資料WG12－1－2

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 ワーキンググループ（第12回）

EU・英国におけるリスク評価の概要

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

2024年4月5日

Envision the value,
Empower the change

目次

EU：リスク評価

3. DSAの全体構成（目次）（1/2）

第Ⅰ章 総則	
第1条	主題
第2条	範囲
第3条	定義

第Ⅱ章 仲介サービス提供者の責任	
第4条	「単なる導管」
第5条	「キャッシング」
第6条	ホスティング
第7条	自主的な調査と法令遵守
第8条	一般的なモニタリング及び積極的な事実調査の義務なし
第9条	違法コンテンツに対する措置命令への対応
第10条	情報提供命令への対応

第Ⅲ章 透明で安全なオンライン環境のためのデューデリジェンス義務	
第1節 すべての仲介サービス提供者に適用される規定	
第11条	加盟国当局、欧州委員会および理事会の連絡窓口
第12条	サービス受領者の窓口
第13条	法定代理人
第14条	利用規約
第15条	仲介サービス提供者に対する透明性報告義務

第Ⅲ章 透明で安全なオンライン環境のためのデューデリジェンス義務（つづき）	
第2節 オンライン・プラットフォームを含むホスティング・サービスの提供者に適用される追加規定	
第16条	通知と行動の仕組み
第17条	理由の通知
第18条	刑事犯罪の疑いに関する通知
第3節 オンライン・プラットフォームの提供者に適用される追加規定	
第19条	零細企業及び中小企業の除外
第20条	内部苦情処理体制
第21条	法定外の紛争解決
第22条	信頼できる警告者
第23条	不正使用に対する措置及び保護
第24条	オンライン・プラットフォームの提供者に対する透明性報告義務
第25条	オンライン・インターフェースの設計と構成
第26条	オンライン・プラットフォームにおける広告
第27条	レコメンダー・システムの透明性
第28条	未成年者のオンラインでの保護
第4節 消費者に取引業者との遠隔契約を可能にするオンライン・プラットフォームの提供者に適用される追加規定	
第29条	零細企業及び中小企業の除外
第30条	トレーダーのトレーサビリティ
第31条	コンプライアンス・バイ・デザイン
第32条	情報を通知される権利

第Ⅲ章 透明で安全なオンライン環境のためのデューデリジェンス義務（つづき）	
第5節 システムリスクを管理するための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）の提供者の追加義務	
第33条	超大型オンライン・プラットフォームと超大型オンライン検索エンジン
第34条	**リスク評価**
第35条	**リスクの軽減**
第36条	危機対応メカニズム
第37条	**独立監査**
第38条	レコメンダー・システム
第39条	オンライン広告の透明性の追加
第40条	データへのアクセスと精査
第41条	コンプライアンス・オフィサー
第42条	VLOP・VLOSEに対する透明性報告義務
第43条	監督料
第6節 デューディリジェンス義務に関するその他の規定	
第44条	標準
第45条	行動規範
第46条	オンライン広告の行動規範
第47条	アクセシビリティの行動規範
第48条	危機管理プロトコル

出所）EU-Lex「Document 32022R2065」
https://eur-lex.europa.eu/eli/reg/2022/2065/oj -

3. DSAの全体構成（目次）（2/2）

第Ⅳ章 実施、協力、制裁及び執行	
第1節 所轄官庁と各国デジタルサービス・コーディネーター	
第49条	所轄官庁とデジタル・サービス・コーディネーター
第50条	デジタルサービス・コーディネーターの要件
第51条	デジタル・サービス・コーディネーターの権限
第52条	罰則
第53条	苦情を申し立てる権利
第54条	報酬
第55条	活動報告
第2節 権限、協調調査及び一貫性メカニズム	
第56条	権限
第57条	相互援助
第58条	デジタルサービスコーディネーターの国境を越えた協力
第59条	欧州委員会への照会
第60条	共同調査
第3節 欧州デジタルサービス会議	
第61条	欧州デジタルサービス会議
第62条	会議の構成
第63条	会議の任務

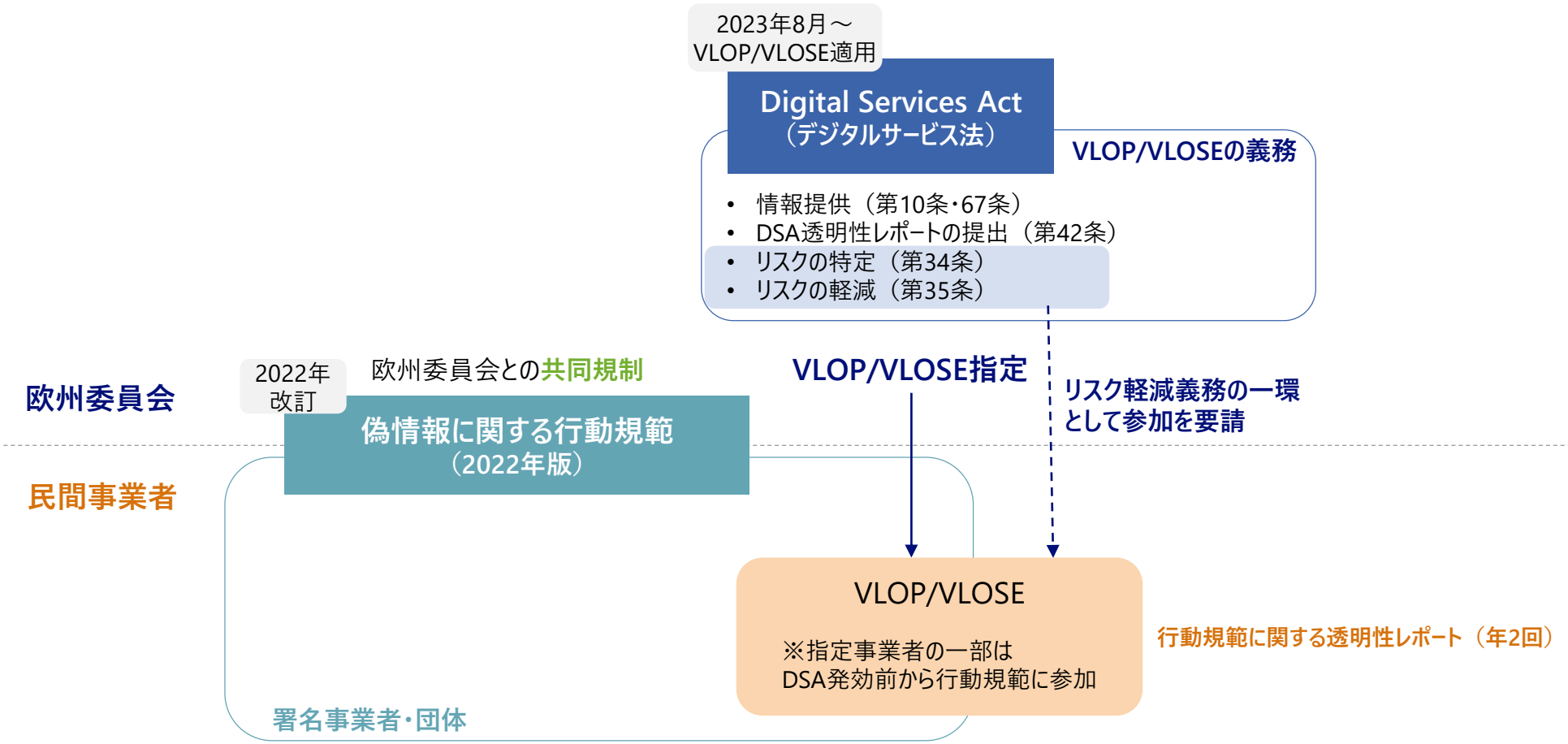
第Ⅳ章 実施、協力、制裁及び執行（つづき）	
第4節 超大規模オンラインプラットフォームおよび超大規模オンライン検索エンジンの提供者に関する監督、調査、遵守及びモニタリング	
第64条	専門知識及び能力の開発
第65条	超大型オンライン・プラットフォームおよび超大型オンライン検索エンジンの提供者の義務の執行
第66条	欧州委員会による手続きの開始と調査への協力
第67条	情報の要求
第68条	聴取・陳述を行う権限
第69条	調査の権限
第70条	暫定措置
第71条	コミットメント
第72条	モニタリング行為
第73条	違反
第74条	制裁金
第75条	第III章第5節に定められた義務違反に対する救済措置の監督強化
第76条	定期的な制裁金の支払い
第77条	制裁金賦課の制限期間
第78条	罰則の執行期限
第79条	聴聞権とファイルへのアクセス権
第80条	決定事項の公表
第81条	欧州連合司法裁判所による審査
第82条	アクセス制限の請求と国内裁判所との協力
第83条	欧州委員会の介入に関する実施法

第Ⅳ章 実施、協力、制裁及び執行（つづき）	
第5節 執行に関する共通規定	
第84条	職業上の秘密
第85条	情報共有システム
第86条	代理
第6節 委任法および実施法	
第87条	委任の発動
第88条	委員会手続き

第Ⅴ章 最終条項	
第89条	指令2000/31/ECの削除
第90条	指令（EU）2020/1828の改正
第91条	評価
第92条	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への適用が予想される
第93条	効力の発生及び適用

出所）EU-Lex「Document 32022R2065」
https://eur-lex.europa.eu/eli/reg/2022/2065/oj -

行動規範への参加は、VLOP/VLOSEのリスク軽減義務の一環に位置づけられる



出所）欧州委員会「Guidance on Strengthening the Code of Practice on Disinformation」https://digital-strategy.ec.europa.eu/en/library/guidance-strengthening-code-practice-disinformation
「Protecting democracy」https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/new-push-european-democracy/protecting-democracy_en

参考：第34条「リスク評価」

- 第34条「リスク評価」は**VLOP・VLOSEに関する追加規定**として明記されている。
- 第34条では、VLOP・VLOSEは、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、またはそのサービスの利用に起因する欧州域内のシステムリスクを真摯に特定・分析・評価をしなければならない、とされている。
  - 第33条第6項第2号で言及されている適用日または**少なくとも1年に1回、いかなる場合においても、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、リスク評価を実施しなければならない。**
  - リスク評価を実施する際、特に、**推奨システムの設計や、コンテンツモデレーションシステム、適用される条件、広告の選択・表示システム、VLOP・VLOSEの慣行に関するデータ等の要素が、システム上のリスクに影響を及ぼすかどうかを考慮する必要**がある。
  - VLOP・VLOSE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のデジタルサービスコーディネーターに伝達しなければならない。

条文（抜粋、仮訳）

第34条
（第1項）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、特に憲章第1条に規定される人間の尊厳、憲章第7条に規定される私生活および家族生活の尊重、憲章第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、憲章第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、憲章第21条に謳われる非差別、憲章第24条に謳われる児童の権利の尊重、憲章第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

出所）EU-Lex「Document 32022R2065」https://eur-lex.europa.eu/eli/reg/2022/2065/oj -

参考：第35条「リスク軽減」

- 第35条「リスク軽減」は**VLOP・VLOSEに関する追加規定**として明記されている。
- 第35条では、VLOP・VLOSEは、第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、当該措置が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、とされている。
  - **当該措置には、レコメンダー・システムや広告システム適合、システムリスク検知への監督強化、第22条・第21条へのコミット、第45条・第48条に基づいた他の提供者との協力、オンライン・インターフェースの適合、年齢認証やペアレンタル・コントロール・ツール、未成年者の支援ツールなど、的を絞った措置等**が含まれる。
  - 欧州デジタルサービス会議は欧州委員会と協力して、年1回、VLOP・VLOSEの「リスク評価」及び「リスク軽減」に関する包括的な報告書を公表するものとされている。

条文（抜粋、仮訳）

第35条

（第1項）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不能化、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、コンテンツ調整プロセスの適応

(d)レコメンダー・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条に従った信頼される旗振り業者との協力、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）EU-Lex「Document 32022R2065」https://eur-lex.europa.eu/eli/reg/2022/2065/oj -

参考：第37条「独立監査」

- 第37条「独立監査」は、**VLOP・VLOSEに関する追加規定**として明記されている。
- 第37条では、VLOP・VLOSEは、自己の費用負担で少なくとも年に1回、DSA第Ⅲ章（第11条～48条）に定める義務と、第45・46条の行動規範及び第48条の自主的な危機プロトコルの遵守状況を評価するために独立監査を受けるものとする、とされている。
  - 独立監査を行う主体の要件は、VLOP・VLOSEから独立・利益相反しないこと、リスク管理等の専門知識を持つこと、客観性・職業倫理を遵守することと定められている。
  - VLOP・VLOSEは監査主体の監査報告書の作成を保証する必要があり、報告書の中で監査主体は、「肯定的」「コメント付き肯定的」「否定的」の三段階の監査意見を提示する。「肯定的」以外の意見の場合、VLOP・VLOSEは監査内容を踏まえた取組報告書を作成しなければならない。

条文（抜粋、仮訳）

第37条 独立監査

（第1項） 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は、自己の費用負担で、少なくとも年に1回、以下の遵守状況を評価するための独立監査を受けるものとする：

(a)第III章に定める義務

(b)第45条および第46条の行動規範ならびに第48条の危機プロトコルに従って実施されるすべての約束

（第3項）3第1項に基づき実施される監査は、次の各号に掲げる組織によって実施されなければならない

(a)当該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のプロバイダおよび当該プロバイダに関係する法人から独立しており、かつ、当該プロバイダと利害関係を有しないこと

・・・（略）・・・

(b)リスク管理、技術的能力、能力の分野で実証された専門知識を有すること

(c)客観性と職業倫理が証明されており、特に業務規範または適切な基準の遵守に基づくこと。

（第4項） 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は、監査を実施する組織が監査ごとに監査報告書を作成することを保証しなければならない。その報告書は、文書で立証され、少なくとも以下を含むものとする；

・・・（略）・・・

(g)監査の対象となった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のプロバイダが、第1項で言及された義務およびコミットメントを遵守したかどうかについての監査意見、すなわち「肯定的」、「コメント付き肯定的」または「否定的」のいずれか；(h)監査意見が「肯定的」でない場合、遵守を達成するための具体的な措置に関する業務上の勧告、および遵守を達成するために推奨される期間。

（第6項）「肯定的」でない監査報告書を受領した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のプロバイダは、それらを実施するために必要な措置を講じることを目的として、それらに宛てられた運営上の勧告を十分に考慮しなければならない。勧告を受けてから1ヶ月以内に、それらの措置を記載した監査実施報告書を採択しなければならない。業務上の勧告を実施しない場合、監査実施報告書において、実施しない理由を正当化し、また、指摘されたコンプライアンス違反の事例に対処するためにとった代替措置を明記しなければならない。

出所）EU-Lex「Document 32022R2065」https://eur-lex.europa.eu/eli/reg/2022/2065/oj -

DSAにおけるリスク評価の枠組み

署名事業者・団体のうち、DSAにおいてVLOP・VLOSEに指定される事業者は、リスク評価を含めたDSA、行動規範それぞれの遵守状況について独立機関から監査を受ける必要がある。

- 署名事業者・団体のうち、VLOP・VLOSEに指定されている事業者は、DSA第37条、行動規範コミットメント44に基づいて、独立した監査機関から行動規範の遵守状況について監査を受けなければならない。
  - 監査主体の要件は、VLOP・VLOSEから独立・利益相反しないこと、リスク管理等の専門知識を持つこと、客観性・職業倫理を遵守することと定められている。
- 欧州委員会はDSA第87条に従い、監査の手順や方法及び報告テンプレートを定める委任法の採択権限を持つ。

独立監査の流れ

VLOP・VLOSEに該当する署名事業者・団体

① システミックリスクの識別・分析・評価（DSA第34条）

- ✓ 評価は年1回以上行う必要がある
- ✓ 具体例は以下　（※システミックリスク評価に含む必要あり）
  - 市民言説・選挙等への悪影響リスク
  - 基本権に対する悪影響リスク
  - 違法コンテンツの拡散リスク
  - 人の心身の幸福へのリスク　　等

② 合理的・比例的かつ有効な軽減措置（DSA第35条）

- ✓ ①の評価内容を踏まえて措置を行う。措置の具体例は以下
  - サービス設計・機能等の工夫
  - 利用規約の工夫
  - コンテンツモデレーション手続の工夫
  - 軽減措置を講じる約束を定めた**行動規範**（※）の策定
  - アルゴリズム、広告表示の工夫　　等

※ 欧州委員会等が作成を奨励・促進。特に、偽情報リスクの場合、事業者が作成した行動規範の支持・遵守が「適切な軽減措置と解され得る」とされる。

④ 監査を踏まえた取組報告書の作成（DSA第37条）

- ✓ コンプライアンス達成のための推奨事項を実施する場合は、具体的な措置を記載する必要がある。
- ✓ 実施しない場合は、実施しない理由と代替措置を記載する必要がある。

遵守状況を監査

独立監査主体

監査主体の要件
①VLOP/VLOSEと独立・利益相反しない
②リスク管理等の専門知識を持つ
③客観性・職業倫理の遵守

③ 監査の実施（DSA第37条、行動規範コミットメント44）

- ✓ 監査対象は以下
  - 「DSA上の義務」の遵守状況
  - **「行動規範を通じて自主的に誓約した事項」の遵守状況**
- ✓ 監査意見は次の三段階で、「肯定的」以外の意見の場合はVLOP・VLOSEに対して報告書の作成が求められる
  - **「肯定的」**
  - **「コメント付き肯定的」**
  - **「否定的」**

「肯定的」以外の意見の場合

※24年8月に最初の独立監査の報告の見込み
（24年3月時点ではリスク評価に関する資料は公表されていない）

VLOP/VLOSEは第34条で偽情報の拡散を含むリスクの特定が義務付けられている

■ 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プロバイダーは、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、当組合における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、そのサービスに特化し、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を広めること；
  - (b)基本的権利、特に憲章第1条に謳われる人間の尊厳、憲章第7条に謳われる私生活および家族生活の尊重、憲章第8条に謳わ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、憲章第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、憲章第21条に謳われる非差別、憲章第24条に謳われる児童の権利の尊重、憲章第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、本人の身体的・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、リスク評価を実施する際、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)レコメンダー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；
- 評価はまた、第1項におけるスクが、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年間は、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デジタルサービスコーディネーターに伝達しなければならない。

VLOP/VLOSEは第35条でリスクの軽減を義務付けられている

■ 前文第86項

- 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、基本的権利を遵守しつつ、リスク評価で特定されたシステミックリスクを真摯に軽減するために必要な手段を展開すべきである。採用される措置は、本規則のデューデリジェンス要件を尊重し、特定された特定のシステミックリスクを軽減する上で合理的かつ効果的でなければならない。これらの措置は、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のプロバイダーの経済的能力、および基本的権利に対する潜在的な悪影響を十分に考慮し、そのサービスの利用に対する不必要な制限を回避する必要性に照らして、相応のものでなければならない。これらのプロバイダーは、表現の自由への影響を特に考慮すべきである。

■ 第35条 リスクの軽減

- 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プロバイダーは、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軽減措置を、当該措置が基本的権利に与える影響を特に考慮して、講じなければならない。かかる措置には、該当する場合、以下が含まれる：
  - (a)オンライン・インターフェースを含む、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)利用規約およびその実施方法を変更すること；
  - (c)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および質を含む、コンテンツ調整プロセスの適合。また、特に違法なヘイトスピーチやサイバー暴力に関して、適切な場合には、通知されたコンテンツの迅速な削除、またはアクセス不能化、ならびにコンテンツ調整のための関連する意思決定プロセスおよび専用リソースの適合を含む、コンテンツ調整プロセスの適合を行うこと；
  - (d) レコメンダー・システムを含むアルゴリズム・システムをテストし、適合させること；
  - (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又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条に従った信頼できる旗振り業者との協力、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) サービスの受け手に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) 適切な場合には、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、児童の権利を保護するための的を絞った措置を講じること；
  - (k) 生成または加工された画像、音声、映像であるか否かを問わず、実在する人物、物、場所、その他の実体または出来事に著しく類似し、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は、オンライン・インターフェースに表示される際、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表示できるような使いやすい機能を提供すること。
- ２．（略）

DSAと行動規範の関係性（1/2）

欧州委員会は行動規範の策定と参加を奨励しており、行動規範の遵守はVLOP/VLOSEのリスク軽減義務の一環に位置付けられる。不参加はDSAの義務違反の考慮要素となりうる

■ 前文第103項

- 欧州委員会および理事会は、本規則の適用に資するため、自主的な**行動規範の策定と、それらの規範の規定の実施を奨励すべき**である。欧州委員会および理事会は、行動規範が、取り組んでいる公益目的の性質を明確に定義し、その目的の達成を独立的に評価する仕組みを含む、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。特に、安全保障、プライバシー、個人情報の保護への悪影響の回避や、一般的な監視義務を課すことの禁止に注意を払うべきである。**行動規範**の実施は測定可能であり、公的な監視の対象となるべきであるが、そのような規範の自発的な性質や、利害関係者が参加するかどうかを決定する自由を損なうことがあってはならない。特定の状況においては、**超大規模オンラインプラットフォームが特定の行動規範の策定に協力し、遵守することが重要**である。本規則のいかなる規定も、他のサービスプロバイダーが同じ行動規範に参加することにより、デューデリジェンスの同じ基準を遵守し、ベストプラクティスを採用し、欧州委員会および理事会が提供するガイドラインの恩恵を受けることを妨げるものではない。

■ 前文第104項

- 本規則は、そのような行動規範のために考慮すべき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、情報操作や虐待行為、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についても検討すべきである。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、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による所定の**行動規範の遵守とコンプライアンスは、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欧州委員会による招へい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に違反したか否かを判断する際に、関連のある範囲で考慮されうる**。

■ 前文第106項

- 本規則に基づく行動規範（Codes of conduct）におけるルールは、「製品安全に関する誓約」、「インターネット上の偽造品販売に関する覚書」、「オンライン上の違法なヘイトスピーチ対策に関する行動規範」ならびに「**偽情報に関する行動規範**」など、欧州連合レベルですでに確立されている自主規制の取り組みの基礎となりうる。**特に後者（偽情報に関する行動規範）については、欧州委員会のガイダンスに従い、欧州民主主義計画で発表されたとおり、偽情報に関する行動規範が強化された。**

出所）https://Aigital-strategy.ec.europa.eu/en/policies/code-practice-disinformation#:~:text=The%20Code%20will%20strengthen%20the,challenges%20related%20to%20such%20techniques

DSAと行動規範の関係性（2/2）

欧州委員会は行動規範の策定と参加を奨励しており、行動規範の遵守はVLOP/VLOSEのリスク軽減義務の一環に位置付けられる。不参加はDSAの義務違反の考慮要素となりうる

■ 第45条 行動規範

- 1. 欧州委員会および理事会は、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. 第34条第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。
- 4. 委員会および理事会は、行動規範が第1項および第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。両委員会は、その結論を公表しなければならない。委員会および理事会はまた、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、委員会および理事会は、**行動規範の署名事業者・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）https://Aigital-strategy.ec.europa.eu/en/policies/code-practice-disinformation#:~:text=The%20Code%20will%20strengthen%20the,challenges%20related%20to%20such%20techniques

DSAと行動規範の関係性―（参考）オンライン広告の行動規範に関連する条項（1/2）

行動規範はVLOP/VLOSEのオンライン広告に関する透明性義務を補完するものであり、欧州委員会はその策定と参加を奨励する

■ 前文第88項

- 超大規模オンラインプラットフォームのプロバイダーや、超大規模オンライン検索エンジンのプロバイダーも、レコメンダーシステムをはじめとするアルゴリズムシステムをテストし、必要に応じて適応させるための措置を講じることに努めるべきである。パーソナライズされたレコメンデーションの悪影響を緩和し、レコメンデーションに使用される基準を修正する必要があるかもしれない。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プロバイダーが使用する広告システムも、システミックリスクの誘因となりうる。これらのプロバイダーは、特定の情報に対する広告収入を中止するなどの是正措置、または権威ある情報源の可視性を向上させる、広告システムをより構造的に適合させるなどの他の措置を検討すべきである。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、特にシステミックリスクの検出に関して、その活動の内部プロセス又は監督を強化し、新たな機能に関連するリスク評価をより頻繁に又は的を絞って実施する必要があるかもしれない。特に、異な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間でリスクが共有される場合、**既存の行動規範またはその他の自主規制措置を開始または参加**することを含め、他のサービスプロバイダーと協力すべきである。また、特に**偽情報**キャンペーンに関連するリスクについては、啓発活動を検討すべきである。

■ 前文第107項

- オンライン広告の提供には、一般に、広告のパブリッシャーと広告主をつなぐ仲介サービスを含む複数の関係者が関与する。**行動規範**は、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**広告に関する透明性義務を支援し、補完するものでなければならない**。これは、特に関連情報の伝達の様式に関して、これらの義務の遵守を促進し、強化するための柔軟かつ効果的なメカニズムを提供するためである。これには、広告の代金を支払う広告主が、オンラインプラットフォームのオンラインインターフェース上で広告を提示する自然人または法人と異なる場合に、広告主に関する情報の伝達を容易にすることを含むべきである。**行動規範**には、データの収益化に関する有意義な情報がバリューチェーン全体で適切に共有されることを確保するための措置も含まれるべきである。幅広い利害関係者が関与することで、**行動規範**が広く支持され、技術的に健全で、効果的であり、透明性の義務がその目的を達成するために最高レベルの使いやすさを提供することが保証されるべきである。幅広い利害関係者が関与することで、**行動規範**が広く支持され、技術的に健全で、実効性があり、透明性義務がその目的を達成するよう確保するために最高レベルの使いやすさを提供できるはずである。**行動規範**の実効性を確保するため、**欧州委員会は行動規範の策定に評価メカニズムを含める**べきである。必要に応じて、欧州委員会は、欧州基本権機関または欧州データ保護監督機関に、それぞれの行動規範について意見を述べるよう求めることができる。

行動規範はVLOP/VLOSEのオンライン広告に関する透明性義務を補完するものであり、欧州委員会はその策定と参加を奨励する

■ 第46条 オンライン広告における行動規範

- 1.欧州委員会は、第26条（オンラインプラットフォームにおける広告）および第39条（オンライン広告の追加的な透明性）の要件を超えて、オンライン広告のバリューチェーンにおける関係者の透明性を高めることに貢献するため、オンラインプラットフォームのプロバイダー、およびオンライン広告仲介サービスのプロバイダー、プログラム広告のバリューチェーンに関与するその他の関係者、またはサービスの受け手を代表する組織、市民社会組織もしくは関係当局などのその他の関係するサービスプロバイダーによる、**欧州連合レベルでの自主的な行動規範の策定を奨励し、促進**するものとする。
- 2.欧州委員会は、行動規範が、EU法および国内法、特に競争法およびプライバシーと個人情報の保護に関する法律に従って、オンライン広告における競争的で透明かつ公正な環境と同様に、すべての関係者の権利と利益を十分に尊重した効果的な情報伝達を追求することを確保することを目指すものとする。欧州委員会は、行動規範が少なくとも以下の事項に対処していることを確認することを目指すものとする：
  - (a)第26条第1項(b)、(c)および(d)に定める要件に関して、オンライン広告仲介業者のプロバイダーがサービスの受け手に対して保有する情報の伝達；
  - (b)第39条に基づき、オンライン広告仲介事業者が保有する情報をリポジトリに送信すること；
  - (c)データの収益化に関する有意義な情報；
- 3.欧州委員会は、2025年2月18日までの当該**行動規範の策定**と、2025年8月18日までのその**適用を奨励**しなければならない。
- 4.欧州委員会は、第1項で言及したオンライン広告のバリューチェーンのすべての関係者に対し、行動規範に記載されたコミットメントに賛同し、それを遵守するよう奨励しなければならない。

行動規範（The 2022 Code of Practice on Disinformation）

行動規範の透明性レポートにおける主要PF事業者のコミットメント内容（2024年3月現在）

■ 2024年3月26日に公表された、偽情報に関する行動規範の最新版の透明性レポート（対象期間：2023年7月1日～12月31日）における主要PF事業者の「Executive Summary」の記載内容について、分野ごとにみられる特徴を整理。

- 本レポートでは、特に「4.サービスの完全性」や「5.ユーザーのエンパワーメント」の分野における取組として、2024年度に世界中で行われる選挙を取り巻くAIの使用について、そのリスクと対抗策に関する記載が目立った。

分野	主要PF事業者の主なレポート記載内容
2.広告表示の精査	• 広告表示におけるポリシーを作成・更新し、ポリシーに違反する広告に対して措置を講じている。
3.政治広告への対応	• Googleはポリシーを更新し、選挙広告への生成AI使用についてラベル付けすることを義務化した。 • TikTokは、政府関係者の広告掲載や収益化機能の利用をポリシー内で許可していない。
4.サービスの完全性	• 選挙に先立って、生成AIコンテンツが与えうるリスクに対抗したアプローチ計画やガイダンスの発表、原則への誓約が行われ、新たな脅威の発見や開示の動きが加速している。 • Google検索では、「About This Image」機能を展開し、Metaは、C2PAの認証情報を使用したContent Credentials as a Serviceによって、生成AIコンテンツの開示を図っている。
5.ユーザーのエンパワーメント	• 特に選挙に関連した偽誤情報に対抗するため、音声・画像・動画の出所や作成元などの情報を開示することでユーザーに対する透明性を高める動きが加速している。 • Microsoftは、Bingにおいて欧州選挙に特化した公式投票情報へのリンクを表示する。 • TikTokは、選挙の完全性確保のためのメディアリテラシーキャンペーンを全EU加盟国で実施予定である。
6.研究団体のエンパワーメント	• 研究団体や大学機関等への助成によって研究団体を支援している。 • また、各社リサーチアクセスプログラムや、APIツールによって研究者による情報のアクセスを許可している。
7.ファクトチェック団体のエンパワーメント	• ファクトチェック団体への助成やサミットの開催によって活動を支援している。 • 複数のファクトチェック団体と提携を結ぶことで複数言語をカバーしている。 • Metaは、自社ファクトチェックプログラムにEFCSN認証の受け入れを開始予定。
8.透明性レポート 9.常設タスクフォース 10.行動規範のモニタリング	• 新たな脅威に先立って、タスクフォース内に新たに設置された選挙に関するワーキンググループや生成AIワーキンググループに対し、主要PF事業者は協力の姿勢を示している。

出所）「Transparency Center」各団体レポート　https://disinfocode.eu/reports-archive/?years=2023

The 2022 Code of Practice on Disinformation

主要5PF事業者の分野ごとのコミットメント内容（2024年3月現在）（1/5）

■ 2024年3月26日に公表された最新版の透明性レポート（対象期間：2023年7月1日～12月31日）における「Executive Summary」の記載内容から一部抜粋し、行動規範の10分野別に分類した。「Executive Summary」の記載内容や粒度は各社によって異なる。

- 本レポートより新たに追加された内容のみ抜粋し、記載。
- なお、X（旧Twitter）は2023年5月に脱退しており、新たなレポートを出していないため、除外。

団体名	2.広告表示の精査 （SCRUTINY OF AD PLACEMENTS）	3.政治広告への対応 （POLITICAL ADVERTISING）
Google	• 2023 年 11 月、Advertising は Political Content Policy を更新し、EU 域内のすべての検証済み選挙広告主に対し、選挙広告に合成コンテンツが含まれている場合、その合成コンテンツが本物ではない、またはリアルに見える人物や出来事を描写している場合は、目立つように開示することを義務付けた。（政治広告への対応でも記載。） • 2023年8月、Google広告ははヘルプセンターに訴求情報を追加し、新しい透明性センターを展開。（政治広告への対応でも記載）	• 2023 年 11 月、Advertising は Political Content Policy を更新し、EU 域内のすべての検証済み選挙広告主に対し、選挙広告に合成コンテンツが含まれている場合、その合成コンテンツが本物ではない、またはリアルに見える人物や出来事を描写している場合は、目立つように開示することを義務付けた。（広告表示の精査でも記載） • 2023年8月、Google広告ははヘルプセンターに訴求情報を追加し、新しい透明性センターを展開。（広告表示の精査でも記載）
Microsoft	• Microsoft広告は、広告コンテンツを提供する広告主、および広告を自社のサービスに表示するBingなどのパブリッシャーの両方と連携し、広告による偽情報の拡散を防止するため、これら2種類のビジネス パートナーのそれぞれについて、個別のポリシーおよび実施手段を採用している。	• （前回レポートから追加なし）
Meta	• 2023年7月1日から同年12月31日までに、EU加盟国のFacebookとInstagramから440万件以上の広告を削除。そのうち、4.4万件以上の広告が誤情報ポリシー違反としてFacebookとInstagramから削除された。 • 同期間にいて、EU加盟国においてFacebookとInstagramの両方で70万件以上の広告に「paid for by」の免責事項を表示した。	• （記載なし）
TikTok	• 広告出稿の精査の効果を高めるため、Global Alliance for Responsible Media (GARM)のメンバーとして、GARMフレームワークを支持し、有害な誤情報の収益化排除にコミットしている。 • コンテンツ収益化機能の参加者は、当社の誠実性と真正性のポリシーを含むコミュニティガイドラインを遵守する必要がある。当のポリシーは、サービスの完全性やユーザーの信頼性を損なう可能性のある行為を禁止することを明確にしている。また、重大な誤解を招く情報を含むコンテンツやクリエイターのアカウント、欺瞞的な行為の削除も明確にしている。	• 政府、政治家、政党アカウント（GPPPA）の広告掲載または収益化機能やキャンペーン資金調達にアクセスすることを禁止している。 • 政治的信条を共有することは許可されているが、当社のポリシーでは、ユーザーがこのようなコンテンツの広告や宣伝にお金を払うことは禁止されている。 • 政治広告の透明性に関するEU規則が最近採択されたことを受け、当社の方針への潜在的な影響を慎重に検討する予定である。

出所）「Transparency Center」各団体レポート　https://disinfocode.eu/reports-archive/?years=2023

The 2022 Code of Practice on Disinformation

主要5PF事業者の分野ごとのコミットメント内容（2024年3月現在） （2/5）

団体名	4.サービスの完全性 （INTEGRITY OF SERVICES）
Google	• Google検索では、スパム検出システムの適用範囲を改善するアップデートを公開。また、2023年10月、「About This Image」機能を全世界に展開し、オンライン上の画像やメタデータの信憑性を確認可能にした。（ユーザーのエンパワーメントにも記載）さらに、生成AIコンテンツに関するガイダンスを発表し、検索における情報の品質とコンテンツ全体の有用性を高水準で維持するためのアプローチを概説した。誤情報への対応を支援するため、透かし、メタデータ、その他の技術における新たなイノベーションを最新の生成モデルに近々統合する予定も公表。 • YouTubeでは、2023年10月、YouTubeは責任あるAIイノベーションへのアプローチを発表し、クリエイターがAIツールを使用して改変コンテンツや合成コンテンツを作成した場合に、それを開示することを求める計画を公表。2023年7月、YouTubeを含むGoogleは、ホワイトハウスで他の大手AI企業とともに、AIシステムの脆弱性の発見を促進することを約束し、信頼できるオンライン情報を促進するための8つの約束に署名。 • ルーマニア、スペイン、ベルギーでの新たなパートナーとの契約を含め、Google は EU 域内で 23 の誤報優先フラガーと合意。2023年10月、GoogleはVulnerability Rewards Program（VRP）を拡張し、生成AIに特化した攻撃シナリオに対して報酬を付与。 • 2018年にAI原則を発表して以来新システムの倫理的なレビューを実施し、この度実行に移すためのガバナンスチームを設立。 • 大手AI企業とともに、G7、OECD、各国政府の取り組みを支援するAI開発における責任ある実践を推進することを共同で約束。今後も、Googleが開発したAIツールの報告と拡張を続け、AIの利益を最大化し、リスクを最小化するために、大胆かつ責任あるAIの推進に取り組む。
Microsoft	• 自社サービスが偽情報の助長に使用されるリスクを軽減するため、高精度ツールとテクノロジーを活用し、より広範な社会と透明性をもって情報を共有することでAI研究のリーダーとしての役割を果たす。その一環として、AIコミュニティ、規制当局、社会の利益のため、脅威検出の取り組みを積極的に公表することを約束し、欺瞞的なAIの使用に対抗するため、6つの重点分野を採択。 • 2024年2月のミュンヘン・セキュリティ会議において、テックセクターと協力し、2024年の選挙におけるAIの欺瞞的使用に対抗するための新しい技術協定を発表した。政治家候補者、選挙関係者、その他の重要な利害関係者の外見、声、行動を改変する動画、音声、画像や、投票方法に関する虚偽の情報に対抗する。 • Tech Accordのコミットメントでは、悪質な行為者のディープフェイク作成をより困難にすると同時に、ユーザーが本物のコンテンツを識別するプロセスを簡素化する。AIコンテンツの生成企業だけでなく、流通企業にも焦点を当て、悪用を防ぐためのリスク評価と管理強化によって、AIサービスの安全性強化を目指す。そのため、Tech Accordのコミットメントを満たすため、コンテンツの出所、報告ルートの確立、検知能力の向上などを実施し、新たな選挙保護にコミットメントする。（ユーザーのエンパワーメントにも記載） • C2PAの認証情報を使用し、政治家候補者がメディアにデジタル署名し認証できるようにするため、Content Credentials as a Serviceを開始。 • BingのCopilotは、検索における長年の安全システムに基づいて構築されており、AIに関する新たなリスクへの追加的な保護によって補完される。当社のResponsible AIチームと緊密に連携し新たなAIリスクに積極的に対処しており、Copilot in Bingにおけるアプローチについて透明性を保っている 。（ユーザーのエンパワーメントにも記載）
Meta	• **生成AI**： 2023年9月に新しい生成AI機能を立ち上げ、内外部の専門家と協力し、人々の安全と責任ある構築を優先した。安全性と責任に関するガイドラインをモデルに教え、全Metaアプリ内ですべての年齢層にとって有害または不適切な反応を共有する可能性を低減している。また、生成AIシステムにおける潜在的なバイアスにも対処している。他のAIモデルと同様に、より多くのフィードバックを共有することで、今後のアプローチの洗練が期待される。 • **生成AI画像へのラベリングによる透明性向上**： コンテンツの一部がAIによって作成されたことを示す共通の技術標準について業界のパートナーと協力しており、ユーザーがFacebookやInstagramに投稿するAI生成画像へのラベル付けが可能になる。2024年2月、このアプローチを世界中で重要な選挙が行われる1年を通して行うことを発表。（ユーザーのエンパワーメントにも記載）
TikTok	• 当社の合成メディア・ポリシーは、生成AIコンテンツの使用に対処する。AIによって生成・操作された場合の積極的な開示をユーザーに求める。この要件の遵守を促進するため、開示のためのAI生成ラベルを開始した。同様に、ユーザーが生成AIコンテンツを識別しやすくするため、AIを使用したTikTokエフェクトには、エフェクト名と対応するエフェクトラベルに「AI」の記載が明示されるようになった。 • 秘密裏に行われる影響力工作（CIO）と闘い続け、アカウントの身元、出所、活動場所、人気、目的について当社PFのシステムやコミュニティを欺きながら世論を動かそうとする試みを禁止する。今後数ヶ月のうちに、透明性、説明責任、業界横断的な共有をさらに高めるため、CIO専用のレポートを導入する予定。

出所）「Transparency Center」各団体レポート https://disinfocode.eu/reports-archive/?years=2023

The 2022 Code of Practice on Disinformation

主要5PF事業者の分野ごとのコミットメント内容（2024年3月現在） （3/5）

団体名	5.ユーザーのエンパワーメント（EMPOWERING USERS）
Google	• Google検索は、2023年10月、「About This Image」機能を全世界のユーザーに展開し、オンライン上の画像や画像のメタデータの信憑性や文脈を簡単に確認可能にした。（サービスの完全性にも記載） • YouTubeは、視聴者がオリジナルチャンネルとファンチャンネルをより簡単に見分けられるよう、チャンネルなりすましポリシーを更新。また、有料コンテンツの透明性を更新し、ユーザーがYouTube上で宣言された有料プロダクトプレースメント、スポンサーシップ、エンドースメントを含む動画を個別に閲覧できるようにした。 • ポーランドの非営利団体School with Classに350万ドルのGoogle.org助成金、ベルギーの非営利団体Bibliothèques Sans Frontières (BSF)に100万ドルの助成金を提供し、デジタルシチズンシッププログラムの支援を継続。 • ジャーナリストが偽情報と闘うためのツールを作成するイタリアを拠点とするLUISS DatalabとFederazione Italiana Editori Giornali（FIEG）に100万ユーロを供与。 • 2023年7月から12月にかけて、Googleは7つの誤報関連イベントを開催し、700人以上が参加した。 • 2023年にGoogleの透明性センターを開設し、Googleのポリシーや実施状況について、ユーザーに詳しく開示をした。
Microsoft	• Tech Accordのコミットメントでは、悪質な行為者のディープフェイク作成をより困難にすると同時に、ユーザーが本物のコンテンツを識別するプロセスを簡素化する。AIコンテンツの生成企業だけでなく、流通企業にも焦点を当て、悪用を防ぐためのリスク評価と管理強化によって、AIサービスの安全性強化を目指す。そのため、Tech Accordのコミットメントを満たすため、コンテンツの出所、報告ルートの確立、検知能力の向上などを実施し、新たな選挙保護にコミットメントする。（サービスの完全性にも記載） • C2PAの認証情報を使って、政治家候補者がメディアにデジタル署名し認証できるようにするため、Content Credentials as a Serviceを開始。（サービスの完全性にも記載） • BingのCopilotは、検索における長年の安全システムに基づいて構築されており、AIに関する新たなリスクへの追加的な保護によって補完される。当社のResponsible AIチームと緊密に連携し新たなAIリスクに積極的に対処しており、Copilot in Bingにおけるアプローチについて透明性を保っている 。（サービスの完全性にも記載） • Bingは、選挙の完全性への取り組みとして、欧州選挙に特化した回答や情報パネルを統合し、公式投票情報へのリンクも提供する。 • 情報リテラシー・プログラムへの投資を継続し、ニュースリテラシー・プロジェクトや、Trust Indicators and Verifiedに関するTrust Projectが主導するコラボレーションなど、主要なニュースおよびメディアリテラシーの非営利団体と提携し、業界研究とベストプラクティスに基づいたキャンペーンを展開。 • Tech Accordの重点分野とコミットメントの一環として、AIリテラシーの取り組みを強化し、出所やその他の信頼性指標について理解を深めるためのパートナーシップを拡大。 • 生成AIの機能についてユーザーを教育し、責任あるAIに対する当社のアプローチに関する情報をより広範な一般市民や研究コミュニティに提供する。
Meta	• **生成AI画像へのラベリングによる透明性向上**： コンテンツの一部がAIによって作成されたことを示す共通の技術標準について業界のパートナーと協力しており、ユーザーがFacebookやInstagramに投稿するAI生成画像へのラベル付けが可能になる。2024年2月、このアプローチを世界中で重要な選挙が行われる1年を通して行うことを発表。（サービスの完全性にも記載） • **メディア・リテラシー**：2023年夏、リトアニア、スロバキア、ブルガリアでメディア・リテラシー・キャンペーンを実施し、人々の誤情報識別・対応を支援。また、ポーランドの現地パートナーによる広告キャンペーンを支援し、選挙期間中の誤報に対するレジリエンス向上支援を実施。2024年には、EFCSN及びEU障害フォーラムと協力して、EU選挙に焦点を当てたキャンペーンを実施予定。 • 2023年第3四半期に1つのネットワークを削除し、2023年第4四半期に1つのネットワークを、1つ以上の欧州諸国を対象としたCIB（Coordinated Inauthentic Behaviour）ポリシーに違反したとして削除した。また、偽アカウント削除のための措置を講じており、第3四半期には8億2,700万、2023年第4四半期には全世界で6億9,100万の偽アカウントに措置を講じた。
TikTok	• 気候変動、ウクライナ戦争、イスラエルとハマスの紛争などへの介入を開発すべきかを常に検討しており、これらのツールを23の公用語で提供している。 • 地方選挙に関して、ファクトチェック・パートナーとの緊密な協力し、様々なメディア・リテラシー・キャンペーンを継続している。オランダ、ポーランド、スロバキア、スペインでの選挙に先立ち、地域限定の選挙完全性運動を展開。2023年には欧州18カ国でメディアリテラシーキャンペーンを展開し、2億2,000万回以上のインプレッションを獲得、TikTokで約5,000万人にリーチした。この活動は今年も継続され、2024年にはさらに9つのキャンペーンが実施される予定。また、2024年6月のEU選挙に先立ち、全EU加盟国で地域限定メディア・リテラシー・アクティベーションが実施予定。 • コラボレーションがPF上の被害や悪用からの保護の取り組み強化に役立つと考えており、選挙スピーカー・シリーズの一環として、特にファクトチェック・コミュニティから外部の専門家を招き、見識や市場に関する専門知識を社内のチームと共有することで、選挙に対する私たちのアプローチに反映させる。 • サハラ砂漠以南のアフリカ地域にも国営メディア・ラベルの提供を拡大し、EUユーザーの透明性をさらに高めた。

出所）「Transparency Center」各団体レポート https://disinfocode.eu/reports-archive/?years=2023

主要5PF事業者の分野ごとのコミットメント内容（2024年3月現在） （4/5）

団体名	6.研究者団体のエンパワーメント（EMPOWERING THE RESEARCH COMMUNITY）	7.ファクトチェック団体のエンパワーメント（EMPOWERING THE FACT-CHECKING COMMUNITY ）
Google	• 欧州メディア情報基金への2,500万ユーロの寄付、テクノロジーにおける信頼と安全（Trust & Safety）に関する研究活動を支援する追加助成金、Google Research ProgramおよびYouTube Researcher Programを通じた助成金など、当社の製品を通じて研究者を支援。 • 2023年10月、 Trust & Safety Research Awards を開始。助成先は、ブルガリアのソフィア大学・コペンハーゲン大学の調査。 • 2023年8月、Google Research Programを通じ、EU研究者は、検索やYouTubeを含むGoogle製品の一部で一般公開されているデータへのアクセスを申請可能に。	• 国際ファクトチェック・ネットワークへの1,320万米ドルの助成金 • 2024年のEU議会選挙のファクトチェックのため、40以上のファクトチェック組織の連合体を構築するため、EFCSNに150万ユーロの助成金を交付。 • 2023年10月、ベルギーのブリュッセルで「ネット上の誤報との闘い」に関するサミットを開催。500人近い参加者が集まり、誤情報をめぐる最も差し迫った緊急の問題について議論し、増大する懸念に対処するための協力的なアプローチを模索することを目的とした。
Microsoft	• （前回レポートから追加なし）	• （前回レポートから追加なし）
Meta	• **研究者のデータアクセス**： 2023年11月、Instagramのクリエイターやビジネスアカウント、Facebookのページ、投稿、グループ、イベントから、リアルタイムの公開コンテンツへのアクセスを提供するコンテンツライブラリとAPIツールの展開を発表。リアクション数、シェア数、コメント数、初めて投稿の閲覧数など、コンテンツに関する詳細も利用可能。研究者は、グラフィカルUIまたはプログラムAPIを通じて、コンテンツ検索、フィルタリングが可能Metaコンテンツライブラリーツールをリリースしてから数ヶ月間、研究者が必要とする一般公開データを効果的に利用可能にするため、研究者から集めたフィードバックに基づき、2024年に、ダウンロードオプションや、Metaコンテンツライブラリ内の新しいデータタイプとしての「コメント」など、いくつかの新しいデータや機能を追加予定。	• 選挙の完全性を確保するため、EFCSNが欧州のファクトチェックコミュニティのために確立した基準を評価し、欧州のMetaファクトチェッキングプログラム検討の前提条件として、EFCSN認証の受け入れを開始予定。 • 独立したファクトチェック・パートナーが当社のアプリ上の偽コンテンツを迅速にレビュー・評価できるよう、グローバルなファクトチェック・プログラムを継続し、EUの22言語をカバーする26のファクトチェック機関と提携。 • 2023年7月1日から同年年12月31日の間に、EU圏のFacebookでは、14万以上のファクトチェック記事及び6,700万以上のコンテンツのラベル付けとバイラリティの低減の両方に利用された。Instagramでは、EU域内で3.6万以上の記事及び110万以上のコンテンツのラベル付けとバイラリティの低減に使用された。
TikTok	• PF運営・保護の方法への透明性をもたらすため、2020年からコミュニティガイドライン実施報告書を公開している。Research API（コンテンツやアカウントに関する公開データへのアクセスを研究者に提供するもの）をヨーロッパに拡大し、多くの改善を展開して以来、EEAの研究者からのAPIアクセスと使用の申請を49件承認した。 • TikTokの有料広告やその他の営利目的のコンテンツに透明性をもたらすCommercial Content APIへのアクセス申請を、50件以上受けた。	• クロアチアとポルトガルでファクトチェック・プログラムを開始し、新たに2社のファクトチェック・パートナーと提携することで、EUの18の公用語とEEAの24カ国の主要言語をカバーした。 • ファクトチェックのカバー範囲を拡大する努力により、2024年のEU議会選挙に先立ち、欧州加盟国の各公式言語を少なくとも1つはカバーできるようになる。

出所）「Transparency Center」各団体レポート https://disinfocode.eu/reports-archive/?years=2023

The 2022 Code of Practice on Disinformation

主要5PF事業者の分野ごとのコミットメント内容（2024年3月現在） （5/5）

団体名	8.透明性センター（TRANSPARENCY CENTRE）	9.常設タスクフォースへのコミット（PERMANENT TASK-FORCE）	10.行動規範のモニタリング（MONITORING OF THE CODE）
Google	• （前回レポートから追加なし）	• （前回レポートから追加なし）	• （前回同様記載なし）
Microsoft	• （前回レポートから追加なし）	• （前回同様記載なし）	• （前回同様記載なし）
Meta	• （前回同様記載なし）	• 過去3回の報告サイクルにわたって透明性レポートを作成し、タスクフォースと継続的に関与した結果、我々の測定基準の粒度が向上し、新しいワーキンググループ（生成 AIや選挙）を通じてタスクフォースのメンバーとの協力が強化された。 • タスクフォースのバランスの取れたアプローチと、構造化された対話と有意義な透明性のために設けるフォーラムを高く評価しており、EU議会選挙の完全性を守るために、欧州委員会、ERGA、EDMO、我々の共同署名者、そしてタスクフォースの他のメンバーとのパートナーシップに全力を尽くす。	• （前回同様記載なし）
TikTok	• （前回レポートから追加なし）	• TikTokは選挙に関するワーキンググループの共同議長であり、新設された生成AIワーキンググループにも参加している。	• （前回同様記載なし）

出所）「Transparency Center」各団体レポート https://disinfocode.eu/reports-archive/?years=2023

参考：主要5PF事業者の分野ごとのコミットメント内容（2023年7月現在）（1/5）

■ 公表されている署名事業者・団体の透明性レポートのうち最新版である2023年7月分（対象期間：2023年1月1日～6月30日）における「Executive Summary」の記載内容から一部抜粋し、行動規範の10分野別に分類した。

- 「Executive Summary」の記載内容や粒度は各社によって異なる。
- なお、Xは2023年5月に脱退したため、2022年12月分（対象期間：2022年7月1日～9月30日）の内容を整理した。

団体名	2.広告表示の精査 （SCRUTINY OF AD PLACEMENTS）	3.政治広告への対応 （POLITICAL ADVERTISING）
Google	• Google のサービス上で最も有害な行為やコンテンツを禁止するルールの開発と実施	• Refreshed Adsの指標を政治広告のセクションに組み込み、EUの新しい政治広告規制に基づいてこの分野の取り組みを更新予定
Microsoft	• Refreshed Adsの指標を政治広告のセクションに組み込み、EUの新しい政治広告規制に基づいてこの分野の取り組みを更新予定	• Microsoft Advertising は、国家が自国のプロパガンダを推し進めるために使用しているメカニズムへの資金提供を停止し、必要に応じてユーザーおよび広告主のポリシーを定期的に評価し、改善する。
Meta	• 2023年1月1日から同年6月30日までに、EU加盟国のFacebookとInstagramから690万件以上の広告を削除。そのうち、2.4万件以上の広告が誤情報ポリシー違反としてFacebookとInstagramから削除された。 • 同期間にいて、EU加盟国においてFacebookとInstagramの両方で68万件以上の広告に「paid for by」の免責事項を表示した。	• 国営メディアを指定する枠組みを構築する際の課題、透明性優先のアプローチに至ったトレードオフ、危機発生時に必要な適応など、国営メディアに関するポリシーの策定について概説した記事をLawfareに掲載した。 • 関連して、GraphikaはFacebookとInstagramのロシア国営メディアに関するレポートを発表し、侵攻から6ヵ月後、ページの投稿量は1年前の同じ日に比べて43％減少し、エンゲージメントレベルは80％低下したことを報告。
TikTok	• 広告表示に関しては当社の厳格なポリシーに準拠する必要があり、全広告が事前にポリシーに照らして審査される。当社の広告ポリシーでは、誤解を招く行為、不正な行為、欺瞞的な行為を特に禁止している。TikTokは継続的なポリシー改善を行い、医療誤情報、危険的誤情報、操作されたメディア、危険な陰謀論、といった4つの有害誤情報広告ポリシーを開発した。 • TikTok上の有料広告やその他の商業的コンテンツに透明性をもたらすために、Commercial Content APIを導入。 • 本レポートでは、違反広告の閲覧数を含む新たな指標を開示することで、当社PF上の広告に関する透明性をさらに高めたことを確認できる。	• 政治家が広告を掲載することを許可しておらず、政治的な広告を掲載することも許可していない。党派的な政治的動機によるものでなければ、政府機関、非営利団体、その他の団体による大義に基づく広告や公共サービス広告は許可している。
X（2023年1月現在）※2023年5月頃脱退	• 操作的な広告やスパム広告を禁止する明確なポリシーを定めている。 • 人々が期待する広告の透明性を提供するため、Twitter広告の透明性センターの再開を予定している。	• 国家が支援する情報操作との闘いに関するページでは、我々がどのように協調的なPF操作に効果的に取り組んでいるかを示している。PF操作とスパムに関するポリシーは堅固であり、脅威破壊チームはコミュニティノートの開発と並行して活動を続けている。

出所）「Transparency Center」各団体レポート https://disinfocode.eu/reports-archive/?years=2023

The 2022 Code of Practice on Disinformation

参考：主要5PF事業者の分野ごとのコミットメント内容（2023年7月現在）　（2/5）

団体名	4.サービスの完全性 （INTEGRITY OF SERVICES）
Google	• 新たにAI原則を更新した。 • Google検索では、生成AIコンテンツに関するガイダンスを発表し、検索上のコンテンツの高い情報品質と全体的な有用性を維持するためのアプローチを概説している。 • Googleは、ウォーターマーク、メタデータ、その他の技術における新たなイノベーションを最新の生成モデルに統合する予定である。 • 他の大手AI企業とともに、G7、OECD、各国政府の取り組みを支援する人工知能の開発における責任ある実践を推進協定を結ぶ。 • また、レポート内において、EEA における検索スパムポリシーの手動およびアルゴリズムによるアクションの加盟国の内訳や、YouTube による TTP 関連のマッピングと測定基準の拡大についても更新している。
Microsoft	• 自社サービスが偽情報の助長に使用されるリスクを軽減するため、高精度ツールとテクノロジーを活用し、より広範な社会と透明性をもって情報を共有することでAI研究のリーダーとしての役割を果たす。 • AIの責任ある実装を確実にするために、2022年6月、製品チーム全体の基本基準とガイダンスを設定するための「責任あるAI標準v.2」と「情報完全性原則」を発表。 • LinkedInは「責任あるAI 原則」のフレームワークを発表。コミットメント15の回答内で、発売した3つの生成AI製品に関して原則をどのように実践したかを説明している。 • TruepicとのProject Providenceを含むツールやCoalition for Content Provenance and Authenticity（C2PA）などを開発し、操作されたメディアや生成AIメディアの台頭に対抗している。 • 今後の主な取り組みとしては、一般市民が生成AIコンテンツを特定し、出所を理解するため、新しい出所確認ツールの導入することや、Microsoftとリンクトインの製品が、Microsoftの「責任あるAI基準」とリンクトインの「責任あるAI原則」に準拠して開発され、ユーザーに安全で信頼できる体験が提供されるようにし、Microsoft製品に含まれるAIシステムに当社の情報完全性の原則が統合されることに注力する予定。 • 新たな脅威やTTPが出現した場合、生成AIに関する学習の共有を継続し、偽情報の傾向とTTPへの対処のベストプラクティスを定期的に評価し、実施し、共有し続ける。
Meta	• Metaは、生成AIツールの普及と導入が、自社PFにおける偽情報の特定と対処方法に影響を与える可能性があることを認識しているため、Partnership on AI's Responsible Practices for Synthetic Mediaに署名し、ユーザーのためにオンライン情報環境の完全性を維持するための業界横断的な協力に取り組む。 • 新しいAI技術に反映させたい原則についてのフィードバックを得ることを目的とした「生成AIに関するコミュニティ・フォーラム」を、スタンフォードのDeliberative Democracy LabとBehavioural Insights Teamと協議して開催する予定で、これはAIモデルを共有するためのオープン・コラボレーション・アプローチの一環である。 • Code's Task Force Working Group on Generative AI のメンバーとして、他のメンバーと今後も協力していく予定である。
TikTok	• 当社のIntegrity and Authenticityポリシー（I&Aポリシー）は、欺瞞的な行為を断固として禁止しており、これらのポリシーを実施するために、さまざまな戦術、技術、および手順を使用している。 • 生成AIコンテンツのPF上での使用に対処するため、合成メディア・ポリシーの更新を開始した。ポリシーにより、ユーザーは、コンテンツがAIによって生成または操作されているが、現実的なシーンを示している場合、開示する必要がある。開示されていない、あるいは当社のポリシーに違反する合成メディアは、当社のプラットフォームで禁止されている。 • 秘密裏に行われる影響力工作（CIO）と闘い続け、アカウントの身元、出所、活動場所、人気、目的について当社PFのシステムやコミュニティを欺きながら世論を動かそうとする試みを禁止する。 • 当社は複雑な欺瞞行為を繰り返し調査・評価し、適切な製品とポリシーによる解決策を開発し続ける。
X（2023年1月現在） ※2023年5月頃脱退	• コミュニティ・ノートの投資を強化し、中央集権的で時間のかかるやり取りを必要とするコンテンツモデレーションへの依存を減らす。 • コミュニティ・ノートで実証されたテクノロジー・ファーストの戦略は、、スピードと規模という2つの課題に常に直面してきた集中型のコンテンツ調整方法にはない利点がある。 • Twitter上でユーザーの身元を認証することで、スパムやウイルスによる偽情報の蔓延を減らすことを目的としている。

出所）「Transparency Center」各団体レポート　https://disinfocode.eu/reports-archive/?years=2023

参考：主要5PF事業者の分野ごとのコミットメント内容（2023年7月現在） （3/5）

団体名	5.ユーザーのエンパワーメント（EMPOWERING USERS）
Google	• 明らかに有害でない限り、低品質なコンテンツを完全に削除するのではなく、不服申し立てメカニズムを通じて表現の自由を保護したり、PF上に現れる可能性のある低品質なコンテンツに対処するために権威あるコンテンツを提起したりするなど、適切かつ比例的な緩和手法を検討する。 • 欧州経済地域全域のユーザーへの情報パネル提供や、メディアリテラシーキャンペーンの実施などを通してユーザーを支援し、権威あるコンテンツにつなげるための取り組みを行う。 • Google Search は、Search Essentialsの一環として、スパムポリシーを更新した。 • Google検索は、検索結果の全体的な品質についてシステムが高い信頼性を持っていない検索に対して、コンテンツアドバイザリーを利用できる範囲をフランス語とドイツ語に拡大した。 • YouTubeの「Hit Pause」キャンペーンがEEA加盟国すべてで開始された。 • その他、誤報ポリシー違反で削除された動画の再生回数、誤報動画削除の復元数等についても報告している。
Microsoft	• MicrosoftのPFや製品を通じてユーザーが取得する情報をよりよく理解できるようにするためのパートナーシップを結ぶ。チェコのカレル大学とVerifyにAzureクレジットを提供し、彼らのサイバースキリング活動を公開し、数百人の高校生のアクセスを可能にした。 • 2022年には、情報リテラシープログラムへの投資を拡大し、非営利団体と提携して業界の研究とベストプラクティスに基づいたキャンペーンを展開した。 • Minecraftのゲーム開発者のよるMinecraft Educationの中でのメディアリテラシーの中核概念を探求するゲーム立ち上げや、Microsoft Teamsで教育者と生徒が信頼できるリソースを特定するのに役立つ無料アプリの立ち上げを支援した。 • Bingの取り組みには、多くの場合、ターゲットを絞ったランキング介入、高オーソリティのソースを指すアンサー、信頼性のシグナル、またはコンテンツの出所表示などのデジタルリテラシーの追加機能などのアクションがコンテンツ削除より効果的であることが判明した。今後も定期的に対策の有効性を検証し改善すべき分野を特定するとともに、社内外の専門家と協力し、ユーザーが検索結果に含まれる有害なコンテンツに意図せずさらされることを防ぐ。 • The New Bingでは、『Our approach to Responsible AI』において、AIについて透明性を示している。今後もユーザーや外部からのフィードバックに基づいて、これらの機能を進化させ続ける。 • 今後もメディアリテラシーとクリティカルシンキングの分野における取り組みを強化し、資金を拡大し、ユーザーが訪問しているサイトやドメインの信頼性を理解するのに役立つツールや機能を追加し、情報源について十分な情報を得た上で判断できるようにする。
Meta	• 2023年夏に、リトアニア、スロバキア、ブルガリアで、一連のメディア・リテラシー・キャンペーンを開始した。これらのキャンペーンは、特にスロバキアで予定されている選挙に関連して、適切でインパクトのあるキャンペーンを構築するため、2023年前半に専門家と話し合ったものである。 • 当社は偽アカウントを削除するための積極的な措置を講じており、危害を加える意図のある偽アカウントを優先的に削除している。第1四半期には、4億2,600万件の偽アカウントに対して対策を講じ、2023年第2四半期には、全世界で6億7,600万件の偽アカウントに対して対策を講じた。
TikTok	• ユーザーが当社のコミュニティガイドライン（CG）違反のコンテンツや違法と疑われるコンテンツに遭遇した場合、アプリ内報告ツールにアクセスしやすく、簡単に使用できる。 • ポリシー違反コンテンツを体系的に削除するとともに、特定のコンテンツに関する背景をユーザーに示し、権威情報へ誘導し、潜在的な誤情報を報告するよう促すアプリ内対策に注力した。 • ファクトチェック・パートナーとの緊密な協力のもと、様々メディア・リテラシー・キャンペーンを継続している。ウクライナ戦争に関連する8つの地域別キャンペーンや、スペイン、ギリシャ、フィンランドの選挙を前にしたキャンペーンを展開した。ポーランドとスロバキアの選挙に向けても展開予定。 • For Youフィードは、ユーザーがTikTokを開いたて最初に表示されるインターフェースだが、これがパーソナライズされたレコメンデーション・システムに基づいていることを明確にし、レコメンデーション・システムの運用方法についてユーザーに提供する情報を刷新した。 • ユーザーにより多くの選択肢を提供するため、For Youフィードを更新できる機能を導入。また、パーソナライゼーションのオフを可能にし、コンテンツを発見する方法を増やした • 推奨システムでは、当社CGに違反する有害な誤報コンテンツの削除だけでなく、一般視聴者にとって不適切な可能性のある特定のカテゴリのコンテンツを推奨しない措置を講じている。一般的な陰謀論、緊急事態や未検証の情報、ファクトチェックの評価を受けてる潜在的有害性の高い誤報など、2023年8月以降、コンテンツが推薦に不適格とされた場合、クリエイターに通知し、異議申し立てができるようにしている。
X（2023年1月現在）※2023年5月頃脱退	• ユーザーをエンパワーするアプローチとして、コミュニティノート機能を中心に据えており、コミュニティノートの広範な取り組みを通じて、ユーザーの偽情報の特定と緩和を支援する。 • コミュニティノートでは、誤解を招くツイートに対してユーザーが有益な注釈を加えられる。また、コミュニティ・ノートへのすべての投稿は公開されるため、誰でもデータ分析が可能。 • Twitterでコンテンツを提案するアルゴリズムと逆時系列フィードを直感的に切り替えられる機能を追加し、タイムラインでコンテンツを推薦するアルゴリズムをオープンソース化する予定。

出所）「Transparency Center」各団体レポート https://disinfocode.eu/reports-archive/?years=2023

The 2022 Code of Practice on Disinformation

参考：主要5PF事業者の分野ごとのコミットメント内容（2023年7月現在）　（4/5）

団体名	6.研究者コミュニティのエンパワーメント（EMPOWERING THE RESEARCH COMMUNITY）	7.ファクトチェック団体のエンパワーメント（EMPOWERING THE FACT-CHECKING COMMUNITY ）
Google	• 欧州メディア・情報基金（European Media and Information Fund）への2,500万ユーロの初回投資（これまでに47件に資金を提供）などを通じて、研究者を支援。	• 欧州メディア・情報基金への2500万ユーロの投資（欧州全域で47のプロジェクトに資金を提供）、国際ファクトチェック・ネットワークへの1320万米ドルの寄付を行う。 • YouTubeが主催する「グローバル・ファクト10」は、ファクトチェッカーの国際的な集まりで、トレンドやテクノロジーについて議論された。
Microsoft	• 偽情報および広範な偽情報の傾向と戦術に関する誠実な研究を支援する。	• ファクトチェック機能を拡張し、信頼性のシグナルとファクトチェックを進化するテクノロジープラットフォームに統合する革新的な方法を模索するため、複数のファクトチェック団体と継続的に協議している。 • EUにおけるファクトチェック適用範囲の拡大のための新たなパートナーシップを構築し、Microsoftのサービス上のコンテンツをユーザーが評価できるようにするためのさらなる方法を引き続き模索する。
Meta	• MetaのPF上の公開コンテンツに対する独立したリサーチをサポートするための、メタ・コンテンツ・ライブラリーとAPIを開発。コンテンツライブラリには、Facebook（公開投稿、ページ、グループ、イベント）とInstagram（クリエイターとビジネスアカウント）のほぼリアルタイムの公開コンテンツが含まれ、ライブラリからのデータは、グラフィカル・ユーザー・インターフェースまたはプログラムAPIを通じて検索、探索、フィルタリングすることが可能。これらを組み合わせることで、これまで研究者に提供してきたツールの中で、FacebookとInstagramの公開コンテンツへの最も包括的なアクセスを提供する。	• 2023年1月1日から同年6月30日の間に、全世界のFacebook上で19万以上の明確なファクトチェック記事が、EU域内の4,000万以上のコンテンツにラベルを付け、バイラリティを低下させるために使用された。Instagramに関しては、全世界で5.2万以上の記事が、EU域内で110万以上のコンテンツにラベル付けされ、そのバイラリティを減少させるために使用された。これは、独立したファクトチェッカーの仕事を拡大する我々のツールの力を実証している。 • 当社は、業界最大規模のグローバル・ファクトチェッキング・プログラムを維持しており、独立系ファクトチェック・パートナーは当社のアプリ上の虚偽コンテンツを迅速にレビュー・評価がが可能で、EUの22の言語をカバーする26のファクトチェック機関と提携。 • ファクトチェックされたコンテンツをシェアする意思を示す人々に特化した、ファクトチェックラベルの影響の指標を共有する。
TikTok	• 当社は2020年からCGエンフォースメント・レポートを公開し、当社PFの運営と保護方法について透明性を高めている。最近では、Research API（当社PFからコンテンツやアカウントに関する公開データへのアクセスを研究者に提供するもの）をヨーロッパに拡大し、多くの改善を展開した。 • DSAの下で指定された審査済みの研究者とデータを共有するプロセスを試行するため、EDMOとのデータアクセス・パイロットに参加している。	• IFCNが主催する年次GlobalFact10サミットのスポンサーを務め、そこでの発表を行った。 • TikTokのファクトチェック・プログラムやファクトチェッカーの意見を、より広範なコンテンツモデレーションに取り入れており、ファクトチェックのアウトプットが効果的に増幅され、偽情報コンテンツやトレンドがより包括的かつ広範に対処される。 • 欧州全域でファクトチェック・プログラムの急速なスケールアップを進め、新たに9つのEU諸国でプログラムを開始した。 • ファクトパートナーがより多くの言語で多くの種類の主張を論破し、追加のアドホック・プロジェクトをサポートできるよう、パートナーとの契約を見直した。また、彼らから潜在的な誤報のフラグを積極的に受け取る機能を強化し、彼らのフィードバックに関する実施データを共有する試験的なスキームを実施。
X（2023年1月現在） ※2023年5月頃脱退	• Twitterは、学術研究のためのデータ共有に関して、プラットフォーム分野で最もオープンなアクターのひとつである。国家による広範な情報操作の詳細を示す大規模なデータセットが、世界の学術コミュニティに提供されている。 • TwitterのAPIプログラムも学術研究者の間で広く利用されている。	• （記載なし）

出所）「Transparency Center」各団体レポート　https://disinfocode.eu/reports-archive/?years=2023

参考：主要5PF事業者の分野ごとのコミットメント内容（2023年7月現在）　（5/5）

団体名	8.透明性センター（TRANSPARENCY CENTRE）	9.常設タスクフォースへのコミット（PERMANENT TASK-FORCE）	10.行動規範のモニタリング（MONITORING OF THE CODE）
Google	• Googleでは、今後も規約で求められている通り、各申告に関連する6ヶ月間のレビュー期間に焦点を当て、本レポートの後続版を隔年で発行する。	• 透明性レポートは、常設タスクフォース（Permanent Task-force）が策定した構成とテンプレートに従い、本規範のコミットメントと章を中心に構成される。 • Googleは、2023年前半まで行動規範の義務を果たし続けてきた行動規範の常設タスクフォースのメンバーとして、引き続きコミットし、生産的な活動を行う。	• （記載なし）
Microsoft	• 既存の調査ツールを強化し、より充実したデータ報告を提供するとともに、偽情報の拡散に関する調査を支援するため、関連するデータと調査の提供を継続する。	• （記載なし）	• （記載なし）
Meta	• （記載なし）	• タスクフォースと緊密に協力し、共に改善を続けていくことを約束する。 • 欧州委員会、ERGA、EDMO、共同署名事業者・団体、およびタスクフォースの他のメンバーとの関与と対話を継続し、我々の慣行と透明性の両方を強化していく。 • 来るべきEU選挙に備え、準備態勢と効果的なマルチステークホルダー協力を確保するために、タスクフォースの一員として協力する。	• （記載なし）
TikTok	• 2024年の次回の行動規範報告書に向けて、当社ポリシーとツールの開発と強化を継続することを約束する。	• 規範のタスクフォースとそのすべてのワーキンググループおよびサブグループに有意義に関与し続ける。 • TikTokは選挙に関するワーキンググループの共同議長であり、TrustLabの構造的指標試験運用のサポートにおいて主導的な役割を果たしている。 • 規範のタスクフォースおよびそのすべての作業部会とサブグループを通じて、業界およびその他のパートナーとの協力関係を継続していく。	• （記載なし）
X（2023年1月現在）※2023年5月頃脱退	• Twitterは10年以上前の2012年に最初の透明性報告書を発表した。それ以来、Twitter Transparency Centreはほぼ毎回、より詳細な報告書を発行しており、現在では法的要請と利用規約違反の両方に関する国レベルのデータを提供している。	• （記載なし）	• （記載なし）

出所）「Transparency Center」各団体レポート　https://disinfocode.eu/reports-archive/?years=2023

2．英国：オンライン安全法におけるリスク評価の考え方

オンライン安全法の概要と目的・経緯

項目	内容
概要	• 英国では、違法又は子供に有害なコンテンツや活動によるリスクを特定・軽減・管理する義務をオンラインサービスの提供者に課し、個人にとってより安全なオンラインサービスの提供を確保することを目的として、**英国オンライン安全法（Online Safety Act, 2023）が2023年10月26日に制定**された • 同法は、有害なコンテンツから児童を保護する一方で、成人に対してはオンラインで閲覧できるコンテンツの選択肢を増やすことを目指すとしている
目的	**5つの政策的な目的**がある • オンラインにおける利用者の安全性を高めること • オンラインにおける言論の自由を維持・強化すること • オンライン上の違法コンテンツに対処する法執行能力を向上させること • 利用者のオンラインにおける安全確保能力の向上 • 被害状況に関する社会の理解を深めること
経緯	• 英国政府はオンライン上の安全性確保・向上を目的に、2019年4月にOnlineHarmsWhitePaperを公表 • その後、パブリックコメントを実施し、2021年5月にはオンライン安全法案（Online Safety Bill、以下OSB）の草案が公表された。 • OSBの草案公表以降、英国議会の合同委員会やDCMS小委員会でOSB草案について検討・議論が行われ、2022年3月17日、上記検討の結果を踏まえ修正されたOSBが英国議会（下院）に提出された • その後、英国議会（下院）での議論・合意を経た後、2023年1月18日上院に提出され、2023年9月12日に上院による修正案を英国議会（下院）で稟議し、2023年9月19日に修正案に同意 • 2023年10月26日、英国オンライン安全法（Online Safety Act, 2023）として制定された

https://commonslibrary.parliament.uk/research-briefings/cbp-9579/

オンライン安全法の目次・構成

項目			条項
Part 1	イントロダクション・全体概要		第1条-2条
Part 2	用語の定義		第3条-5条
Part 3	ユーザー間サービスや検索サービスに課される義務	1章：イントロダクション	第6条
		2章：ユーザー間サービスの注意義務	第7条-23条
		3章：検索サービスの注意義務	第24条-34条
		4章：子供のアクセス評価	第35条-37条
		5章：不正広告に関する義務	第38条-40条
		6章：行動規範とガイダンス	第41条-54条
		7章：Part3の解釈	第55条-63条
Part 4	ユーザー間サービスや検索サービスに課される更なる義務	1章：本人確認	第64条-65条
		2章：子供の性的搾取と虐待に関するコンテンツの報告	第66条-70条
		3章：利用規約：透明性、説明責任、表現の自由	第71条-74条
		4章：死亡した子供の利用者	第75条-76条
		5章：透明性レポート	第77条-78条
Part 5	ポルノコンテンツを提供する事業者に課される義務		第79条-82条
Part 6	違反時の罰則（罰金）		第83条-90条

項目			条項
Part 7	OFCOMの権力と義務	1章：一般義務	第91条-93条
		2章：規制対象となるユーザー間サービスおよび検索サービスのカテゴリー登録	第94条-97条
		3章：検索サービスの注意義務	第98条-99条
		4章：インフォーメーション	第100条-120条
		5章：テロ・コンテンツおよびCSEAコンテンツに対処するための通知	第121条-129条
		6章：執行権限	第130条-151条
		7章：委員会、調査及びレポート	第152条-164条
		8章：メディアリテラシー	第165条-166条
Part 8	不服申し立てと苦情	1章：不服申し立て	第167条-168条
		2章：苦情	第169条-171条
Part 9	規制サービスに関する国務長官の機能		第172条-178条
Part 10	通信に関する犯罪		第179条-191条
Part 11	補足		第192条-225条
Part 12	解釈と最終規定		第226条-241条

https://www.legislation.gov.uk/ukpga/2023/50/contents/enacted
https://www.OFCOM.org.uk/__data/assets/pdf_file/0027/211986/understanding-online-false-information-uk.pdf

対象事業者・対象コンテンツ

項目	内容	条項
対象事業者	● 法律の対象となる特定のPFプロバイダー ✓ **ユーザー間サービス**–ユーザーがコンテンツを作成して共有したり、相互にやり取りしたりできるサービス。例としては、あらゆるソーシャルメディアやアプリ、写真/ビデオ共有デバイス、インスタントメッセージサービス、オンラインゲームサービスなど ✓ **検索サービス**–ユーザーが他のウェブサイトやデータベースを検索できるサービス ● 「ユーザー間サービス」又は「検索サービス」については、それが英国外から運営されている場合であっても、「英国との関連性を有する」サービスである限り、英国オンライン安全法の域外適用があるとされている	● 3条1項 ● 3条4項 ● 4条2項a号

項目	内容	条項
違法コンテンツ（Illegal content）	同法における違法コンテンツとは、テロ、児童の性的搾取、自殺勧奨、自傷行為、ハラスメント、ヘイトクライム、支配行為、薬物犯罪、武器関連犯罪、入国管理法違反、人身売買、成人の性的搾取、過激なポルノ、親密な画像の乱用、犯罪収益、詐欺、外国干渉などを対象とする犯罪を示す。 ※後述するOFCOMのコンサルテーション（ガイダンス・行動規範）の中で、違法危害（illegal harms）で15種類の犯罪を指定している	● 59条
子供に有害な内容	「子供に有害な内容」とは、 (a) 子供に有害な最優先コンテンツ • ポルノコンテンツを含み、自殺を助長し、自傷行為を助長し、摂食障害または摂食障害に関連する行動を助長するコンテンツ (b) 子供に有害な優先コンテンツ • 人種、宗教、カースト、性別などを虐待したり、標的にしたりするコンテンツ • 憎悪を扇動したり、暴力を助長したり、いじめのコンテンツや、重傷を負わせる可能性のあるスタントを奨励するコンテンツ (c) (a) または (b) 以外のコンテンツで、英国のかなりの数の子供たちに重大な危害を及ぼす重大なリスクをもたらす種類のコンテンツ。（a material risk of significant harm to an appreciable number of children in the United Kingdom）	● 60条 ● 61条 ● 62条 ● 63条

https://www.legislation.gov.uk/ukpga/2023/50/contents/enacted
https://www.OFCOM.org.uk/__data/assets/pdf_file/0027/211986/understanding-online-false-information-uk.pdf

オンライン安全法における偽誤情報の位置づけ

偽誤情報の定義

■ オンライン安全法における違法コンテンツの定義の中に、偽情報は含まれていない。

- 参考：英国政府は、「偽情報（disinformation）を、人々に危害を与えるため、あるいは政治的、個人的、金銭的利益を得るために、人々を欺き、誤解させることを意図した虚偽の情報および／または操作された情報を意図的に作成し、広めること」と定義している。また、「誤情報（misinformation）とは、不注意による虚偽の情報の拡散である」と定義している

オンライン安全法における偽誤情報に関連する項目

■ 同法の中で、**偽誤情報に関連す項目としては、大きくは三つ**あり、偽誤情報のアドバイザリー委員会の設置（後頁参照）と、新たな虚偽通信罪の規定、OFCOMのメディアリテラシー義務に関する規定である。

- 虚偽通信罪については、同法の179条で規定されており、虚偽であると知っている情報を、情報が心理的または身体的危害を与えることを意図していた場合、および、その情報を送信することについて合理的な理由がない場合に違反となるとしている。
- OFCOMのメディアリテラシー義務については、同法の165条において、 規制対象サービスを利用する際に、自分自身や他人を守ることができる方法について、一般市民の認識と理解を高めるための措置を講じることをOFCOMに求めており、例示として「偽情報と誤報の性質と影響」を理解することが挙げられている。

オンライン安全法における偽誤情報の位置づけに対する評価

■ 英国のファクトチェック団体であるFull Factは、オンライン安全法は利用規約にどのような内容を盛り込み、どのように対処・監督するかの規定が不足しているとし、誤情報の拡散を防ぐために十分ではないとの意見を表明している。

https://commonslibrary.parliament.uk/research-briefings/cdp-2024-0003/
https://www.legislation.gov.uk/ukpga/2023/50/contents/enacted
https://www.OFCOM.org.uk/__data/assets/pdf_file/0027/211986/understanding-online-false-information-uk.pdf

参考：“Advisory committee on disinformation and misinformation” について

■ 同法152条において、偽誤情報のアドバイザリー委員会の設置が義務付けられている。

- OFCOMは、OFCOM法（ Office of Communications Act 2002）の別表第14項に基づく権限を行使し、本項に規定する助言を提供する委員会を設置し、維持しなければならないと規定されている
- OFCOM法の別表第14項は、「Committees of OFCOM and advisory committees」であり、OFCOMの職務遂行に関する事項について、委員会や諮問委員会をOFCOMが設置できること、また、必要に応じてOFCOM以外のメンバーから委員会を構成できることが明記されている

偽誤情報のアドバイザリー委員会の概要

項目	概要
設置根拠・主体	• オンライン安全法の152条にもとづく（上記の通り、オンライン安全法152条における委員会の設置は、OFCOM法におけるOFCOMの委員会設置の権限に関する規定にもとづく） • OFCOMが委員を指名し、設置する
委員会の構成	• 委員長と構成員（OFCOMの指名により構成される） • 構成員には、(a)規制対象となるサービス利用者の代表者(b)規制サービスの代表者(c)オンライン上の偽誤情報の防止および処理に関する専門知識を有する者　が含まれることが望ましいとされている
委員会の役割	• OFCOMに対して、以下に関する助言を提供することが求められている • 規制対象サービスの当該サービス上の偽誤情報への対処に関する助言 • 偽誤情報に関して、規制対象サービスに課す透明性レポート（同法77条）やメディアリテラシー（同法165条、ならびに通信法11条）に対するOFCOMの権利行使に関する助言
レポートの提出義務	• アドバイザリー委員会は設置から18か月以内に報告書を公表すること、また、その後定期的に報告書を公表することが義務付けられている

https://www.legislation.gov.uk/ukpga/2023/50/contents/enacted
https://www.OFCOM.org.uk/__data/assets/pdf_file/0027/211986/understanding-online-false-information-uk.pdf

事業者に課される義務・違反時の罰則の概要

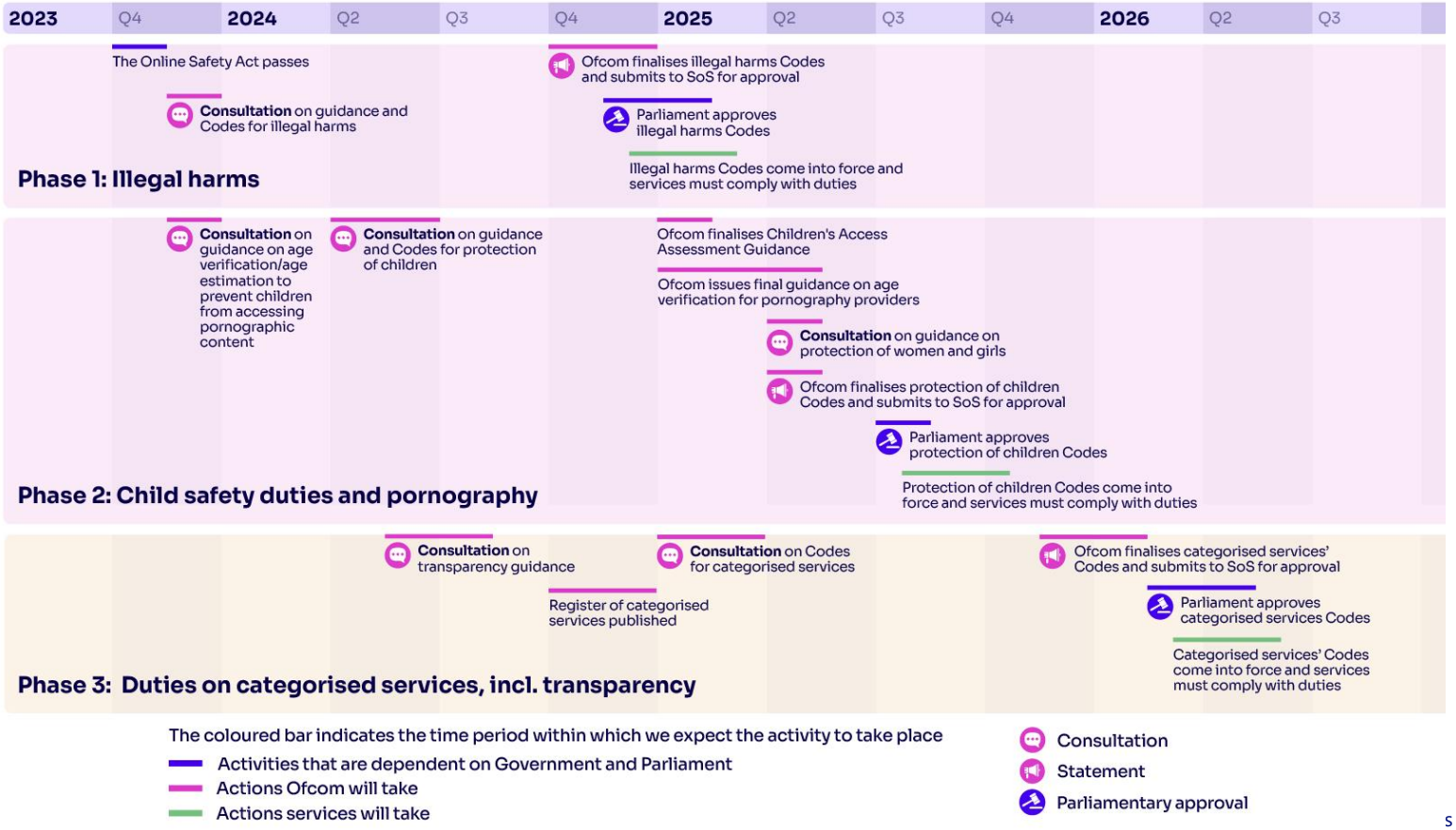
項目	内容	条項
事業者に課される主な義務	● 違法コンテンツ（Illegal content）に関するリスク評価の義務：違法コンテンツが個人に及ぼすリスク（ 9条5項(b)the level of **risk of individuals** who are users of the service encountering the following by means of the service）に関する評価を実施し、サービスに重要な変更を加える場合を含めて、アップデートを実施 ● 違法コンテンツに関する安全義務：サービスのデザイン及び運用に関して、違法コンテンツに関するリスクを効果的に管理・軽減する方策の採用、違法コンテンツを速やかに除去するためのシステム及びプロセスの導入 、利用規約又は公表文書において違法コンテンツからの保護に関する措置の開示 ● コンテンツ報告及び不服申立てに関する義務：ユーザーや影響を受ける個人が違法コンテンツや（子供がアクセス可能な場合）子供に有害なコンテンツを容易に報告できる仕組みの設定、ユーザー等からの不服申立て手続の整備及び不服申立てに対する対応の実施	● 9条、26条 ● 10条、27条 ● 20条、21条、31条、32条
子供がアクセスする可能性の高いサービスに対しての追加義務	● 子供に関するリスク評価の義務：コンテンツが子供に及ぼすリスクに関する評価を実施し、サービスに重要な変更を加える場合を含めて、アップデートを実施 ● 子供の保護に関する安全義務： サービスのデザイン及び運用に関して、子供に及ぼすリスクを効果的に管理・軽減する方策を採用、コンテンツのリスクから子供を保護するためのシステム及びプロセス（「ユーザー間サービス」では年齢認証・推計）の導入 、利用規約又は公表文書において子供の保護に関する措置の開示	● 11条、28条 ● 12条、29条
特定のカテゴリーのサービスに対しての追加義務	特定のカテゴリーのサービスは今後、別途定められる予定の規則によって規定される ● ユーザーエンパワーメントに関する義務：ユーザーエンパワーメントに関する評価を実施し、サービスに重要な変更を加える場合を含めて、アップデートを実施、大人のユーザーがコンテンツコントロールを行える仕組みを採用、可能な最も早い機会において、デフォルト設定を維持するか、変更するかを選択できるシステム及びプロセスの導入、利用規約において利用可能なコントロール機能及び直近のユーザーエンパワーメントに関する評価の要旨を開示 ● 詐欺的広告の防止義務：詐欺的広告を速やかに除去するためのシステム及びプロセスを導入し、そのために用いている技術を利用規約において開示 ● 本人確認の義務：大人のユーザーに対して、サービスの利用に本人確認が不要の場合でも、本人確認のオプションを付与し、利用規約において開示	● 14条、15条 ● 38条 ● 64条
違反時の罰則	● 1,800万ポンド、または、当該企業の最終事業年度における全世界売上高の10％のいずれか高い額を上限とする制裁金が課される可能性がある	● スケジュール13の4条1項、5条3項

https://www.legislation.gov.uk/ukpga/2023/50/contents/enacted
https://www.theverge.com/23708180/united-kingdom-online-safety-bill-explainer-legal-pornography-age-checks
西村・あさひ法律事務所「英国オンライン安全法（Online Safety Act）の解説-その適用範囲と要対応事項の概要-」

監督・執行体制と執行に向けたスケジュール

- オンライン安全法の監督・執行はOFCOMが担う
  - オンライン安全法の施行に際しては、**OFCOMに対して、PF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（同法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（同法52条、53条、54条等）**
- OFCOMは3つの段階に分けて具体的な施行に向けたガイダンスと行動規範を公表するとしており、その第一段階として、違法な危害に対するガイダンスと行動規範に関するコンサルテーションを公表し、2024年2月23日までパブリックコメントを受け付けた

オンライン安全法の施行に向けたスケジュール



OFCOMのコンサルテーション “Protecting people from illegal harms online”の概要

- OFCOMは最初のコンサルテーション（パブリックコメント）として、**“Protecting people from illegal harms online”を公開**
  - 24年2月23日までパブリックコメントを受け付けた
- **コンサルテーションの中では、行動規範についてのOFCOM案も提示している**

コンサルテーション“Protecting people from illegal harms online”の構成・概要

項目		概要
Volume 1	背景	• オンライン安全法の策定の背景・概要とともに、対象となるサービスを説明 • サービスのタイプに応じて対応を変える必要があること、OFCOMが提示する行動規範は時の変化とともにアップデートされるものであること、サービスやリスクに応じて様々な対応をとる必要があることに言及
Volume 2	オンライン被害の原因と影響	• オンライン被害の原因と影響に関するOFCOMの調査結果とそれに基づく分析の概要を提示 • 具体的には過去3年に渡る調査結果をもとにオンラインでの違法被害に対する原因と影響を概説（成人のインターネット利用者の87%がオンライン上での詐欺等に遭遇したことがあること等を例示）
Volume 3	オンライン上のリスクの評価方法	• サービス事業者がリスク管理のためにどのようなガバナンスを敷くべきか、違法被害のリスクを評価するために何をすべきか、記録保持と報告義務をどのように果たすべきかについてOFCOMとしての方針を提示 • リスク評価のプロセスやリスク評価を行う際のエビデンスについてのガイダンスを提示
Volume 4	リスクを軽減するための方法（行動規範）	• 違法コンテンツによる被害を軽減するためにサービス事業者がとるべき推奨措置について説明 • 大きくは、ユーザー間サービス（User to User)と検索サービス（Search）に分けて対策を項目別に提示 ※ガバナンスとアカウンタビリティはVolume3の中で提示。ただし、行動規範に含まれるとしている。
Volume 5	違法コンテンツの判別方法	• オンライン安全法にもとづくコンテンツの判別方法についてのアプローチを説明 • 違法コンテンツの判定ガイダンス（ Illegal Content Judgements Guidance “ICJG”）にもとづく判定や自社での規約にもとづくアプローチ等を提示
Volume 6	執行と監督へのアプローチ	• OFCOMの各事業者に対する情報収集の権限の概要と基本的なアプローチを説明 • OFCOMの執行権限とその行使に対する基本的なアプローチを説明

出所：Protecting people from illegal harms online

コンサルテーションの中で、リスク評価については、4つのステップで行うことが提唱されている。

■ コンサルテーションの中で、リスク評価の具体の進め方についてのOFCOM案を提示している

オンライン安全法におけるリスク評価のプロセス（OFCOM提示案）

	主な内容	アウトカム
Step1 危険性の理解	• OFCOMが提示するリスクプロファイル等に沿ってリスク要因を考慮したうえで、評価が必要な違法な危険性を認識する	• 15種類の違法な危害についての理解
Step2 リスクの評価	• 各種の違法な危害の可能性とその影響を評価し、危害のリスクレベルを割り当てる	• 自社サービスに対する、15種類の違法な危害への評価結果 • 15種の危害それぞれに対するリスクの程度（低・中・高）の割り当て
Step3 対策の実行と記録	• OFCOMが提示する行動規範を含めた適切なリスク軽減措置の決定 • 対策を実施するとともに、リスク評価の結果を記録する	• リスク評価と対策の記録の完成
Step4 レポートとリスク評価の監視と更新	• 関連するガバナンスの経路を通じてリスク評価と対策について報告をする • 効果を監視するとともに、リスク評価へのレビューを行うこと	• リスク評価の年次レビューのサイクルの確立（年に一度のレビューを提示） • リスク評価のレビューのきっかけ・要因となる要素の理解

出所：Protecting people from illegal harms online

参考：15種類の違法な危害

■ オンライン安全法はあらゆる種類の違法コンテンツを対象とするが、考慮すべき特定の犯罪リストとして15の犯罪をコンサルテーションの中で提示している

1. テロ：terrorism offences;
2. 児童の性的搾取・虐待：child sexual exploitation and abuse (CSEA) offences, including grooming and child sexual abuse material (CSAM);
3. 自殺ほう助：encouraging or assisting suicide (or attempted suicide) or serious self-harm offences;
4. ハラスメント・ストーキング・脅迫・虐待：harassment, stalking, threats and abuse offences;
5. 憎悪：hate offences;
6. CCB犯罪：controlling or coercive behaviour (CCB) offence;
7. ドラッグ：drugs and psychoactive substances offences;
8. 銃器・武器：firearms and other weapons offences;
9. 不法移民・人身売買：unlawful immigration and human trafficking offences;
10. 成人の性的搾取sexual exploitation of adults offence;
11. 極端なポルノextreme pornography offence;
12. 親密画像の悪用：intimate image abuse offences;
13. 犯罪行為による収益：proceeds of crime offences;
14. 詐欺・金融サービス犯罪：fraud and financial services offences; and
15. 外国干渉罪：foreign interference offence (FIO)

https://www.ofcom.org.uk/online-safety/information-for-industry/guide-for-services

参考：リスク評価の枠組み

■ 現状はコンサルテーションのため、事業者へのリスク評価の義務はないが、ガイダンスの最終版が公表されてから、3か月以内にリスク評価を行うこととされている

リスク評価の概要

What do you need to assess?	What should you take into account?	How do you make judgements about risk?
The **likelihood and impact** of each of the 15 kinds of illegal priority harm	Ofcom's Risk Profiles – which help you identify your **risk factors** – and any relevant **characteristics** of your service, including user base, functionalities, algorithmic systems, business model, any user protection or risk mitigation measures, and other relevant aspects of the service's design and operation, and the way the service is used	Review **evidence** about how harm could be experienced on your service and how your service's characteristics increase or decrease risks
Outcome: an assessment of low, medium, or high risk for each kind of illegal harm Our detailed guidance offers help on how to make this decision.		

出所：Protecting people from illegal harms online

参考：リスクプロファイル

- OFCOMはリスクの最終的な評価は各事業者が行うこととしているが、リスク評価のガイドとしてリスクプロファイルを提示している
- 具体的には、自社サービスがどのリスク要因に該当する可能性があるかを判別するための参考として、ユーザー間サービス（User to User)と検索サービス（Search）別に評価項目リストを提示
  - ユーザ間サービス、検索サービスそれぞれに3段階でのリスク評価のガイドを提示

OFCOMが提示するリスクプロファイルの手順

	ユーザ間サービス	検索サービス
①サービス特性の評価	• サービスタイプ（ソーシャルメディア、ゲーミング等）や特定の機能への該当性など8項目で設問を設定し、サービスの該当性を確認	• サービスタイプ、子供のアクセス可能性、特定機能の3つの項目で設問を設定し、該当性を確認
②特定のリスク要因への評価	• 上記の8項目ごとに、リスク要因や違法な危害等について提示	• 上記の3項目ごとに、リスク要因や違法な危害等について提示
③一般的なリスク要因へ評価	• ユーザ間サービス・検索サービスに共通するリスク要因として、ユーザ人口構成、ビジネスモデル、サービス・ビジネスの成熟度（ステージ）の3つの観点からリスク要因を提示	

出所：Protecting people from illegal harms online

参考：ユーザ間サービスに向けたリスクプロファイル（一部抜粋）

①サービス特性の評価（一部を抜粋）

Select Yes (Y) or No (N) for the following questions about your U2U service.	
1. Is my service any of the following service types? Select all that apply:	
a. **Social media service** (services which connect users and enable them to build communities around common interests or connections)	Y / N
b. **Messaging service** (services that are typically centred around allowing users to send messages that can only be viewed or read by a specific recipient or group of people)	Y / N
c. **Gaming service** (services which allow users to interact within partially or fully simulated virtual environments)	Y / N
d. **Adult service** (services which are primarily used for the dissemination of user-generated adult content)	Y / N
e. **Discussion forum or chat room service** (services which allow users to send or post messages that can be read by the public or an open group of people)	Y / N
f. **Marketplace or listing service** (services which allow users to buy and sell their goods or services)	Y / N
g. **File-storage and file-sharing service** (services whose primary functionalities involve enabling users to store digital content and share access to that content through links)	Y / N
2. Does my service allow child users to access some or all of the service?[33]	Y / N
3. Does my service have any of the following functionalities related to how users identify themselves to one another? Select all that apply:	
a. Users can display identifying information through a user profile that can be viewed by others (e.g. images, usernames, age)	Y / N
b. Users can share content anonymously (e.g. anonymous profiles or access without an account)[34]	Y / N
4. Does my service have any of the following functionalities related to how users network with one another? Select all that apply:	
a. Users can connect with other users[35]	Y / N
b. Users can form closed groups or send group messages	Y / N
5. Does my service have any of the following functionalities that allow users to communicate with one another? Select all that apply:	
a. Livestreaming (either open or closed channels)	Y / N
b. Direct messaging (including ephemeral direct messaging)	Y / N
c. Encrypted messaging	Y / N
d. Commenting on content	Y / N
e. Posting or sending images or videos (either open or closed channels)	Y / N
f. Posting or sending location information	Y / N
g. Re-posting and forwarding content	Y / N

②特定のリスク要因への評価（ソーシャルメディアサービスの例）

自社のサービスがソーシャルメディアサービスに該当する場合は、ソーシャルメディアサービスが持つリスク要因に対する理解と自社サービスのリスクの評価が求められる

Specific risk factors

U2U services with relevant characteristics should take account in their risk assessment.

1. Service type factors

☐	**1a** Social media services	• **Risk factor:** Social media services • **Key kinds of illegal harm*:** Your service is likely to have an increased risk of nearly all kinds of illegal harm. Many social media services are designed to maximise engagement between users. If your service is a social media service, you should consider how potential perpetrators may exploit this design for illegal purposes. For example, potential perpetrators may exploit the likelihood of virality to share illegal content with very large groups of people. Social media services can also be used by potential perpetrators of grooming to target young users by sending out many messages. These services are also used in large-scale foreign interference campaigns to spread disinformation. Research shows that social media services can increase the risk of nearly all kinds of illegal harm, except for firearms and other weapons offences where we do not currently have evidence. This may be due to more research on social media services, or greater probability of risk due to the wide range of functionalities and features on many social media services.

ソーシャルメディアサービスが持つリスク要因の説明の中で、偽情報を拡散するキャンペーンとしてソーシャルメディアサービスが利用される可能性についても言及している

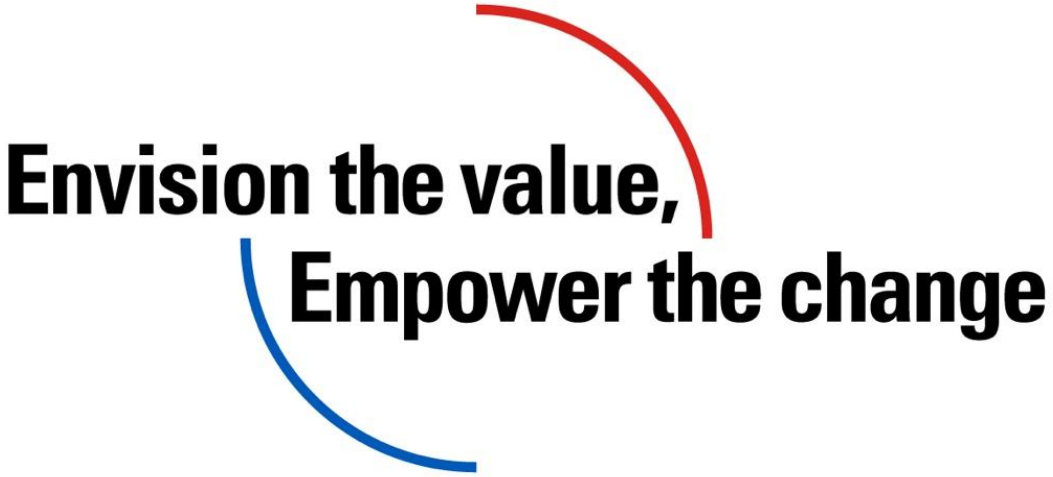
出所：Protecting people from illegal harms online

参考：一般的なリスク要因への評価の概要

③一般的なリスク要因におけるリスクプロファイルの概要

項目	ユーザ間サービス	検索サービス
ユーザー構成	• ユーザーの性別や年齢がリスク評価に影響を与えること（特に女性・女児はリスクが高くなること） • 人種、宗教、年齢など複数のユーザの特徴を踏まえてリスク評価を行うこと	• 検索サービスは幅広いユーザ属性によって利用されており、あらゆる違法危害のリスクに影響を与える可能性があること • 特に複数の保護すべき属性をもつユーザは違法コンテンツによる危害を経験する可能性が高いこと
ビジネスモデル	• 収益モデルによって違法な危害のリスクを高める危険性があること • 例えば、エンゲージメントを高めるための設計が違法コンテンツへの関与を促す可能性があること • 広告ターゲティングなどによって有害な活動を助長する環境を作り出す可能性があること	• 一般的な検索サービスは広告が表示される可能性があり、ユーザが違法行為に及ぶ可能性のある商品や情報に触れる可能性があること • サービス設計がリスクにどのように影響するかを評価することを期待するとしている
サービス・ビジネスの成熟度	• アーリーステージのビジネスは技術的なスキルや財務的なリソースが限定的であるため、違法危害の可能性が高まること • 利用者数が急成長する場合には、リスク源が変化しうること	

出所：Protecting people from illegal harms online


Envision the value,
Empower the change