Lite Rewards programme from the EU to comply with the Digital Services Act」
TikTok commits to permanently withdraw TikTok Lite Rewards (europa.eu)

リスク評価・軽減措置の遵守状況等について、第三者機関の監査を受けることを義務付け。また、監査での指摘への対応や、リスク評価の結果・軽減措置の報告をDSC等に行うことを事業者に義務付け

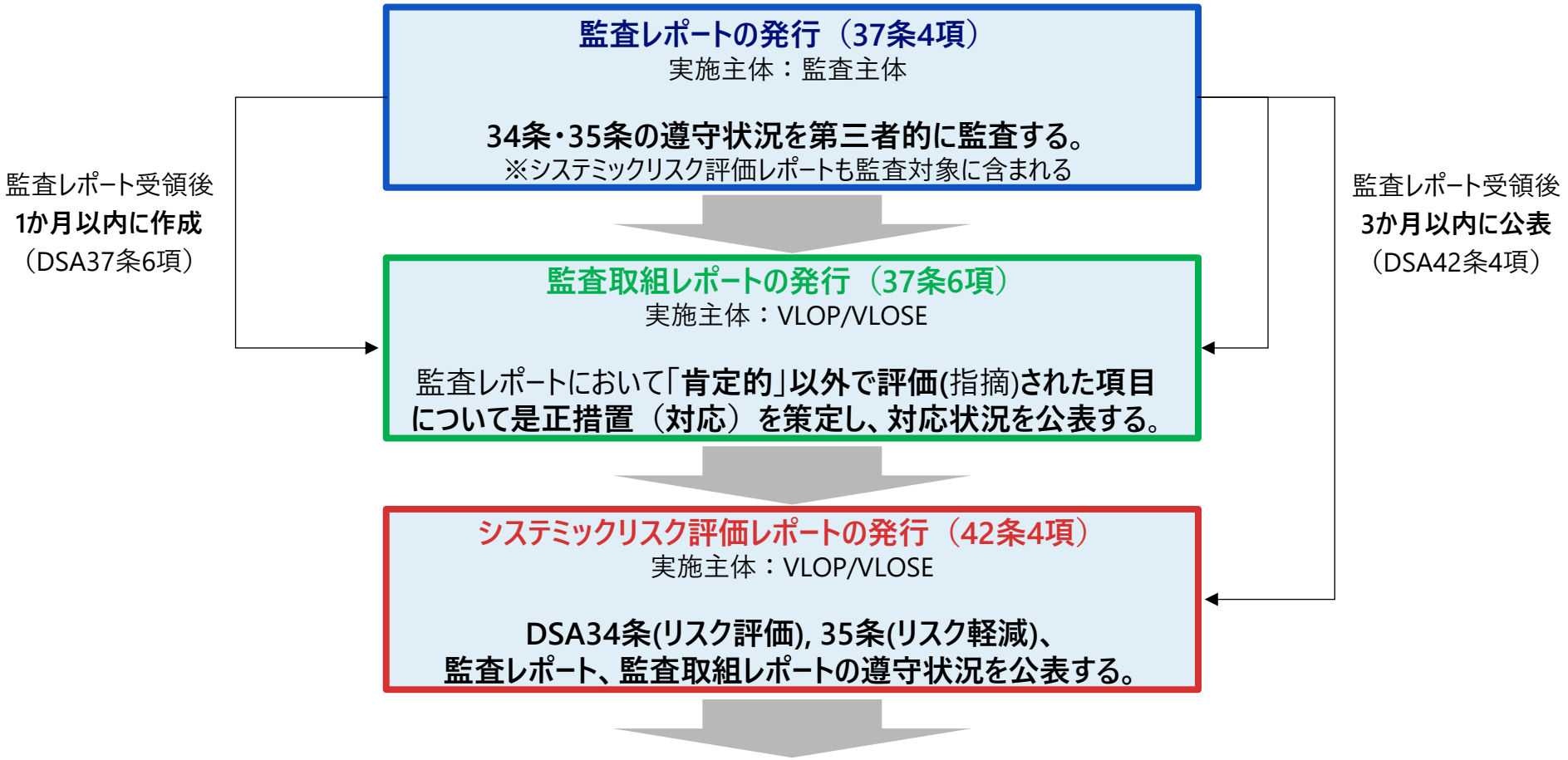
DSA37条・42条に基づいてVLOP/VLOSEが公表するレポート

	関連するDSA条文	発行主体	公表頻度	内容
監査レポート	• 37条（独立監査）4項 • 42条4項(c)	監査主体	年1回以上	• 監査主体は、リスク評価（34条）、軽減措置（35条）を含むVLOP/VLOSEのDSA第3章の義務の遵守状況、行動規範（45条・46条）の遵守状況、危機プロトコル（48条）に従ったコミットメント状況を評価。 • 評価結果は「肯定的」、「コメント付き肯定的」、「否定的」の3段階で評価（37条）。
監査取組レポート	• 37条（独立監査）6項 • 42条4項(d)	VLOP/ VLOSE	年1回以上	• 監査レポートで「コメント付き肯定的」・「否定的」と評価された項目について、監査レポートを受け取ってから１月以内に監査取組レポートを作成。 • 当該項目について、規定のフォーマットに従い、①監査レポートにおける評価、②監査結果を受けて今後実施する予定の取組/既に実施済みの措置、③推奨措置を実施しない理由等を記載
システミックリスク評価レポート	• 34条（リスク評価） • 35条（リスク軽減） • 42条4項(a)(b)	VLOP/ VLOSE	年1回以上	• 以下の情報を遅滞なくDSC・欧州委員会に提出するとともに、監査レポートの受領後、３ヶ月以内に公開 ・第34条に基づくリスク評価の結果 ・第35条に基づいて実施された軽減措置 ・第34条に基づくリスク評価の結果 ・第37条4項に基づく監査報告 ・第37条6項に基づく監査取組報告

出所）DSA37条・42条を基にNRI作成
https://www.eu-digital-services-act.com/Digital_Services_Act_Article_37.html
https://www.eu-digital-services-act.com/Digital_Services_Act_Article_42.html

DSAの遵守状況は監査レポートを通して第三者に監査され、不備があれば監査取組レポートで是正措置を報告する必要があり、事業者はシステミックリスク評価レポートで遵守状況を報告。

DSA34条・35条違反による法的手続き開始までの流れの例



出所）DSA37条・42条を基にNRI作成
https://www.eu-digital-services-act.com/Digital_Services_Act_Article_37.html
https://www.eu-digital-services-act.com/Digital_Services_Act_Article_42.html

（参考）事業者の取組状況｜監査レポート概要

各社の監査レポート及び監査取組レポートにおける35条1項に関する記載内容（抜粋）

監査レポート、監査取組レポートにおける35条1項に関する記載（仮訳、抜粋）

肯定的、コメント付き肯定的、否定的

	YouTube（監査主体：EY）（2023年8月-2024年5月）	Meta（監査主体：EY）（2023年8月-2024年5月）	TikTok（監査主体：KPMG）（2023年8月-2024年6月）	X（監査主体：FTI Consulting）（2023年8月-2024年8月）
監査レポート（37条4項）	被監査事業者は監査対象期間中、すべての重要な点において該当義務を遵守した。	監査期間中、MetaはDSA35条1項のを遵守しているかどうかについて政府機関による調査が継続中である。（中略）従って、当監査主体は、監査事項に関して結論を表明しない。	欧州委員会は、デジタルサービス法第35条第1項への潜在的な不遵守を評価するため、2024年2月19日付、2024年4月22日付でTikTokに対する正式な手続きを開始した。（中略）結果、監査不可。	Xは、システムが特定できなかった措置対象コンテンツのうち、ユーザーから報告されたコンテンツがどれだけ存在するかによってリスク管理の効果を測定していると述べている。しかし、監査中、これ以外のリスク管理効果の測定方法は確認されなかった。そのため、X はリスク軽減戦略が実際にリスクを軽減しているかを確認できていないと考えられる。
監査取組レポート（37条6項）	-（「肯定的」なためコメントなし）	–（監査結果がないため、記載なし）	–（監査結果がないため、記載なし）	Xは現在、欧州委員会による調査を受けている。この条文の範囲と解釈に関する規制上の不確実性は依然として大きく、監査主体の調査結果が欧州委員会の決定と矛盾するリスクがある。そのため、Xは監査主体の勧告を実施するつもりはなく、調査における欧州委の決定を待ちたい。

※DSA35条2項・3項は欧州デジタルサービス会議と欧州委を対象にした条文であるため、1項のみが監査の対象である。

出所）P.27の監査レポート、監査取組レポートを参照

（参考）事業者の取組状況｜事業者による軽減措置の対応：Google, TikTok, X (1/2)

システミックリスク評価レポートにおける35条１項の各号に対応する記載内容（抜粋）

3つの事業者がレポート内で言及している軽減措置

対応条文	条文内容	Google（2023年7月-2024年6月）	TikTok（一部抜粋）（2022年4月-2023年3月）	X（2023年7月31日時点）
(a)	サービスの設計や機能の調整	• コンテンツモデレーションでサポートされる言語の数増加に注力(Maps) • 監督ツールの強化(YouTube)	• 低年齢者のアプリダウンロードを制限[1] • 「Know the Facts」[2]による検索の制限 • 「国家統制メディア」[3]の表示	• コミュニティノートの運用によるコンテンツモデレーションへのユーザーの巻き込み • 報告メカニズムの採用 • 苦情システムの採用 • アカウントの認証拡大
(b)	利用規約やその適用方法の調整	• 広告悪用抑制のためのセーフガード改善 • ヘイトスピーチ、ハラスメント、ネットいじめに関する方針の更新(以下、YouTube) • 誤情報に関するポリシーの更新 • アプリの敵対的乱用を禁止	• コミュニティガイドラインを用いた禁止事項の明記（リスクや利用可能年齢） • 利用規約とコミュニティ・ガイドラインの併用	• 特定のリスクに対するポリシーの改善 • 不正使用に対する措置と保護
(c)	コンテンツモデレーションの調整	• アプリレビューで見られる俗語の削減(Play)	• For You Feed(FYF)の適格性基準維持 • 特にヌード、性行為コンテンツの消去 • 年齢による閲覧制御	• ユーザーの報告を用いたモデレーションの改善 • 違反判断時に別のスタッフとの意見交換 • 事前検知の強化
(d)	アルゴリズム（レコメンダシステムを含む）のテスト・調整	• 個人的苦難（例：経済的な問題、深刻な健康状態）への対処支援改善(Search) • 偽コンテンツ拡散防止の強化(Search) • 暴力描写のぼかし機能改善(Search) • 暴力の自動検出を強化(Search) • ジェンダーに基づく暴力による虐待の評価(Search) • 誤情報の自動検出強化(YouTube)	• 特定のハッシュタグやリスクの検索の禁止 • アップロード時点での違反の検出 • コンテンツが一定の人気を集めた段階で、そのコンテンツを人為的に監視 • ユーザー自身が推奨される動画を管理できるシステムの構築	• アルゴリズムシステムの検証 • 重大なシステムの変更に対して、個人情報要件に準拠しているかの検証実施

※Metaについても軽減措置をレポート内で記載しているが、Google, TikTok, Xと異なる形式で軽減措置を公表しているため、本ページには記載していない。

1) 登録時に年齢を記載（自己申告）が必要
2) ファクトチェックパートナーがコンテンツが偽情報か否かを結論付けられなかった場合に、コンテンツに付与されるバナー
3) 編集や意思決定が政府の影響を受けているメディア事業体によって運営されていると判断したアカウント

出所）P.27記載のシステミックリスク評価レポートより、**事業者自身が該当条文に直接紐づけている箇所をNRI抜粋。実際の取組有無を評価しているものではないことに注意。**

（参考）事業者の取組状況｜事業者によるリスク軽減措置の対応：Google, TikTok, X (2/2)

システミックリスク評価レポートにおける35条１項の各号に対応する記載内容（抜粋）

3つの事業者がレポート内で言及している軽減措置

対応条文	条文内容	Google（2023年7月-2024年6月）	TikTok（一部抜粋）（2022年4月-2023年3月）	X（2023年7月31日時点）
(e)	広告システムの調整		• 広告に関するガイドラインの発行	• 新しい広告透明性センターの運用
(f)	システミックリスクの監督強化	• グーグルの各サービス間の情報共有の改善と詐欺検知の向上 • セキュリティ、詐欺・スキャム対策の強化 • 機密データへのアクセス防止策強化(Search) • 悪意のある（ポリシー違反等）アプリの削除(Play)	• チャイルド・セーフティ・チームや安全リスク分析チーム、法執行対応チームによる監視 • 各システミックリスクへの軽減措置適応（特に未成年、13歳未満）	• データの監視 • 新規事業の評価 • プライバシー・ポリシーに関する研修の強化 • 特定分野の専門家の増員 • 特定リスクの調査、安全機能の拡充
(g)	Trusted Flaggerとの協力/ADRの決定適用		• コミュニティ・パートナー・チャンネル[4]の運営	• Trusted flaggersによる報告の優先処理
(h)	行動規範や危機対応プロトコルを通じた他のVLOP/VLOSEとの協力		• 他PF事業者との協力 • 業界共有のハッシュリスト[5]を利用した監視	• 社外との連携強化（例：違法情報のデータ共有、市民団体からの情報共有） • ステークホルダー（外部組織や専門家）や業界との連携
(i)	サービスの受け手に多くの情報を提供する		• コミュニティガイドラインやオンライン上での未成年保護に関するプログラムの実施	• 理由通知(DSA17条)による透明性向上 • 広告関連の情報を公開するサイトの運用 • ユーザーへのポリシーに関する教育強化
(j)	未成年者の保護強化		• 安全性と児童の権利/自由のバランスの維持 • 憲章24条(児童の権利)等を参考	
(k)	誤解を生むAI生成・操作コンテンツの明確化		• なりすましでTikTokアカウントを活用することの禁止	

※Metaについても軽減措置をレポート内で記載しているが、Google, TikTok, Xと異なる形式で軽減措置を公表しているため、本ページには記載していない。

4) DSAのTrusted Flaggerに類似したTikTok独自の制度

5) TikTokは既知の有害コンテンツをまとめたInternet Watch Foundation等のハッシュリストを用いてコンテンツの削除・報告を行っている

出所） P.27記載のシステミックリスク評価レポートより、**事業者自身が該当条文に直接紐づけている箇所をNRI抜粋。実際の取組有無を評価しているものではないことに注意。**

（参考）（出所）本資料作成時に用いたレポート

レポート	PF事業者	リンク
監査レポート	YouTube (Google)	https://storage.googleapis.com/transparencyreport/report-downloads/dsa-audit-youtube_2023-8-28_2024-5-31_en_v1.pdf
	Instagram (Meta)	https://transparency.meta.com/sr/dsa-independent_audit_report_Aug_2023-Jun_2024_instagram
	TikTok	https://sf16-va.tiktokcdn.com/obj/eden-va2/zayvwlY_fjulyhwzuhy[/ljhwZthlaukjlkulzlp/DSA_H2_2024/TikTok-DSA-Audit-Assurance-Report-Sep-2024.pdf
	X	https://transparency.x.com/content/dam/transparency-twitter/dsa/dsa-audit/TIUC-DSA-Audit-Report-2024-08-27.pdf
監査取組レポート	Google	https://storage.googleapis.com/transparencyreport/report-downloads/dsa-audit-google-implementation_2023-8-28_2024-5-31_en_v1.pdf
	Instagram (Meta)	https://transparency.meta.com/sr/dsa-audit_implementation_report_Sept_2024_instagram
	TikTok	https://sf16-va.tiktokcdn.com/obj/eden-va2/zayvwlY_fjulyhwzuhy[/ljhwZthlaukjlkulzlp/DSA_H2_2024/TikTok-DSA-Audit-Implementation-Report-Sep-2024.pdf
	X	https://transparency.x.com/content/dam/transparency-twitter/dsa/dsa-audit/TIUC-DSA-Audit-Implementation-Report-2024-09-27.pdf
システミックリスク評価レポート	Google	https://storage.googleapis.com/transparencyreport/report-downloads/dsa-risk-assessment_2024-8-28_2024-8-28_en_v1.pdf
	Instagram (Meta)	https://transparency.meta.com/sr/dsa-sra_results_report-2024-instagram
	TikTok	https://sf16-va.tiktokcdn.com/obj/eden-va2/zayvwlY_fjulyhwzuhy[/ljhwZthlaukjlkulzlp/DSA_H2_2024/TikTok-DSA-Risk-Assessment-Report-2023.pdf
	X	https://transparency.x.com/content/dam/transparency-twitter/dsa/dsa-sra/TIUC-DSA-SRA-Report-2023.pdf

1－2．英国におけるOSA （Online Safety Act）の リスク評価・軽減措置

リスク評価・軽減措置の規定

OSAでは、リスク評価と安全義務（軽減措置）を、それぞれ9条、10条で定めている

Part3「ユーザー間サービスや検索サービスに課される義務」 2章：ユーザー間サービスの注意義務

セクション	条項	タイトル	カテゴリー２サービス	カテゴリー1サービス
ユーザー間サービス：義務の範囲	7	ユーザー間サービスの提供者：注意義務	●	●
	8	注意義務の範囲	●	●
ユーザー間サービスの違法コンテンツ義務	9	違法コンテンツのリスク評価義務	●	●
	10	違法コンテンツに関する安全義務	●	●
子供にアクセスされる可能性の高いサービス	11	子供のリスク評価義務	●	●
	12	子供を保護するための安全義務	●	●
	13	子供を保護するための安全義務：解釈	●	●
カテゴリー1サービス ※以下のいずれかの条件を満たすサービス。 ①コンテンツレコメンダシステムを有し、英国人口の約50%に相当する3,400万人以上の英国ユーザーを有する ②ユーザー生成コンテンツの転送又は再共有をユーザーに許可し、コンテンツレコメンダシステムを有し、英国人口の10%を占める700万人以上の英国ユーザーを有する	14	アセスメント義務：ユーザーのエンパワーメント		●
	15	ユーザーのエンパワーメント義務		●
	16	ユーザーのエンパワーメント義務：解釈		●
	17	民主的に重要性のあるコンテンツを保護する義務		●
	18	ニュースパブリッシャーのコンテンツを保護する義務		●
	19	ジャーナリスティックコンテンツを保護する義務		●
コンテンツに関する報告および苦情処理手続きに関する義務	20	コンテンツ報告に関する義務	●	●
	21	苦情処理手続きに関する義務	●	●
横断的義務	22	表現の自由とプライバシーに関する義務	●	●
	23	記録保持とレビュー	●	●

出所）https://www.legislation.gov.uk/ukpga/2023/50/contents/enacted

（参考）OSAにおける施行スケジュール

OSAにおける各種義務は、義務の対象となるサービスの種類に応じて、3段階に分けて施行する予定となっており、Ofcomには、情報の種類等に応じて（フェーズ1~3）、行動規範およびガイダンスの発行が義務付けられている

■ OSAの監督・執行は、Ofcomが担う。

- OSAの施行に際しては、**Ofcomに対して、オンラインサービス事業者に課される義務に対する行動規範（Code of Practice）の公表が義務付けられている（41条）**
- また、PFサービス事業者が同法が定める義務の遵守を支援するための**ガイダンスを発行することも義務付けている（52条、53条、54条等）**

Ofcomによる、行動規範やガイダンスの整備

	フェーズの概要	ステータス　灰字：今後の予定
フェーズ1: 全てのサービスに課される義務	• 全てのサービスに課される義務に関する行動規範やガイダンスを整備 • リスク評価義務（9条）、安全措置義務（10条）、ユーザーからのコンテンツ報告・苦情受付義務（20条、21条）、テロコンテンツ等への対処通知義務（121条）、CSEAコンテンツのNCAへの報告義務（66条）等が該当	• （2023年11月）行動規範・ガイダンスに関するパブコメを公表 • （2024年12月）パブコメを受け、行動規範・ガイダンスを確定 • （2025年3月17日~）事業者は、リスク軽減措置を講じる義務を負う
フェーズ2: 子供にアクセスされる可能性が高いサービスに課される義務	• 子供にアクセスされる可能性が高いサービスに課される義務に関する行動規範やガイダンスを整備	• （2024年5月）行動規範・ガイダンスに関するパブコメを公表 • （2025年4~6月）行動規範・ガイダンスを確定予定
フェーズ3: 大規模サービスに課される義務	• 大規模サービス（特定カテゴリーサービス）に課される義務に関する行動規範やガイダンスを整備 • 本資料で取り上げている義務では、本人確認義務（64条、65条）が該当	• （2024年3月）行動規範・ガイダンス作成のためのエビデンス募集を開始 • （2025年1~3月）行動規範・ガイダンスのパブコメ予定 • （2025年10~12月）行動規範・ガイダンスを確定予定

出所）Statement: Protecting people from illegal harms online – Ofcom等よりNRI作成

リスク評価に関する制度｜第９条「違法コンテンツのリスク評価義務」

■ 第9条（2）では、すべてのユーザー間サービスの提供事業者に対し、
「**違法コンテンツのリスク評価**」として、以下の事項等を評価することを義務付けている。

■ **リスク評価は、サービス開始から３ヶ月以内※1、サービスの設計・運用に重大な変更をするとき、OFCOMが策定するリスクプロファイル※2を更新したときに実施する。**

※1 違法コンテンツのリスク評価に関するガイダンスが策定される前に既にサービスを行っている場合は、最初のガイダンス策定後 3 月以内に実施することとされている。当該ガイダンスは2024年12月16日に策定されたため、既存のサービス提供事業者は2025年3月16日までにリスク評価が必要となる。

※2 OFCOMがサービスによってもたらされるリスクを特定・評価し、サービスの特性やリスクレベルを整理したもの。

■ サービス提供者が評価する必要がある事項は、以下のとおり。

評価事項	条文
ユーザーベース	9条5項(a)
当該サービスの利用者が違法コンテンツ（優先違法コンテンツ、その他の違法コンテンツ）に遭遇するリスクレベル	9条5項(b)
当該サービスが優先犯罪の実行または助長に使用されるリスクレベル	9条5項(c)
異なる種類の違法コンテンツや、優先犯罪の実行または助長のためのサービスの使用が個人に与える危害	9条5項(d)
サービスの機能が違法コンテンツの存在や拡散、または優先犯罪の実行・助長を促進するリスクレベル	9条5項(e)
サービスの使用方法の多様性、およびその使用が個人に生じうる危害のリスクレベルに与える影響	9条5項(f)
(b) から (f) で特定された事象によって、個人が被る可能性のある危害の性質および深刻さ	9条5項(g)
サービスの設計および運用が、特定されたリスクを低減または増加させる可能性	9条5項(h)

出所）https://www.ofcom.org.uk/online-safety/illegal-and-harmful-content/statement-protecting-people-from-illegal-harms-online/
Online Safety Act 2023よりNRI作成

（参考）9条　違法コンテンツのリスク評価義務

1. 本条は、すべての規制対象ユーザー間サービスに適用される、リスク評価に関する義務を定める。
2. 適切かつ十分な違法コンテンツのリスク評価を、附則3に定める時期またはそれに基づき実施する義務。
3. リスク評価を最新の状態に保つための適切な措置を講じる義務。これには、オフコムが該当するサービスのリスクプロファイルに重大な変更を加えた場合も含まれる。
4. サービスの設計または運用のあらゆる側面に重大な変更を加える前に、提案される変更の影響に関連する適切かつ十分な違法コンテンツのリスク評価を実施する義務。
5. 特定の種類のサービスにおける「違法コンテンツのリスク評価」とは、以下の事項を評価することを指し、当該サービスのリスクプロファイルを考慮に入れるものとする：
   (a) 利用者層の特性、
   (b) 当該サービスを介して以下の内容に接触するリスクの程度：
      (i) 優先的違法コンテンツの各種（それぞれ別個に評価）、
      (ii) その他の違法コンテンツ、
      ※特に、サービスで使用されるアルゴリズム、およびコンテンツがどの程度容易に、迅速に、広範囲に拡散されるかを考慮すること。
   (c) サービスが優先的犯罪の実行または支援に使用されるリスクの程度、
   (d) 違法コンテンツの種類やサービスが優先的犯罪の実行または支援に使用されることによる個人への害のリスクの程度、
   (e) サービスの機能性が違法コンテンツの存在または拡散、もしくは優先的犯罪の実行または支援を助長するリスクの程度を高める方法を特定し評価すること、
   (f) サービスがどのように使用されるかの多様性、およびその使用が個人に対する害のリスクに与える影響、
   (g) 上記(b)～(f)で特定された事項に起因する個人に対する害の性質および重大性、
   (h) サービスの設計および運用（ビジネスモデル、ガバナンス、積極的な技術の使用、ユーザーのメディアリテラシー向上措置、安全な利用を促進する措置、その他のシステムやプロセスを含む）が特定されたリスクを増加または軽減する方法。
6. 本条における「リスクプロファイル」は、第98条に基づき公開されている、違法コンテンツが個人に与える害のリスクに関連するプロファイルを指す。
7. 本条に関連して以下も参照のこと：
   (a) 第23条第2項および第10項（リスク評価の記録）、
   (b) 附則3（提供者による評価時期）。

出所）Online Safety Act 2023よりNRI作成

OSAに基づきOfcomは、2024年12月にリスク評価の要件などを定めたガイダンスを策定

- Ofcomに対して、OFCOMが、ユーザー間サービスなどがもたらす危険を評価し、評価結果を各リスクごとに、サービスのリスクプロファイル（危険度などの特徴）を作成・公表を義務付けており（98条）、リスクプロファイルの公表後、違法コンテンツ等のリスク評価義務を果たすのを支援するためのガイダンスを作成・公表を義務付け（99条）。
- OSAに基づきOfcomは2024年12月16日、**違法コンテンツのリスク評価の要件やプロセス、リスクプロファイルなどを定めたガイダンス（"Risk Assessment Guidance and Risk Profiles"）を公表。これを受けて、既存のサービス提供事業者は2025年3月16日までにリスク評価が必要となる。**ガイダンスの内容は以下のとおり。

「適正かつ十分な」リスク評価の要件

No	要件
1	• ユーザーが17種類の優先違法コンテンツおよびその他の**違法コンテンツに遭遇するリスク**、およびユーザー間サービスが**優先犯罪の実行に使用されるリスク**を評価すること。
2	• Ofcomの**リスクプロファイルを考慮**すること。
3	• **サービスの特性を考慮**すること。 例：ユーザーベース、機能、アルゴリズムシステム（およびコンテンツをいかに容易に、迅速に、広範囲に配信するか）、ビジネスモデル等。
4	• **サービス設計および運営に関するその他の関連事項を考慮**すること。 例：ガバナンス、先進技術の利用、ユーザーのメディアリテラシーおよび貴社のサービスの安全な利用を促進するための措置、リスクレベルに影響を与える可能性のあるその他のシステムやプロセス等。
5	• **サービスがどのように利用されているか**を考慮すること。

リスク評価から軽減措置実施までのプロセス

	アクティビティ
Step 1. 評価すべき違法コンテンツの種類の理解	• **リスクプロファイルを参照し、優先違法コンテンツについて、サービスに関連する主なリスク要因を特定する** • その他の違法コンテンツが発生するリスクがあるかどうかを特定する
Step 2. 危害リスクの評価	• 優先違法コンテンツ、およびその他の違法コンテンツについて、関連するエビデンスを基に、**発生可能性と影響を個別評価する** • リスクのレベルに影響を与える可能性のある、既存の管理対策の有用性を評価する • **違法コンテンツにリスクレベルを割り当てる**
Step 3. 対策の決定・実施・記録	• 適切な対策を決定する（既存の措置の有効性評価、追加対策の検討） • 対策を適切に実施し、結果を記録する
Step 4. 報告・レビュー	• リスク評価と対策について、適切なルートで報告する • 対策の有効性を監視する • 適切な対策が実施された後のリスクの露出レベルを監視する • リスク評価の見直し・更新を行う

出所）https://www.ofcom.org.uk/online-safety/illegal-and-harmful-content/statement-protecting-people-from-illegal-harms-online/ よりNRI作成

Step1：サービスプロバイダーは、Ofcomが策定したリスクプロファイルを使用し、自社サービスが持ついかなるリスク要因が、違法コンテンツに紐づくのかを把握可能

サービスプロバイダーによる、リスクプロファイルの使用イメージ

質問への回答

例）「あなたのサービスは、ゲーミングサービスに該当しますか？」等の質問に対し、Yes/Noで回答

あなたのユーザー間サービスについて、以下の質問にYesかNoで答えてください。	
• *私のサービスは、以下のサービスタイプに該当する。* *（当てはまるものすべてを選択）* *a.ソーシャルメディアサービス* *b.メッセージングサービス* *c.ゲーミングサービス*	*Y/N*

違法コンテンツに紐づく、リスク要因の把握

例）「ゲーミングサービス」というリスク要因が、CSEAや憎悪等の違法コンテンツに紐づくことが把握できる

- ***リスク要因***：*ゲーミングサービス*
- ***主な違法コンテンツの種類***：*あなたのサービスは、以下の種類の違法コンテンツのリスクを高く抱えている可能性が高い*
  テロ・CSEA・憎悪・ハラスメント

出所）https://www.ofcom.org.uk/online-safety/illegal-and-harmful-content/statement-protecting-people-from-illegal-harms-online/ よりNRI作成

Step2：特定した違法コンテンツごとに、リスクの影響度と発生可能性の2軸を評価し、リスクレベルを高・中・低・リスクなしに割り当てることが求められている

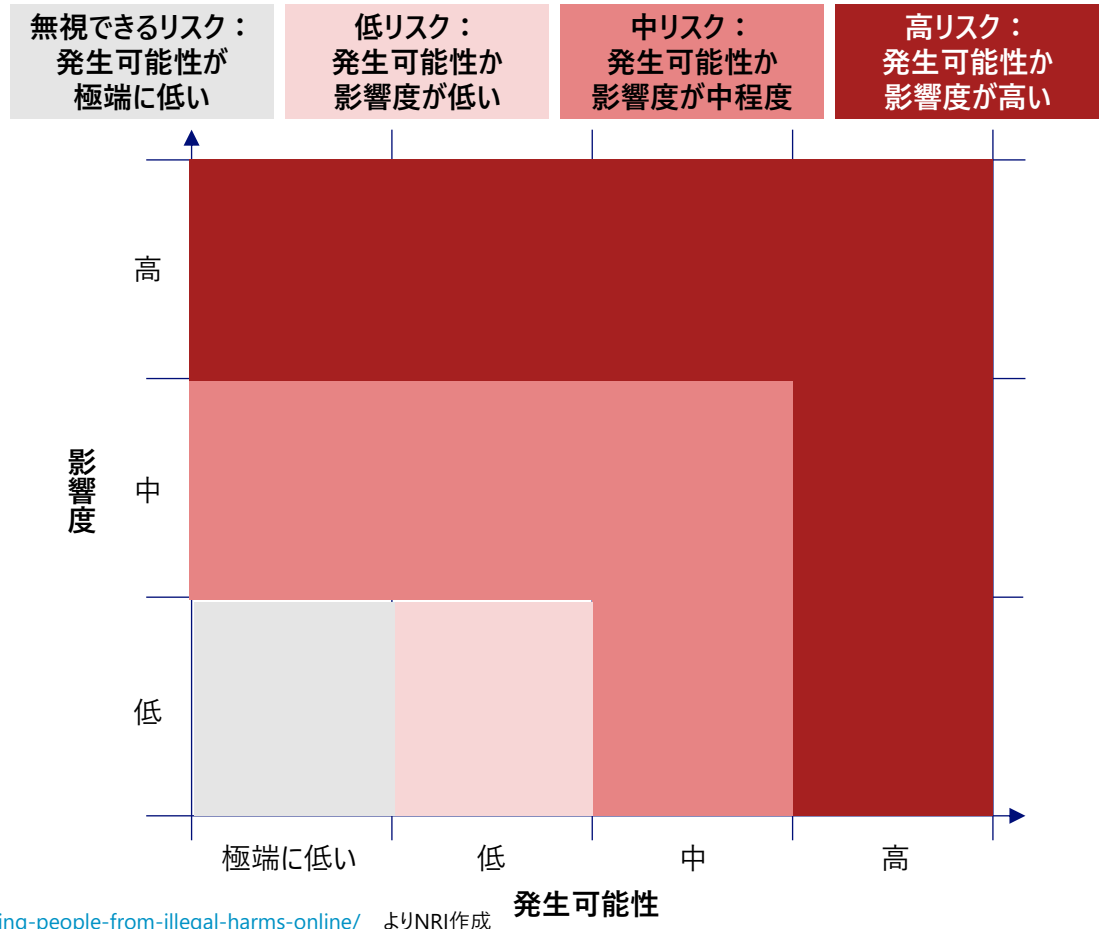
Step2. 危害リスクの評価

リスクの発生可能性と影響度の評価

- サービス提供者は、違法コンテンツの種類ごとに、リスクの**発生可能性と影響度**を評価する。
- その際、以下のような**エビデンス**をそれぞれインプットとして取り入れ、評価を行うべきであるとされている。

類型	エビデンス
コアインプット（Core inputs）：すべてのサービスプロバイダーが考慮すべきエビデンス	• リスクプロファイル（ステップ1）を通じて特定されたリスク要因 • ユーザーからの苦情および報告 • ユーザーデータ（年齢、言語等） • 被害事例の分析 • その他の関連情報（被害のリスクを増大または低減させる可能性のある貴社のサービスのその他の特性を含む）
強化インプット（Enhanced inputs）：大規模サービスプロバイダー等が考慮すべきエビデンス	• 製品テストの結果 • コンテンツモデレーションの成果 • 社内の専門家との協議・独立専門家の見解

リスクレベルの割り当て



出所）https://www.ofcom.org.uk/online-safety/illegal-and-harmful-content/statement-protecting-people-from-illegal-harms-online/ よりNRI作成

安全義務に関する制度｜第10条「違法コンテンツに関する安全義務」

■ **ユーザー間サービスや検索サービスを提供する事業者に対して**「違法コンテンツに関する安全義務」を**義務付け** (10条) 。

措置内容	条文
サービスの設計または運営に関連する以下の措置を講じる義務	10条 2 項
・優先的違法コンテンツに個人が接触することを防止 ・優先的犯罪の実行または支援に使用されるリスクを、最新の違法コンテンツリスク評価に基づき効果的に軽減・管理 ・最新の違法コンテンツリスク評価で特定された個人への害のリスクを効果的に軽減・管理	(a)～(c)
サービスを次のように運営する義務	10条 3 項
・優先的違法コンテンツが存在する期間を最小限に抑えるために適切なシステムおよびプロセスを利用してサービスを運用 ・違法コンテンツの存在を通知された場合、または存在を認識した場合、そのコンテンツを迅速に削除するようサービスを運用	
上記の義務（10条2項・3項）について、以下の分野において措置を組み込む義務	10条 4 項
・リスクの管理 ・機能・アルゴリズム設計 ・利用規約 ・特定コンテンツのアクセス管理 ・コンテンツモデレーション ・ユーザーサポート ・スタッフの方針と実務	(a)～(h)
以下をサービスの利用規約に含め、適用する義務	10条5項、6項
・違法コンテンツから個人を保護する方法	(a)
プロアクティブ技術に関する情報（技術の種類、使用タイミング、機能の仕組みを含む）をサービスの利用規約に含める義務	10条 7 項
利用規約の規定が明確で、アクセスしやすいことを確保する義務	10条 8 項
リスク評価の結果をサービス利用規約に要約し、利用規約に記載する義務(大規模なユーザー間サービス事業者（カテゴリー 1 サービス)のみ義務）	10条 9 項

出所）Online Safety Act 2023よりNRI作成

（参考）10条　違法コンテンツに関する安全義務

1. 本条は、規制対象ユーザー間サービスに適用される違法コンテンツに関する義務を定める。

すべてのサービスに適用される義務

2. サービスの設計または運用に関連して、以下の事項を達成するために適切な措置を講じる義務：
   (a) サービスを介して優先的違法コンテンツに個人が接触することを防止する、
   (b) サービスが優先的犯罪の実行または支援に使用されるリスクを、最新の違法コンテンツリスク評価に基づき効果的に軽減および管理する、
   (c) 最新の違法コンテンツリスク評価で特定された個人への害のリスクを効果的に軽減および管理する（第9条第5項(g)参照）。
3. サービスを次のように運営する義務：
   (a) 優先的違法コンテンツが存在する期間を最小限に抑えるために適切なシステムおよびプロセスを用いる、
   (b) 人から違法コンテンツの存在を通知された場合、またはその他の方法でその存在を認識した場合、そのコンテンツを迅速に削除する。
4. 第2項および第3項に定める義務は、サービスの設計、運用、使用方法、およびサービス上に存在するコンテンツすべてに適用される。この義務には、以下の分野で措置を講じることが含まれる（適切な場合）：
   (a) 規制遵守およびリスク管理の取り組み
   (b) 機能性、アルゴリズム、その他の設計要素
   (c) 利用規約に関する方針
   (d) サービスまたは特定のコンテンツへのアクセスを管理する方針（ユーザーのアクセスをブロックする場合を含む）
   (e) コンテンツのモデレーション（削除を含む）
   (f) ユーザーが遭遇するコンテンツを制御する機能
   (g) ユーザーサポートに関する措置
   (h) スタッフの方針および実務
5. 以下を明記する規定をサービスの利用規約に含める義務：
   (a) 違法コンテンツから個人を保護する方法（第3項各号に対応）、
   (b) 第3項(a)に関しては、テロリズム関連コンテンツ、CSEA（児童性虐待）コンテンツ（第59条および附則6参照）およびその他の優先的違法コンテンツをそれぞれ個別に明記。
6. 第5項で言及される利用規約の規定を一貫して適用する義務。
7. 第2項または第3項の義務に従う目的で使用される積極的技術に関する情報（技術の種類、使用タイミング、機能の仕組みを含む）を提供する規定をサービスの利用規約に含める義務。
8. 第5項および第7項で言及される利用規約の規定が明確であり、アクセスしやすいことを確保する義務。

カテゴリー1サービスに適用される追加義務

9. サービスにおける最新の違法コンテンツリスク評価の結果（リスクのレベルや、個人への潜在的な害の性質と重大性を含む）を要約し、利用規約に記載する義務。

出所）Online Safety Act 2023よりNRI作成

Ofcomは、OSA第41条に基づき安全義務の詳細を記した行動規範を策定

- Ofcomは、PFサービス事業者に課される義務に対する行動規範の公表が義務付けられており（41条）、PF事業者は、当該行動規範を参照しながら、安全義務を履行する。
- 同法に基づきOfcomは、**違法コンテンツに対する安全義務の詳細を記した行動規範(”Illegal content Codes of Practice for user-to-user services”)を公表**（2024年12月）。

ユーザー間サービスを対象とした安全義務（全41項目）

大項目	小項目数
ガバナンスと説明責任 →リスク管理活動の年次レビューなど	7
コンテンツモデレーション →違法コンテンツの疑いがあるものを確認・評価するコンテンツ監視機能の保持など	8
自動化されたコンテンツモデレーション →ハッシュ照合を使用したCSAMの検知・削除など	2
レコメンダシステム →コンテンツレコメンダシステムのオンプラットフォームテスト中の、安全性指標の収集など	1
設定、機能とユーザーサポート →子供ユーザーの安全デフォルトなど	2
利用規約 →使用条件の明記など	3
ユーザーアクセス →禁止された組織のアカウントの削除など	1
ユーザーコントロール →ユーザーブロックとミュートなど	3
報告と苦情 →見つけやすく、アクセスしやすく、使いやすい苦情処理システムとプロセスの保持など	14

出所）https://www.ofcom.org.uk/siteassets/resources/documents/online-safety/information-for-industry/illegal-harms/illegal-content-codes-of-practice-for-user-to-user-services-24-feb.pdf?v=391889

（参考）安全義務に関する行動規範

安全義務の適用は、サービスの規模・リスクの程度により判別される

■ Ofcomはユーザー間サービスを6種類（2×3）に分け、安全義務の適用範囲を整理している。

安全義務の適用の考え方

サービスの規模	• **大規模** • **小規模**	• 大規模：英国の月間ユーザー数が700万人（英国人口のおよそ10%）を超えるサービス • 小規模：上記未満のサービス
✖		
リスクの程度	• **低リスク** • **単一リスク** • **マルチリスク**	• 低リスク：すべての違法な危害について、プロバイダーが低リスクと評価したサービス • 単一リスク：1種類の違法な危害について、プロバイダーが高リスクまたは中リスクと評価したサービス • マルチリスク： 2種類以上の違法な危害について、プロバイダーが高リスクまたは中リスクと評価したサービス

出所）https://www.ofcom.org.uk/siteassets/resources/documents/online-safety/information-for-industry/illegal-harms/illegal-content-codes-of-practice-for-user-to-user-services-24-feb.pdf?v=391889

安全義務に関する行動規範｜ユーザー間サービスにおける行動規範の内容①

適用される　一部適用される

大項目	内容	サービスの規模とリスクの程度					
		小規模サービス			大規模サービス		
		低リスク	単一リスク	マルチリスク	低リスク	単一リスク	マルチリスク
ガバナンスと説明責任	**リスク管理活動の年次レビュー** サービス提供者の最高統治機関は、毎年リスクレビューを実施し、記録しなければならない。						
	違法コンテンツの安全確保義務、報告義務、苦情対応義務を負うポジションの設定 サービス提供者は、違法コンテンツ安全義務および報告・苦情の義務の遵守に責任を負う個人を指名すべきである。						
	責任に関する書面による陳述 サービス提供者は、英国の個人に関連する違法な危害のリスク管理について、意思決定を行う上級管理職の責任を明記した書面を作成する必要がある。						
	内部監査および保証 サービス提供者は、リスク評価で特定された個人に対する危害のリスクを軽減・管理するために講じた対策が継続的に有効であることを独立的に保証するための、内部監視および保証機能を備えるべき。						
	違法な被害の証拠の追跡 サービス提供者は、自社のサービス上で発生する新たな種類の違法コンテンツの証拠や、特定の違法コンテンツまたは違法コンテンツの代替手段の異常な増加を追跡するべき。						
	利用者に対する違法な被害からの保護に関する行動規範の作成 サービス提供者は、英国のユーザーを違法な危害のリスクから保護することに関する基準と期待事項を定めた行動規範（Code of Conduct）を策定する必要がある。						
	サービスのコンプライアンスに対するアプローチに関する訓練 サービス提供者は、自社のサービスの設計や運用管理に関与する従業員が、違法コンテンツの安全義務および報告・苦情対応義務への対応方針について十分な研修を受け、それらの義務を適切に履行できるようにする必要がある。						

安全義務に関する行動規範｜ユーザー間サービスにおける行動規範の内容②

適用される　一部適用される

大項目	内容	サービスの規模とリスクの程度					
		小規模サービス			大規模サービス		
		低リスク	単一リスク	マルチリスク	低リスク	単一リスク	マルチリスク
コンテンツモデレーション	**違法コンテンツの疑いがあるものを確認・評価するコンテンツ監視機能の保持** 違法コンテンツである可能性があると疑われるコンテンツを確認・評価するためのシステムおよびプロセスを設計し、運用する必要がある。						
	違法コンテンツを迅速に削除できるコンテンツ監視機能の保持 違法コンテンツおよび違法コンテンツの代替手段（プロキシ）を迅速に削除するためのシステムおよびプロセスを設計・運用する必要がある。						
	内部コンテンツポリシーの設定 内部コンテンツポリシーを策定し、記録する必要がある。						
	パフォーマンス目標の設定 コンテンツ・モデレーション機能に関するパフォーマンス目標を設定する必要がある。						
	優先順位付け コンテンツの優先的な審査に関するポリシーを策定し、適用する必要がある。						
	コンテンツモデレーションに対するリソース投入 コンテンツ・モデレーション機能に適切なリソースを確保する必要がある。						
	従事する個人（非ボランティア）へのトレーニングおよび資料の提供 コンテンツ・モデレーション業務に従事する従業員（ボランティアを除く）が、適切に業務を遂行できるよう、研修や資料を提供する必要がある。						
	ボランティアへの資料提供 コンテンツ・モデレーション機能に従事するボランティアが、適切に業務を遂行できるよう、研修や資料を提供する必要がある。						
	ハッシュ照合を使用したCSAMの検知・削除*1 ハッシュマッチング技術を使用して児童性的虐待資料（CSAM）を検出し、削除する必要がある。						
	リストされたCSAM URLに一致するコンテンツの検知・削除*2 サービス上で公開されているコンテンツの中から、過去に児童性的虐待資料（CSAM）をホストしていると特定されたURLのリストと一致するものを検出し、削除する必要がある。						

注1）画像ベースのCSAMのリスクが中程度または高い大規模サービスのプロバイダ、または画像ベースのCSAMのリスクが高く、(a) 英国の月間アクティブユーザー数が700,000人を超える、または(b) ファイルストレージおよびファイル共有サービスに適用される。
注2）CSAM URL のリスクが中程度または高い大規模サービスのプロバイダー、または英国の月間アクティブユーザー数が 700,000 人以上で、CSAM URL のリスクが高いサービスに適用される。

安全義務に関する行動規範｜ユーザー間サービスにおける行動規範の内容③

適用される　一部適用される

大項目	内容	サービスの規模とリスクの程度					
		小規模サービス			大規模サービス		
		低リスク	単一リスク	マルチリスク	低リスク	単一リスク	マルチリスク
報告と苦情	**苦情の受付** 英国のユーザー、被害を受けた人が適切な措置を講じてもらえるように、関連する苦情を申し立てることができる苦情処理システムおよびプロセスを備える必要がある。						
	見つけやすく、アクセスしやすく、使いやすい苦情処理システムとプロセスの保持 苦情処理手続きを設計・運用する際に、利用者が簡単に見つけ、アクセスし、使用できるようにする必要がある。						
	苦情の提出前の情報提供*1 特定のコンテンツに関する報告ツールを通じて、苦情の情報を他のユーザーと共有するかどうか、また共有する場合はどの情報が共有されるのかについて、苦情申立人が確認できるようにする必要がある。						
	見込みのタイムラインの送信 苦情を受理したことを申立人に通知し、苦情が解決されるまでのおおよその時間枠を提示する必要がある。						
	苦情の処理方法に関する追加情報の送信*2 苦情の可能な結果について申立人に通知し、サービスがその結果について申立人に報告するかどうかを含めて知らせる必要がある。						
	苦情申し立て後の、連絡のオプトアウト 申立人がサービス提供者からの一時的でない苦情に関する通知を受け取らないよう選択できる、オプションを提供する必要がある。						
	違法コンテンツの疑いに関する関連クレームへの適切な対応 疑わしい違法コンテンツに関する苦情を、コンテンツの優先審査プロセスおよびコンテンツ・モデレーション機能に従って処理するか、これらの優先順位や目標に関する推奨事項が該当しない場合は速やかに対応する必要がある。						

注1）子供がアクセスする可能性が高いサービスを提供する事業者に適用される。　注2）子供がアクセスする可能性が高いサービスを提供する事業者に適用される。
注3）詐欺のリスクが中程度または高い大規模なサービスを提供する事業者に適用される。

安全義務に関する行動規範｜ユーザー間サービスにおける行動規範の内容④

適用される　一部適用される

大項目	内容	サービスの規模とリスクの程度					
		小規模サービス			大規模サービス		
		低リスク	単一リスク	マルチリスク	低リスク	単一リスク	マルチリスク
報告と苦情	**判断時間と正確性の監視** 異議申し立てに該当する苦情を判断し、その判断に要する時間および正確性についてパフォーマンス目標に基づいて監視する必要がある。			適用される	適用される	適用される	適用される
	異議申し立てへの対応 異議申し立てに該当する苦情を速やかに判断する必要がある。	適用される	適用される				
	異議申し立てによって判断を覆した場合の対応 異議申し立てに該当する苦情について、プロバイダーがコンテンツが違法であるとの判断を覆した場合、これまでに講じた措置を取り消す必要がある。	適用される	適用される	適用される	適用される	適用される	適用される
	自動検出技術の使用に関する苦情への対応 利用規約に違反する形でのプロアクティブ技術（自動検出技術）の使用に関する苦情について、プロバイダーは申立人に対し、プロバイダーが講じる可能性のある措置および申立人が訴訟を起こす権利について通知する必要がある。	適用される	適用される	適用される	適用される	適用される	適用される
	担当者の指名 他の関連するすべての苦情を適切な個人またはチームに振り分け、処理を確実に行うために、担当者またはチームを指名する必要がある。	適用される	適用される	適用される	適用される	適用される	適用される
	明確に根拠のない苦情に対する例外 明らかに根拠がないと判断できる情報や属性を示すポリシーを策定している場合、関連する苦情を無視することができる。	適用される	適用される	適用される	適用される	適用される	適用される
	Trusted Flaggerが不正を通報するための専用報告チャネルの設定*3 信頼されたフラッガー（Trusted Flaggers（少なくとも特定の公的機関を含む）が詐欺を報告できる専用の報告チャネルを確立する必要がある。					一部適用される	一部適用される

注1）子供がアクセスする可能性が高いサービスを提供する事業者に適用される。
注2）子供がアクセスする可能性が高いサービスを提供する事業者に適用される。
注3）詐欺のリスクが中程度または高い大規模なサービスを提供する事業者に適用される。

安全義務に関する行動規範｜ユーザー間サービスにおける行動規範の内容⑤

適用される　一部適用される

大項目	内容	サービスの規模とリスクの程度					
		小規模サービス			大規模サービス		
		低リスク	単一リスク	マルチリスク	低リスク	単一リスク	マルチリスク
レコメンダシステム	**コンテンツレコメンドシステムのオンプラットフォームテスト中の、安全性指標の収集*1** プラットフォーム上でコンテンツのレコメンダーシステムをテストする際、設計調整を行う際に追加の安全性指標を収集する必要がある。						
設定、機能とユーザーサポート	**子供ユーザーの安全デフォルト*2** 特定の機能を対象とし、子どもユーザーと他のユーザーとの可視性や交流を制限する安全デフォルト設定を子どもユーザーのアカウントに適用する必要がある。						
	子供ユーザーのサポート*2 子どもユーザーがより適切な判断を下せるように、重要なタイミングで適切なサポート情報を提供する必要があります。また、そのメッセージが目立つように表示され、明確で子どもにも分かりやすいものであることを確保する必要がある。						
利用規約	**軽減措置や苦情の処理に関する記載** 利用規約および声明の中に、個人を違法コンテンツから保護するための対策、使用されるプロアクティブ技術、および苦情の処理・解決方法に関する条項を含める必要がある。						
	リスク評価の結果の記載 直近の違法コンテンツに関するリスク評価の結果を、利用規約および声明の中で要約する必要がある。						
	透明性の確保 利用規約および声明に記載する違法コンテンツからの保護に関する条項が、明確で分かりやすく、誰でもアクセスしやすいものであることを確保する必要がある。						

注1）レコメンダーシステムのオンプラットフォームテストを実施するサービスプロバイダーで、2種類以上の特定の違法な被害のリスクが中程度または高い場合に適用される。
注2）グルーミングのリスクが高いすべてのサービス、またはグルーミングのリスクが中程度で規模の大きいサービスを提供するプロバイダーに適用される。
注2）カテゴリー1サービスの提供者に適用される。

安全義務に関する行動規範｜ユーザー間サービスにおける行動規範の内容⑥

適用される　一部適用される

大項目	内容	サービスの規模とリスクの程度					
		小規模サービス			大規模サービス		
		低リスク	単一リスク	マルチリスク	低リスク	単一リスク	マルチリスク
ユーザーアクセス	**禁止された組織のアカウントの削除** 英国政府によって禁止されたテロ組織が運営している、またはその代理で運営されていると推測できる合理的な根拠がある場合、そのユーザーアカウントをサービスから削除する必要がある。						
ユーザーコントロール	**ユーザーブロックとミュート*1** すべての登録ユーザーに対し、他のユーザーアカウントをブロックおよびミュートするオプションを提供する必要がある。						
	コメントの無効化*2 すべての登録ユーザーに対し、自身のコンテンツへのコメントを無効にするオプションを提供する必要がある。						
	著名ユーザと収益化ラベル*3 ユーザーが「著名なユーザー制度」や「収益化制度」に基づいてプロフィールがどのように、なぜラベル付けされているのかを理解できるよう、情報を提供する必要がある。						

注1）以下のような違法行為の1つまたは複数について、中程度または高リスクである大規模なサービスを提供するプロバイダーに適用される。（a）グルーミング、（b）自殺（または自殺未遂）の奨励または幇助、（c）憎悪、（d）嫌がらせ、ストーキング、脅迫および虐待、（e）支配的または強制的な行動。
注2）大規模なサービスを提供するプロバイダに適用される。
注3）詐欺または外国からの干渉のリスクが中程度または高い大規模なサービスのプロバイダーに適用される。

義務履行の確認方法

ユーザー間サービス提供者等は第77条に基づき、Ofcomに透明性レポートを毎年提出する必要がある。
第78条ではOfcomに対し、透明性レポートに関するガイダンスを作成することを求めている

透明性レポート提出義務（OSA77条）

- Ofcomは、毎年、サービス提供者に対して、透明性レポートの作成を求める通知を送付する必要がある。
- 通知を受けた**サービス提供者は、以下の要件を満たし、透明性レポートを作成しなければならない**。

A) 通知に記載された種類の情報を含めること。
B) 通知で指定されたフォーマットで作成すること。
C) 通知で指定された期日までにOFCOMへ提出すること。
D) 通知で指定された方法および期日までに公開すること。

Ofcomに対する、透明性レポートに関するガイダンス作成義務（78条）

- OFCOMは、以下の事項に関するガイダンスを作成しなければならない

A) OFCOMが透明性レポートに求める情報の決定方法
B) 透明性レポートの情報の活用方法
C) OFCOMが透明性レポートの作成・公開に関して関連があると考えるその他の事項

出所）Online Safety Act 2023よりNRI作成

当局の執行権限

Ofcomは、リスク評価と安全義務を含めた違反の疑義に対する調査を経て、違反仮通知を発出。事業者の意見陳述を経てなお義務違反と判断した場合、確認決定を通知し違反を確定する

OSAにおける違反確定までの流れ

調査開始・情報収集	•Ofcomは、サービス提供者がOSA上の義務に違反している場合に調査を開始。 •調査の中で、Ofcomは情報の要求（第100条）、インタビュー要求（第106条）、立ち入り検査（第107条）等の権限行使が可能。
違反仮通知の発出	•調査の結果、事業者が義務に違反しているとOfcomが判断した場合、違反仮通知（provisional notice of contravention）をサービス提供者に通知。（第130条） •仮通知には、違反の内容や根拠などが記される。
サービス提供者の意見陳述	•Ofcomは違反仮通知の中で、サービス提供者が意見陳述できる期間を明示する必要がある。（第130条） •サービス提供者はこの期間で、Ofcomの指摘に対する説明や訂正などを主張する。
違反の決定	•意見陳述期間が終了し、なお義務違反と判定される場合、Ofcomは確認決定（confirmation decision）をサービス提供者に通知する。（第132条） •確認決定の中で、違反があった旨およびその詳細な理由、サービス提供者に対して科される措置命令や罰金が具体的に指示される。

出所）Online Safety Act 2023よりNRI作成