資料22－1－2

プラットフォーム事業者ヒアリングの結果（案）

【凡例】
「全ての」：全ての事業者が該当する場合を指す。
「ほぼ全ての」：一社以外の事業者が該当する場合を指す。
「多くの」：「ほぼ全ての」に該当しないが、半数を超える事業者が該当する場合を指す。
「半数の」：半数の事業者の場合を指す。
「一部の」：半数に満たない場合の事業者を指す。

※黄色ハイライト部分は前回（第 19 回会合）からの更新部分

項目	結果（○：質問に対する一定の回答があったもの　●：質問に対して回答が不十分なもの）
１　ヒアリング対象サービスの規模	○　半数の国内事業者及び一部の海外に本社を置く事業者（以下「国外事業者」）は、**日本国内における最近のＭＡＵ（月間アクティブユーザー数（１か月間に対象サービスに１度でもアクセスした利用者の数））**を回答。 ●　一部の国内事業者及び多くの国外事業者からは、**日本国内における最近のＭＡＵの回答なし**。その理由として、**非公表であること、対象サービスの利用にログインが不要であることや、回答しない理由自体が不明**。 ●　全ての国内事業者及び国外事業者**（ただし、対象サービスのうち、利用者登録が不要なものがない事業者を除く）**は、**日本国内における最近の月間合計投稿数の回答なし**。 【日本国内の最近の月間アクティブユーザー数の回答状況】（参考資料22-1-1 及び22-1-2　Q1-1、Q2－1、Q2－4）
２　偽・誤情報の流通・拡散への対応方針	○　一部の国内事業者及びほぼ全ての国外事業者は、**偽・誤情報と重複する場合がある偽アカウント・なりすまし・合成コンテンツ・欺瞞行為・詐欺、選挙・市民活動や医療等の分野とは別**に、**コンテンツモデレーション等**（削除、アカウント停止、表示順位の降格、収益不能化、アクセス不能化、警告表示・ラベリング、ファクトチェック結果の表示、投稿時の注記の義務付け等。以下同じ。）**による対応の対象となる偽・誤情報の範囲（定義）や類型（例示等）**について、**個別に策定**し**日本語で公表**している**ポリシー、ガイドライン等（以下「偽・誤情報ポリシー等」）**を回答。 **【偽・誤情報の範囲（定義）や類型（例示等）［例］】** ■「Yahoo!ニュースコメントポリシー」 https://news.yahoo.co.jp/info/comment-policy ・　健康被害等をもたらす可能性のある偽情報であって、ファクトチェックにより反真実であることが明らかな投稿（明らかな偽情報）（「新型コロナウイルスのワクチンを接種すると、流産する。不妊になる。」、「ワクチン接種された実験用の動物が全て死亡した。」、「ワクチンを接種することでコロナウイルスに感染する。」など）

	■Yahoo!知恵袋「利用のルール」 https://chiebukuro.yahoo.co.jp/topic/guide/rule/ ・ 明らかに事実と異なり社会的に混乱を招く恐れのある投稿(明らかな偽情報)(「(そのような事実がないにもかかわらず)昨日、○○(地名)で大地震があったけど、、、」、「コロナワクチン殺人計画は本当ですか？接種したネズミは 3 年以内に死亡したそうです。」、「トランプ大統領、コロナで亡くなったみたいですね。やはり突然の重症化、恐ろしいウイルスです」など) ■ファイナンス掲示版「ヘルプページ」https://support.yahoo-net.jp/SccFinance/s/article/H000011273 ・ 明らかな偽情報として、架空の出来事のでっちあげや虚偽の内容の投稿(風説の流布に該当する場合違法行為となる可能性)(「○○会社が製造したワクチンを接種された実験用の動物が全て死亡した。」など) ■LINE オープンチャット「安心・安全ガイドライン」https://openchat-jp.line.me/other/guideline ・ 真偽不明の情報の拡散(新型コロナウイルス関連:「コロナワクチンによって不妊になる」、「コロナは人口削減のために人工的に作られた」など、災害関連:「能登半島地震は人工地震である」など) ■「LINE VOOM コミュニティガイドライン」 https://terms.line.me/line_voom_community_guideline?lang=ja&country=JP ・ 誤情報の拡散:当社または第三者になりすまし行為や、フェイクニュースなど虚偽の情報、身体に影響を及ぼす食品、医療、医療薬などの虚偽の情報を意図的に発信・拡散・流布させる行為(新型コロナウイルス関連「コロナワクチンによって不妊になる」、「コロナは人口削減のために人工的に作られた」など、災害関連「能登半島地震は人工地震である」など) ■YouTube「誤った情報に関するポリシー」https://support.google.com/youtube/answer/10834785?hl=ja ・ 特定の種類の誤解を招くコンテンツまたは虚偽が含まれるコンテンツで、重大な危害を及ぼす可能性のあるもの(現実の世界で危害を与える可能性がある特定の種類の誤った情報、技術的に操作された特定の種類のコンテンツ、民主的な手続きを妨害するコンテンツが含まれる) ・ 国勢調査の妨害: 国勢調査に関する時間、場所、方法、資格要件について参加者を誤解させることを目的としたコンテンツ、または国勢調査を著しく妨げる可能性のある虚偽の主張(国勢調査の参加方法に関して誤った手順を示す。回答者の在留資格が法執行機関に報告されるという誤った主張により、国勢調査への参加を妨げる行為。) ・ 改ざんされたコンテンツ: ユーザーの誤解を招くように技術的に操作または改ざんされ(前後関係を無視してクリ

	ップを切り抜く以上の操作が多い)、重大な危害を及ぼす可能性のあるコンテンツ(地政学的緊張を悪化させ、重大な危害を及ぼす可能性のある不正確に翻訳された動画の字幕。政府関係者の死を装うために技術的に操作(前後関係を無視してクリップを切り抜く以上の操作が多い)された動画。重大な危害を及ぼす可能性のある事件を捏造するために技術的に操作(前後関係を無視してクリップを切り抜く以上の操作が多い)された動画。) ・　虚偽のコンテンツ: 過去の事象の古い映像を現在の事象のものであると虚偽の主張をすることで、重大な危害を及ぼす可能性のあるコンテンツ(実際は別の地域や事象に関するコンテンツを、特定の地域の人権侵害の記録として不正確に提示したコンテンツ。実際は数年前の映像であるにもかかわらず、そのコンテンツが現在の事象のものであるという虚偽の主張とともに、抗議活動への参加者に対する軍事的弾圧を示したコンテンツ。) **■Meta「偽情報」ポリシー** https://transparency.meta.com/ja-jp/policies/community-standards/misinformation/ ・　実際の危害や暴力:弊社は、人々に対する差し迫った暴力または実際の危害のリスクに直接つながる可能性が高いと、専門家のパートナーが判断した偽情報および検証できない噂を削除します。偽情報とは、信頼できる第三者が虚偽であると判断する主張を含むコンテンツと定義されます。検証できない噂とは、専門家のパートナーによる情報元の特定が極めて困難であるか不可能な主張、信頼できる提供元がない主張、その内容を証明するための具体性が不十分な主張、またはその内容があまりに信じがたい、もしくは不合理で信用できない主張と定義されます。弊社は、一見無害に思われる偽情報でも、特定の文脈では、死や深刻な怪我といった実際の危害の高いリスクにつながる可能性がある暴力的脅威など、オフラインでの危害のリスクにつながる場合があることを認識しています。弊社は、こうした各地域の動向について専門知識を有する非政府組織(NGO)、非営利団体、人道支援組織、国際機関のグローバルネットワークと連携しています。社会的暴力のリスクが高まっている国では、どの虚偽の主張が差し迫った身体的危害のリスクに直接つながるかを理解するために、現地のパートナーと積極的に協力しています。その上で、弊社のプラットフォームでそのような主張をするコンテンツを特定し削除します。例えば現地の専門家と協議して、文脈から切り離して虚偽の主張をし、暴力行為、暴力の被害者や加害者、武器や軍事用具を描写するメディアを削除する可能性があります。) ・　有害な健康関連の偽情報:弊社では、公衆衛生と安全に対する差し迫った危害に直接的な害をもたらす可能性の高い健康誤情報を特定するために、主要な保健機関と協議しています。弊社が削除する有害な健康関連の偽情報には次のようなものがあります。ワクチンに関する偽情報(弊社は、主にワクチンに関する偽情報について、保健当局がその情報は虚偽であり、差し迫ったワクチン接種の拒否を促進する可能性が高いと判断した場合

	は、その情報を削除します。）、公衆衛生上の緊急事態の間の偽情報（偽情報については、公衆衛生当局が、その情報が虚偽であり、差し迫った身体的危害のリスクに直接つながる可能性が高いと判断した場合、公衆衛生上の緊急事態の間、弊社はその情報を削除します。これには、個人が有害な疾病に感染したり、それを拡散させたりするリスクや、関連するワクチンを拒否するリスクにつながるものが含まれます。弊社では、世界と地域の保険機関と連携して公衆衛生上の緊急事態を特定しています。現在対象となるものとしては、ウイルスに関する緊急事態宣言が引き続き公に発表されている国における、新型コロナウイルス感染症関連の特定の虚偽の主張などがあります。新型コロナウイルス感染症およびワクチンについて、弊社が許可していない偽情報の種類をまとめたルール一覧についてはこちらをクリックしてください。）、健康上の問題に対する有害な「奇跡的な治療法」の宣伝または擁護（これには、医療の観点から推奨される使用法が、重傷または死亡のリスクに直接つながる可能性が高く、正当な医療上の使用法がない治療が含まれます(例: 漂白剤、消毒剤、黒色軟膏、苛性ソーダ)。） ・ 投票者または国勢調査への干渉：選挙や国勢調査の健全性を促進する取り組みとして、弊社は、このようなプロセスに人々が参加する能力を妨害するリスクに直接つながる可能性が高い偽情報を削除します。これには次のようなものが含まれます。投票や有権者登録、国勢調査への参加に関する日付、場所、時間、方法に関する偽情報、投票できる人、投票資格、投票の有効性、投票するために提供しなければならない情報や書類に関する偽情報、候補者が立候補するか否かに関する偽情報、国勢調査への参加資格や、国勢調査へ参加するために提供しなければならない情報や書類に関する偽情報、個人の国勢調査の情報が別の(国勢調査の実施を担当していない)政府機関と共有されると述べること(ただし該当する場合)など、政府の国勢調査への関与についての偽情報、米国移民関税執行局(ICE)が投票所に待機しているとの偽りのまたは根拠のない主張、投票プロセスに参加すると新型コロナウイルス感染症(または別の伝染病)に感染するとする明確な偽の主張、選挙当局によって検証された、米国の投票所の現状に関して投票ができなくなるような偽りの主張。弊社では、暴力の呼びかけ、違法な参加の促進、選挙への協調的干渉の呼びかけを対象とする追加ポリシーを設けており、これらは弊社のコミュニティ規定の他のセクションに記載されています。 ・ 加工されたメディア：メディアはさまざまな方法で編集することができます。多くの場合、このような変更は、芸術的な理由による切り取りもしくは短縮、または音楽の追加など、無害なものです。ただ、特に動画コンテンツの場合、中には加工の有無がはっきりせず誤解を与えてしまう可能性が生じることもあります。このようなコンテンツはすぐに拡散されやすく、専門家の助言によれば、加工されたメディアに関する誤った思い込みはさらなる議論で

	修正できないことが多々あるため、弊社ではこのようなコンテンツを削除します。次の一定の条件を満たす動画については、本ポリシーに基づき削除します。(1)明瞭さや画質・音質の調整にとどまらず、動画の被写体が実際には発言していない言葉を言ったと一般の利用者に誤解させるような編集または合成が、一般の利用者にはわからない形で行われている動画、および(2)ディープラーニング技術を含む人工知能または機械学習(例: 人為的なディープフェイク)によって制作され、特定の動画に他のコンテンツを統合、結合、置換、重ね合わせるなどして動画が本物に見えるように作成されている動画。 ・ 偽・誤情報ポリシー等に違反していないにもかかわらず、プラットフォームの信頼性と完全性(authenticity and integrity)を損なう誤情報については、独立した第三者ファクトチェック・パートナーのネットワークと協力して、誤情報の配信を減らし、強力な警告ラベルを表示し、誤情報に出くわした人、それを共有しようとした人、またはすでに共有した人に通知するというアプローチをとっている。1 つのファクトチェックに基づき、誤りであることを暴いたストーリーと重複するものを特定する類似性検出方法を発動し、特定されたポストに対して、フィード上の流通を減らし、警告ラベルを表示し、利用者に通知するという同じペナルティを適用することができる。 **■TikTok「コミュニティガイドライン」** https://www.tiktok.com/community-guidelines/ja-jp/integrity-authenticity/?enter_method=left_navigation#1 ・ 意図にかかわらず、個人や社会に重大な危害を及ぼし得る不正確な、誤解を招く、または虚偽のコンテンツ(重大な危害には、身体的、心理的、または社会的危害、および物的損害が含まれる。商業上の損害や風評被害はこれには含まれず、単なる不正確な情報や作り話も対象外) ・ 公共の安全に危険を及ぼしたり、危機または緊急事態についてパニックを引き起こし得たりする誤情報(以前に行われた攻撃の過去の映像を現在進行中であるかのように使用することや、特定の場所で基本的な生活必需品(食料や水など)が在庫切れを起こしていると誤った主張をすることなど) ・ 誤った医療情報(ワクチンに関する誤解を招く発言、生命を脅かす疾患に対して適切な治療を受けることを思いとどまらせる不正確な医療上の助言、公衆衛生に危険を及ぼすその他の誤情報など) ・ 確立した科学的コンセンサスを阻害する気候変動に関する誤情報(気候変動やその一因となっている要素の存在を否定することなど) ・ 暴力的またはヘイトに満ちた危険な陰謀論(暴力的な行動を呼びかける、過去の暴力行為と関連付ける、十分に立証されている暴力事件を否定する、保護属性を持つグループに対する偏見を引き起こすなど)

	・　個人を名指しして攻撃する、特定の陰謀論 ・　現実世界の出来事について人に誤解を与えるようなやり方で編集、接合、または合成された素材（動画や音声など） ・　根拠がなく、特定の出来事や事態が「政府」や「秘密結社」などの秘密の集団や強力な集団によって引き起こされたとする一般的な陰謀論 ・　詳細がまだ明らかになっていない緊急事態や展開中の出来事に関連する未確認情報 ・　ファクトチェックの審査中の、危険性が高い可能性のある誤情報 **■LinkedIn「プロフェッショナルコミュニティポリシー」** https://jp.linkedin.com/legal/professional-community-policies? ・　虚偽または誤解を招く内容。事実として提示された特定の主張が、明らかに虚偽であるか実質的に誤解を招くものであり、害を及ぼす恐れがある場合は、これを削除。虚偽であるか実質的に誤解を招くものではあるものの、害を及ぼす恐れがないコンテンツの場合、投稿者のネットワーク外での配信は認められない。 **■LinkedIn ヘルプセンター「虚偽または誤解を招くコンテンツ」** https://www.linkedin.com/help/linkedin/answer/a1340752/ ・　まもなく実施されるか終わったばかりの選挙の投票の時間、場所、手段、資格要件に関する虚偽の、または実質的に誤解を招く情報が含まれたコンテンツ ・　被害を引き起こす恐れのある脅威、暴力、危険についての根拠のない主張など、緊急時パニックを誘発したり、安全対策を講じる意欲を失わせたりするような主張（例:「森林火災が発生している地域で略奪者が横行している」など） ・　特定の場所における人権侵害または軍事紛争の証拠として提示された、実際には別の場所、出来事、期間の、不正確なコンテンツ ・　実際の出来事を歪曲した、対象者、他の個人やグループ、または社会全体に害を及ぼす恐れのある、加工された画像や動画などの合成または操作されたメディア ・　有害な治療法や奇跡の治療法を宣伝するコンテンツ、または専門的な医学的アドバイスを求めたり、聞き入れたりすることを妨げるコンテンツ ・　地域の保健当局や世界保健機関（WHO）の医療ガイダンスと相反する主張や記述

	・　COVID-19 の治療、予防、変異株、感染に関する医療上の誤情報（医療用、非医療用を問わず、マスクは身体に悪影響を及ぼす、イベルメクチンまたはヒドロキシクロロキンは COVID-19 の治療または予防に有効である、承認された COVID-19 のワクチンは、COVID-19 を含む感染症による死亡、不妊、流産、自閉症、または筋収縮を引き起こす恐れがある、承認された COVID-19 のワクチンは、追跡/監視装置を埋め込んだり、患者に磁気を発生させたりする、COVID-19 のワクチン接種者は、ワクチン未接種者に比べ、ウイルスを拡散させる可能性が高い、COVID-19 は、特定の世界的な指導者、公人、または世界各地の保健当局からの資金提供または支援を受けて開発された） ・　COVID-19 の有病率または重症度に関する医学的誤報（COVID-19 は存在せず、デマである、ウイルスとその変異株が根絶され、パンデミックは終息した、COVID-19 の症状、死亡率、感染率は季節性インフルエンザよりも深刻なものではない、COVID-19 に感染して、死亡または重症化した人はいない） **○　一部の国内事業者及び全ての国外事業者は、<u>偽・誤情報ポリシー等に定める禁止事項に違反した場合</u>や<u>同ポリシー等違反でない場合</u>における偽・誤情報への対応として、<u>モデレーション等の具体的な対応方</u>法を回答。** **【モデレーション等［例］】** ・　投稿の削除 ・　アカウントの停止・削除 ・　再取得したアカウントにおける投稿制限 ・　サービスの利用停止 ・　他のサービスや機能へのアクセス制限 ・　プロフィールの編集要請 ・　第三者が閲覧又はアクセスできない状態化 ・　表示の降格 ・　検索結果の上位表示の回避 ・　おすすめやトレンド等の対象外化 ・　投稿時の注意喚起や警告ラベルの表示

	・ 閲覧を望まない第三者への非表示 ・ 第三者に対する変化の激しい出来事ガイド、信憑性未確認ラベルやファクトチェック結果の表示 **○ 一部の国内事業者及びほぼ全ての国外事業者は、<u>偽・誤情報に対するコンテンツモデレーション等についての考え方</u>や、コンテンツモデレーション等のうち<u>投稿の削除の対象となる偽・誤情報についての例示等</u>を回答。** **【考え方［例］】** ・ 偽情報である旨が明らかである投稿を禁止。 ・ 特定の種類の誤解を招くコンテンツまたは虚偽が含まれるコンテンツで、重大な危害を及ぼす可能性のあるものは許可されない。これには、現実の世界で危害を与える可能性がある特定の種類の誤った情報、技術的に操作された特定の種類のコンテンツ、民主的な手続きを妨害するコンテンツが含まれる。 ・ 偽情報は、包括的な禁止事項を明示する方法がなく、他の種類の発言とは異なる。例えば、過度な暴力描写やヘイトスピーチついて禁じている発言はポリシーに規定されているため、そのポリシーに賛成しない人でも従うことができるが、偽情報については、そのような方針を提供することができず、世界は絶え間なく変化し続けているため、ある時点では真実であっても、次の瞬間には真実でなくなることがあり、また、自身の周りの世界について異なるレベルの情報を有しているため、真実でない情報も真実だと信じてしまうことがある。ポリシーとして、偽情報についてさまざまなカテゴリーを明確にし、対象となる発言を見つけたときの対処法を示した明確なガイダンスを設けるよう努めている。 ・ グローバルなコミュニティでは人々がさまざまな意見を持つのは自然なことであるが、事実や現実の共有に基づいて行動するべく努めているため、個人や社会に重大な危害を及ぼし得る不正確な、誤解を招く、または虚偽のコンテンツは、意図にかかわらず許可しない。重大な危害には、身体的、心理的、または社会的危害、および物的損害が含まれ、商業上の損害や風評被害はこれには含まれず、単なる不正確な情報や作り話も対象外となる。 **【投稿の削除の対象に関する例示等［例］】** ・ 国勢調査の妨害: 国勢調査に関する時間、場所、方法、資格要件について参加者を誤解させることを目的としたコンテンツ、または国勢調査を著しく妨げる可能性のある虚偽の主張（国勢調査の参加方法に関して誤った手順を示す。回答者の在留資格が法執行機関に報告されるという誤った主張により、国勢調査への参加を妨げる

	行為。） ・　改ざんされたコンテンツ: ユーザーの誤解を招くように技術的に操作または改ざんされ（前後関係を無視してクリップを切り抜く以上の操作が多い）、重大な危害を及ぼす可能性のあるコンテンツ（地政学的緊張を悪化させ、重大な危害を及ぼす可能性のある不正確に翻訳された動画の字幕。政府関係者の死を装うために技術的に操作（前後関係を無視してクリップを切り抜く以上の操作が多い）された動画。重大な危害を及ぼす可能性のある事件を捏造するために技術的に操作（前後関係を無視してクリップを切り抜く以上の操作が多い）された動画。） ・　虚偽のコンテンツ: 過去の事象の古い映像を現在の事象のものであると虚偽の主張をすることで、重大な危害を及ぼす可能性のあるコンテンツ（実際は別の地域や事象に関するコンテンツを、特定の地域の人権侵害の記録として不正確に提示したコンテンツ。実際は数年前の映像であるにもかかわらず、そのコンテンツが現在の事象のものであるという虚偽の主張とともに、抗議活動への参加者に対する軍事的弾圧を示したコンテンツ。） ・　実際の危害や暴力：弊社は、人々に対する差し迫った暴力または実際の危害のリスクに直接つながる可能性が高いと、専門家のパートナーが判断した偽情報および検証できない噂を削除します。偽情報とは、信頼できる第三者が虚偽であると判断する主張を含むコンテンツと定義されます。検証できない噂とは、専門家のパートナーによる情報元の特定が極めて困難であるか不可能な主張、信頼できる提供元がない主張、その内容を証明するための具体性が不十分な主張、またはその内容があまりに信じがたい、もしくは不合理で信用できない主張と定義されます。弊社は、一見無害に思われる偽情報でも、特定の文脈では、死や深刻な怪我といった実際の危害の高いリスクにつながる可能性がある暴力的脅威など、オフラインでの危害のリスクにつながる場合があることを認識しています。弊社は、こうした各地域の動向について専門知識を有する非政府組織(NGO)、非営利団体、人道支援組織、国際機関のグローバルネットワークと連携しています。社会的暴力のリスクが高まっている国では、どの虚偽の主張が差し迫った身体的危害のリスクに直接つながるかを理解するために、現地のパートナーと積極的に協力しています。その上で、弊社のプラットフォームでそのような主張をするコンテンツを特定し削除します。例えば現地の専門家と協議して、文脈から切り離して虚偽の主張をし、暴力行為、暴力の被害者や加害者、武器や軍事用具を描写するメディアを削除する可能性があります。） ・　有害な健康関連の偽情報：弊社では、公衆衛生と安全に対する差し迫った危害に直接的な害をもたらす可能性の高い健康誤情報を特定するために、主要な保健機関と協議しています。弊社が削除する有害な健康関連の偽情報には次のようなものがあります。ワクチンに関する偽情報（弊社は、主にワクチンに関する偽情報につい

	て、保健当局がその情報は虚偽であり、差し迫ったワクチン接種の拒否を促進する可能性が高いと判断した場合は、その情報を削除します。)、公衆衛生上の緊急事態の間の偽情報(偽情報については、公衆衛生当局が、その情報が虚偽であり、差し迫った身体的危害のリスクに直接つながる可能性が高いと判断した場合、公衆衛生上の緊急事態の間、弊社はその情報を削除します。これには、個人が有害な疾病に感染したり、それを拡散させたりするリスクや、関連するワクチンを拒否するリスクにつながるものが含まれます。弊社では、世界と地域の保険機関と連携して公衆衛生上の緊急事態を特定しています。現在対象となるものとしては、ウイルスに関する緊急事態宣言が引き続き公に発表されている国における、新型コロナウイルス感染症関連の特定の虚偽の主張などがあります。新型コロナウイルス感染症およびワクチンについて、弊社が許可していない偽情報の種類をまとめたルール一覧についてはこちらをクリックしてください。)、健康上の問題に対する有害な「奇跡的な治療法」の宣伝または擁護(これには、医療の観点から推奨される使用法が、重傷または死亡のリスクに直接つながる可能性が高く、正当な医療上の使用法がない治療が含まれます(例: 漂白剤、消毒剤、黒色軟膏、苛性ソーダ)。) ・ 投票者または国勢調査への干渉:選挙や国勢調査の健全性を促進する取り組みとして、弊社は、このようなプロセスに人々が参加する能力を妨害するリスクに直接つながる可能性が高い偽情報を削除します。これには次のようなものが含まれます。投票や有権者登録、国勢調査への参加に関する日付、場所、時間、方法に関する偽情報、投票できる人、投票資格、投票の有効性、投票するために提供しなければならない情報や書類に関する偽情報、候補者が立候補するか否かに関する偽情報、国勢調査への参加資格や、国勢調査へ参加するために提供しなければならない情報や書類に関する偽情報、個人の国勢調査の情報が別の(国勢調査の実施を担当していない)政府機関と共有されると述べること(ただし該当する場合)など、政府の国勢調査への関与についての偽情報、米国移民関税執行局(ICE)が投票所に待機しているとの偽りのまたは根拠のない主張、投票プロセスに参加すると新型コロナウイルス感染症(または別の伝染病)に感染するとする明確な偽の主張、選挙当局によって検証された、米国の投票所の現状に関して投票ができなくなるような偽りの主張。弊社では、暴力の呼びかけ、違法な参加の促進、選挙への協調的干渉の呼びかけを対象とする追加ポリシーを設けており、これらは弊社のコミュニティ規定の他のセクションに記載されています。 ・ 加工されたメディア:メディアはさまざまな方法で編集することができます。多くの場合、このような変更は、芸術的な理由による切り取りもしくは短縮、または音楽の追加など、無害なものです。ただ、特に動画コンテンツの場合、中には加工の有無がはっきりせず誤解を与えてしまう可能性が生じることもあります。このようなコンテンツは

	すぐに拡散されやすく、専門家の助言によれば、加工されたメディアに関する誤った思い込みはさらなる議論で修正できないことが多々あるため、弊社ではこのようなコンテンツを削除します。次の一定の条件を満たす動画については、本ポリシーに基づき削除します。(1)明瞭さや画質・音質の調整にとどまらず、動画の被写体が実際には発言していない言葉を言ったと一般の利用者に誤解させるような編集または合成が、一般の利用者にはわからない形で行われている動画、および(2)ディープラーニング技術を含む人工知能または機械学習(例: 人為的なディープフェイク)によって制作され、特定の動画に他のコンテンツを統合、結合、置換、重ね合わせるなどして動画が本物に見えるように作成されている動画。 ・ 公共の安全に危険を及ぼしたり、危機または緊急事態についてパニックを引き起こし得たりする誤情報（以前に行われた攻撃の過去の映像を現在進行中であるかのように使用することや、特定の場所で基本的な生活必需品（食料や水など）が在庫切れを起こしていると誤った主張をすることなど） ・ 誤った医療情報（ワクチンに関する誤解を招く発言、生命を脅かす疾患に対して適切な治療を受けることを思いとどまらせる不正確な医療上の助言、公衆衛生に危険を及ぼすその他の誤情報など） ・ 確立した科学的コンセンサスを阻害する気候変動に関する誤情報（気候変動やその一因となっている要素の存在を否定することなど） ・ 暴力的またはヘイトに満ちた危険な陰謀論（暴力的な行動を呼びかける、過去の暴力行為と関連付ける、十分に立証されている暴力事件を否定する、保護属性を持つグループに対する偏見を引き起こすなど） ・ 個人を名指しして攻撃する、特定の陰謀論 ・ 現実世界の出来事について人に誤解を与えるようなやり方で編集、接合、または合成された素材（動画や音声など）。 ・ まもなく実施されるか終わったばかりの選挙の投票の時間、場所、手段、資格要件に関する虚偽の、または実質的に誤解を招く情報が含まれたコンテンツ ・ 被害を引き起こす恐れのある脅威、暴力、危険についての根拠のない主張など、緊急時パニックを誘発したり、安全対策を講じる意欲を失わせたりするような主張（例:「森林火災が発生している地域で略奪者が横行している」など） ・ 特定の場所における人権侵害または軍事紛争の証拠として提示された、実際には別の場所、出来事、期間の、不正確なコンテンツ

	・　実際の出来事を歪曲した、対象者、他の個人やグループ、または社会全体に害を及ぼす恐れのある、加工された画像や動画などの合成または操作されたメディア ・　有害な治療法や奇跡の治療法を宣伝するコンテンツ、または専門的な医学的アドバイスを求めたり、聞き入れたりすることを妨げるコンテンツ ・　地域の保健当局や世界保健機関（WHO）の医療ガイダンスと相反する主張や記述 ・　COVID-19 の治療、予防、変異株、感染に関する医療上の誤情報（医療用、非医療用を問わず、マスクは身体に悪影響を及ぼす、イベルメクチンまたはヒドロキシクロロキンは COVID-19 の治療または予防に有効である、承認された COVID-19 のワクチンは、COVID-19 を含む感染症による死亡、不妊、流産、自閉症、または筋収縮を引き起こす恐れがある、承認された COVID-19 のワクチンは、追跡/監視装置を埋め込んだり、患者に磁気を発生させたりする、COVID-19 のワクチン接種者は、ワクチン未接種者に比べ、ウイルスを拡散させる可能性が高い、COVID-19 は、特定の世界的な指導者、公人、または世界各地の保健当局からの資金提供または支援を受けて開発された） ・　COVID-19 の有病率または重症度に関する医学的誤報（COVID-19 は存在せず、デマである、ウイルスとその変異株が根絶され、パンデミックは終息した、COVID-19 の症状、死亡率、感染率は季節性インフルエンザよりも深刻なものではない、COVID-19 に感染して、死亡または重症化した人はいない） ● ほぼ全ての国内事業者は、**偽・誤情報ポリシー等を個別に策定せず、偽・誤情報に対するコンテンツモデレーション等を行っていない**旨、多くがインプレッションや収益目的で投稿内容がその時々の大衆の興味に合わせたものになるという理由で**偽・誤情報について内容による類型定義を行なっていない旨**、**類型に応じて対応方法を変えずに投稿や行為の悪質性・公益性・投稿者のサービス利用状況に応じて偽・誤情報への対応を決定している旨**を回答。 ● ほぼ全ての国内事業者及び国外事業者は、**偽・誤情報ポリシー等違反におけるコンテンツモデレーション等の類型ごとの具体的な対応方法**について、コンテンツモデレーション等のうち**投稿の削除の対象となる偽・誤情報の例示や同ポリシー等違反を繰り返す場合のアカウント停止等に関する基準等を除き**、**どのような場合にどのモデレーション等の対象となるかや、一部の事業者においては公共の利益等の観点から例外と**

	してコンテンツモデレーション等の対象とならない場合等に関する具体的な考え方や基準に関する回答なし。 ●　一部の国外事業者は、**偽・誤情報が同ポリシー等に定める禁止事項ではない場合におけるコンテンツモデレーション等**について、**おすすめの対象外化**等の対象となる偽・誤情報の例示、自らのプラットフォームの信頼性と完全性を損なう場合や信憑性の低い場合における**第三者のファクトチェック機関との連携**を除き、**どのような場合にどのモデレーション等の対象となるか等に関する具体的な考え方や基準に関する回答なし**。また、**偽・誤情報ポリシー等がグローバルなものであり、言語や地域等を問わずに施行される旨**を回答。 【偽・誤情報ポリシー等の策定・公表状況等】(参考資料22-1-1 及び22-1-2Q3－1、Q3－2) ○　一部の国内事業者及び国外事業者は、偽・誤情報ポリシー等について、**定期的な見直しと第三者によるレビューの両方**の実施、また、ほぼ全ての国内事業者及び国外事業者は、**第三者機関によるレビューの実施**を回答。 ●　ほぼ全ての国内事業者及び国外事業者は、**ポリシー等の定期的な見直しに関する回答なし**。また、一部の国外事業者は、**偽・誤情報ポリシー等がグローバルなものであり、言語や地域等を問わずに施行される旨**を回答。 【偽・誤情報ポリシー等の見直し・外部有識者等によるレビュー状況】(参考資料22-1-1 及び22-1-2　Q3－3)
3　偽・誤情報の流通・拡散に対するモデレーション等の手続・体制	○　全ての国内事業者及び国外事業者は、**法令違反やポリシー等に違反するコンテンツ**について、権利を侵害されている者及び発信者（投稿者）以外の**第三者からの日本語による通報を受け付ける窓口（以下「第三者通報受付窓口」）の設置**を回答。 ●　一部の国内事業者及び多くの国外事業者は、**第三者通報受付**にあたり、、**サービスの ID やアカウントを取得しているユーザやサービスにログイン可能なユーザ等通報可能な主体に限定**がある旨を回答。 【日本語による第三者通報受付窓口の設置状況】(参考資料22-1-1 及び22-1-2　Q5－1) ○　一部の国内事業者及び国外事業者は、**第三者通報受付窓口**において、**公的機関によるファクトチェック済みの情報を添える**等により、**偽・誤情報ポリシー等に違反する偽・誤情報を選択して通報が可能**であることを回答。 ●　一部の国内事業者及び多くの国外事業者は、**第三者通報受付窓口について、偽・誤情報ポリシー等に違反**

	する偽・誤情報を選択して通報可能であるかどうかに関する明確な回答なし。 【日本語による第三者通報受付窓口の設置状況】(参考資料22-1-1 及び22-1-2 Q5－1) ○ 全ての国内事業者は、第三者通報受付窓口について、**日本語通報への対応が可能な人数**を回答。 ● 全ての国外事業者からは、**第三者通報受付窓口における日本語通報に対応可能な人数について、随時変動しうることやセキュリティ及びビジネス上の理由から非公表とする等、明確な回答なし**。 【通報受付窓口において、日本語通報対応が可能な人数】(参考資料22-1-1 及び22-1-2 Q5－2－(1)) ○ 全ての国内事業者及び一部の国外事業者は、**第三者通報受付窓口における日本語通報対応**について、**AI その他の機械的手段は利用していない**と回答。 ● 一部の国外事業者は、**第三者通報受付窓口における日本語通報対応における AI その他の機械的手段の利用について、当該手段の概要や利用手順（どのようなケースで用いるのか等）に関する明確な回答なし**。 【日本語通報対応における AI 等の利用状況】(参考資料22-1-1 及び22-1-2 Q5－2－(3)) ○ 全ての国内事業者及び一部の国外事業者は、**日本語通報対応結果を通報者に通知等**していると回答。 ○ 半数の国内事業者及び一部の国外事業者は、**通報者からの不服申立等に対応**していると回答。 ● 多くの国外事業者は、**日本語通報対応結果を通報者に通知等しているかどうかに関する明確な回答なし**。 ● 半数の国内事業者及び多くの国外事業者は、**通報者からの不服申立等にどのように対応しているかどうかに関する明確な回答なし**。 【日本語通報対応結果の第三者通報者への通知等状況】(出典:参考資料22-1-1 及び22-1-2 Q5－2－(6)) 【日本語通報対応結果に関する第三者通報者からの不服申立等の専用窓口等の状況】(参考資料22-1-1 及び22-1-2 Q5－2－(7)) ○ ほぼ全ての国内事業者は、**第三者通報受付窓口を通じた日本語による通報をどの程度の期間で処理しているか**について、**２４時間以内や２営業日以内等の目標期間**を回答。 ● 全ての国外事業者は、処理の目標期間について、報**告される問題やトピックの複雑さによって異なることや非公表とする等、明確な回答なし**。 【日本語による通報への処理期間の目標の設定状況】(参考資料22-1-1 及び22-1-2 Q5－2－(5))

	○　一部の国内事業者及び多くの国外事業者は、**政府機関（法務省、警察、違法・有害情報相談センター（総務省）等）やNGO/NPO（自殺予防に取り組む日本の非営利団体OVA等）など特定の第三者通報主体からの通報の優先的取扱いの実施**を回答。なお、一部の国外事業者は、特定の第三者通報主体からの通報の優先的取扱いの効果について、深い知識を各分野に渡って持つパートナーでありシステムを補完する重要な存在であることや日本固有の状況について特に審査における判断が困難なケースについても、信頼できる第三者の知見を反映した審査を実施できることを回答。 ●　ほぼ全ての国内事業者及び一部の国外事業者は、**政府機関やNGO/NPOなど特定の第三者通報主体からの通報の優先的取扱いは実施していないことを回答**。 ●　全ての国内事業者及び国外事業者は、偽・誤情報ポリシー等に違反する偽・誤情報について、**特定の第三者通報主体からの通報の優先的取扱いを実施しているかどうかに関する回答なし**。 【特定の第三者通報主体からの日本語による通報の優先的取り扱い】(出典:参考資料22-1-1 及び22-1-2 Q5－2－(8)、同22-5-2 Q3-6)【自国内の偽・誤情報と比較して判断が難しいことについての工夫の状況】(参考資料22-5-2 Q3－6) ○　一部の国内事業者及び国外事業者は、第三者通報ではなく**自社による検知・対応**について、**24時間365日稼働で違反投稿のパトロール・検知・判定**、**投稿した瞬間と一定の再生回数に達した段階で検知・対応**、**随時の目視確認**、**技術による自動的な研修と人による組み合わせ**等を回答。 ●　一部の国内事業者及び国外事業者は、自社による検知・対応ではなく、**当事者や関係者からの問い合わせ・通報**や**ファクトチェック団体との連携等**により検知・対応していると回答。 【第三者通報を待たずに自社による検知・対応状況】(参考資料22-1-1 及び22-1-2 Q5－3(1)) ○　一部の国内事業者及び多くの国外事業者は、**関連キーワードの抽出・自動表示**、**システムと人間の両方による検知・対応**、**自動モデレーション**において、**ＡＩ等の機械的手段を利用**していると回答。 ●　多くの国内事業者及び一部の国外事業者は、**ＡＩ等の機械的手段を利用していない**と回答。 【自社による検知・対応におけるAI等の機械的手段の利用状況】(参考資料22-1-1 及び22-1-2 Q5－3(2)) ○　ほぼ全ての国内事業者及び一部の国外事業者は、**自社で検知してからモデレーション等を実施するまでの目標期間**について、**作業上の目安を設定し事案に応じた対応**を行う旨、**検知後速やかに対応**する旨や**検知・**

対応の結果 24 時間以内に削除された動画の割合を公表（23 年 7～9 月期に日本で削除された動画の総数に占める割合は 76.8%）している旨等を回答。

● 一部の国内事業者及びほぼ全ての国外事業者は、処理の目標期間について、**明確な回答なし**。

【自社による検知・対応における処理期間の目標の設定状況】（参考資料22-1-1 及び22-1-2 Q5－2－(5)）

○ 一部の国内事業者及び国外事業者は、**個別のコンテンツに対する審査の結果、偽・誤情報ポリシー等違反の有無に関する疑義が生じた場合のプロセス・判断**について、**社内におけるエスカレーション**、**法的な判断が難しい場合における裁判所の判断への依拠**、**外部評価者との連携**等を回答。

● 多くの国外事業者は、**個別のコンテンツに対する審査の結果、偽・誤情報ポリシー等違反の有無に関する疑義が生じた場合のプロセス・判断について明確な回答なし**。

【疑義が生じた場合のプロセス・判断状況】（参考資料22-5-1 Q1－4 及び 参考資料22-5-2 Q1－10、Q1-14）

【取組［例］】

- 投稿監視をするオペレータが判断に迷う場合は、シフトリーダーへエスカレーションし、シフトリーダーが判断。シフトリーダーが判断に迷う場合は、マネージャーにエスカレーションし、マネージャーが判断 （ここまでは業務委託先により実施）。マネージャーが判断に迷う場合は、CX 推進室にエスカレーションし、CX 推進室で判断。CX 推進室で判断に迷う場合は、マネージャーにエスカレーションし、マネージャーが判断。法的な判断が難しい場合は、自分たちで法律や事実を評価することはせず、裁判所の判断に従う。
- ポリシーに違反しないものの、それに近いコンテンツ（いわゆる「ボーダーラインコンテンツ」）は視聴されているコンテンツのごく一部。機械学習を用いて、潜在的に有害な誤情報を含むこの種のコンテンツが視聴者の目に触れる機会を減少。どのコンテンツが、ボーダーラインコンテンツに該当するかを判断することは困難であり、動画の品質について重要な情報提供を世界各地にいる外部評価者に依頼。外部評価者は、一般に公開されている詳細な評価ガイドラインに従ってトレーニングを受けている。ボーダーラインコンテンツであると決定するには、評価者はコンテンツが不正確、誤解を招く、欺瞞的、無神経、不寛容、有害または害をもたらす可能性があるなどの要素について評価。その結果を総合して、動画に有害な誤情報が含まれている可能性、またはボーダーラインである可能性をスコア化し、そのコンテンツをチャンネル登録していないユーザーにおすすめすることを制限。一方で、信頼性について判断するために、評価者はいくつかの重要な点を確認。コンテンツの裏付けは取れているか、目的を達成しているか、動画が目的を達成するにはどのような専門知識が必要か、動画に登場するスピーカーやそのチ

	ャンネルの評判はどうか、動画の主なトピックは何か（ニュース、スポーツ、歴史、科学など）、コンテンツは主に風刺を意図しているのか、これらの点を確認した上で動画の信頼性を判断。ニュースや情報コンテンツの場合、スコアが高い動画ほどより宣伝される。 ・　コミュニティガイドライン違反について疑義があるコンテンツについては、自動モデレーション技術と人間の両方で対応。違反の可能性が検出された場合には、自動モデレーションシステムは追加審査のために、人間の審査担当者、つまりモデレーターに情報を送信してさらに審査をする場合や、自動モデレーション技術による審査の段階でコンテンツがコミュニティガイドラインに違反しているという確証の度合いが高い場合には、自動的にコンテンツを削除する場合もある。コミュニティガイドラインに違反するか否かの判断が困難な場合、ただ形式的に判断するのではなく、それぞれの地域の文化や背景に配慮してモデレーションチーム内でどのように審査すべきかを議論する体制を構築。 **○　一部の国外事業者は、日本の災害に関する偽誤情報が海外において外国語で流通・拡散している場合に実施している工夫について、世界各国のモデレーションチームと各国の状況等について随時意見交換やフィードバック等を回答。** **●　多くの国外事業者は、日本の災害に関する偽誤情報が海外において外国語で流通・拡散している場合に実施している工夫について明確な回答なし。** 【自国内の偽・誤情報と比較して判断が難しいことについての工夫の状況】（参考資料22-5-2 Q1－16） **【取組［例］】** ・　各国で見られた新たな問題やトレンドをグローバルチームに随時共有し、グローバルで行っているポリシー開発や対策に活用。 ・　担当部門と密接に連携して、自社のプラットフォームで組織的な影響力の行使が行われている可能性がある事例を特定し調査。悪意のある人物を特定してそのチャンネルとアカウントを停止し、他のテクノロジー企業と協力してインテリジェンスとベストプラクティスを共有し、脅威に関する情報を法執行機関に報告。 ・　日本を含む世界各国にモデレーションチームが存在しており、常にモデレーションを担当するチーム内において情報共有を行ない、他国のモデレーターとも意見交換ができる環境を構築

	○　一部の国内事業者及び国外事業者は、**透明性レポートとしての定期的な公表**や**自社のウェブページでの掲載**と回答。 ●　ほぼ全ての国内事業者及び国外事業者は、**非公開等と回答**。 【第三者通報対応や自社による検知・対応の公開状況】(参考資料22-1-1 及び22-1-2 Q5－4)
4　偽・誤情報の流通・拡散への対応状況	○　半数の国内事業者及びほぼ全ての国外事業者は、偽・誤情報ポリシー等への違反を理由として**投稿の削除や非表示等**を行った**日本国内における全体の件数**を回答。 ●　全ての国内事業者及び国外事業者は、偽・誤情報ポリシー等への違反を理由としたコンテンツモデレーション等について、**投稿の削除以外のモデレーション等を行った日本国内における件数についての明確な回答なし**。 【偽・誤情報ポリシー等違反に関するモデレーション等の実施件数】(参考資料22-1-1 及び22-1-2 Q6－1(1)) ○　一部の国外事業者は、**偽・誤情報ポリシー等違反のコンテンツの検知・対応に従事する人数**について、**日本語のコンテンツに対応可能な人数**を回答。 ●　ほぼ全ての国外事業者は、**偽・誤情報ポリシー等違反のコンテンツの検知・対応に従事する人数のうち日本語のコンテンツに対応可能な人数**について、**全世界のモデレーターのうちの日本語を主要言語とする割合のみの回答**や、**国境を超えて担当者が知識や経験を持ち寄ることで適切かつ必要な対応を行なうこと等、明確な回答なし**。 【日本語のコンテンツに対応可能な方の人数の状況】(参考資料22-5-2 Q1－11) ○　一部の国内事業者及び国外事業者は、**偽・誤情報ポリシー等への違反を理由として投稿の削除等を行った日本国内における全体の件数のうちＡＩ等の機械的手段のみによって検知・対応した件数**又は**ＡＩ等の機械的手段と人間による組み合わせにより検知・対応した件数**を回答。 ●　ほぼ全ての国内事業者及び国外事業者は、偽・誤情報ポリシー等への違反を理由として投稿の削除等を行った日本国内における全体の件数のうちＡＩ等の機械的手段のみによって検知・対応した件数又はＡＩ等の機械的手段と人間による組み合わせにより検知・対応した件数について、**件数を取得していないことや公表していないこと等、明確な回答なし**。 【AI 等の機械的手段を利用して検知した件数】(参考資料22-1-1 及び22-1-2 Q6－1－(2)～(3))

	○　一部の国外事業者からは、**ＡＩ等の機械的手段のみによって検知・対応した検知した結果に誤りがあることが事後的に判明した件数**を回答。 ●　ほぼ全ての国内事業者及び国外事業者は、**ＡＩ等の機械的手段のみによって検知・対応した検知した結果に誤りがあることが事後的に判明した件数**について、**件数を取得していないこと、削除済みコメントを全件チェックすることは困難であることや公表していないこと等、明確な回答なし**。 【AI等の機械的手段による判定結果に誤りがあることが事後的に判明した件数】(参考資料22-1-1 及び22-1-2　Q6－1－(4)) ○　一部の国内事業者は、**第三者通報の総数**、**第三者通報を契機としたモデレーション等の実施件数**、**第三者通報の受付からモデレーション等実施までの平均期間**を回答。 ●　全ての国外事業者は、**第三者通報の総数、第三者通報を契機としたモデレーション等の実施件数、第三者通報の受付からモデレーション等実施までの平均期間**について、また、全ての国内事業者及び国外事業者は、**モデレーション等の有無・内容に関する通報者からの不服申立等の件数**について、**明確な回答なし**。 【第三者通報の総数、それを契機としたモデレーション等の実施件数、それの受付からモデレーション等実施までの期間、モデレーション等の有無・内容に関する通報者からの不服申立等の件数】(参考資料22-1-1 及び22-1-2　Q6－1－(9)～(12)) ○　一部の国内事業者及び国外事業者は、**モデレーション等の対象となった投稿者（発信者）からの日本語による苦情や不服申立てを受け付ける窓口（以下「苦情等受付窓口」）を通じた苦情・不服申立て件数**、**そのうち実際にモデレーション等の撤回につながった件数**を回答。 ●　ほぼ全ての国内事業者及び国外事業者は、**苦情等受付窓口を通じた苦情・不服申立て件数、そのうち実際にモデレーション等の撤回につながった件数等について、明確な回答なし**。 【投稿者(発信者)からの日本語による苦情・不服申立て件数、苦情・不服申立てに基づくモデレーション等を撤回した件数、苦情・不服申立ての受付からモデレーション等の撤回までの期間、苦情・不服申立ての再審査要求の件数】(参考資料22-1-1　及び22-1-2　Q6－1(13)～(17)) ●　ほぼ全ての国内事業者及び全ての国外事業者は、**投稿者（発信者）からの日本語による苦情・不服申立てのうち特定の発信者（投稿者）による苦情・不服申立ての件数等について、明確な回答なし**。 【投稿者(発信者)からの日本語による苦情・不服申立て件数のうち特定の発信者(投稿者)による苦情・不服申立ての件数、当該苦情・不服申立ての受付からモデレーション等の撤回までの期間、モデレーション等の有無・内容に関する発信者(投稿者)・通報者との間で

	訴訟等に発展した件数】(参考資料22-1-1 及び22-1-2 Q6－1⑱～⑳) ○　一部の国外事業者は、**日本におけるコンテンツの削除に関する特徴的な理由**について、**災害に関するものや新型コロナウィルス感染症に関する内容**が存在すると回答。 ●　多くの国外事業者は、日**本におけるコンテンツの削除に関する特徴的な理由について、明確な回答なし**。 【日本におけるコンテンツの削除に関する特徴的な理由】(参考資料22-5-2 Q1-5、Q1-6) ○　全ての国内事業者及び国外事業者は、**ポリシー違反コンテンツに対するモデレーション以外の対応**について、**捜査機関への情報開示**や、**自殺予告・児童の性被害・殺害予告等について警察等への通報**等を回答。 ●　多くの国内事業者及び国外事業者は、ポ**リシー違反コンテンツに対するモデレーション以外の対応**について、**警察等への通報件数等について非公表であること等、明確な回答なし**。 【ポリシー違反コンテンツに対するモデレーション以外の対応】(参考資料22-5-1 Q1-8、参考資料22-5-2 Q1-15) ○　半数の国内事業者及び全ての国外事業者は、**信頼できる情報のプロミネンス**について、**ニュースサービスにおけるトピックスやトップニュースとしての掲載**、**おすすめ動画や情報パネル等における信頼できる情報源からの高品質な情報の表示順位の上位化**、**サードパーティの偽情報インデックス指標を用いて信頼性の高いものの表示順位の上位化**、**政府機関・警察・消防等に対する災害の場合における特別なグレーバッジの提供**や**緊急告知をサポートするための無償広告枠の活用**等を回答。 ●　半数の国内事業者は、**信頼できる情報のプロミネンスについて、優先表示は行っていない旨の回答や、明確な回答なし**。なお、一部の国内事業者は、**公的機関が優先表示しやすい形で情報を発信している場合は関連トピックに優先的に枠を作る等今後の取組としては検討**と回答。 ●　**ほぼ**全ての国内事業者及び全ての国外事業者は、**2022 年中及び 2023 年中に実施したポリシー違反によるアカウント停止や投稿の削除等件数のうち日本の政府・自治体・伝統メディアのアカウントに関するものの状況**について、**該当する事例がない**旨の回答や、**政府・自治体・伝統メディアのアカウントか否かの判別や属性毎の管理をしていないことやそのようなアカウントがないこと等、明確な回答なし**。 【信頼できる情報のプロミネンスに関する取組状況】(参考資料22-2-1 及び22-2-2 Q19、同22-2-1 及び22-2-2Q19、同22-5-1Q2及び22-5-2Q4)

	【取組［例］】 **■Yahoo!ニュース** ・ トピックスは、「公共性」と「社会的関心」などを軸として、政治や経済など広く伝えるべき重要度の高いニュースや、スポーツなど多くの人々の関心を集めるニュースといった多様なニュースをユーザーに提供。 ・ トピックスで取り上げるコンテンツは、記事提供元より配信されるものから社内の編集部員が選択。トピックスの掲載場所は、Yahoo! JAPAN トップページや Yahoo!ニュースのページなどで、見出し（最大 15.5 文字）あり。 ・ Yahoo! JAPAN トップページでは、「いま話題になっている重要ニュース」を中心に、トピックスを選び、掲載。トップページにおけるトピックスの並べ方は、ニュースの「公共性」と「社会的関心」のバランスを加味。災害や有事に関する情報や国政選挙など、ユーザーの生命・財産や生活に大きく関わるもの、あるいはオリンピックといった著しく社会的関心が高まる国民的なイベントなどについては、トピックスで集中的に扱う場合も。 ・ トピックスの見出しを押すと、複数のコンテンツがまとめられたページが表示。このページでは、ニュースの基本情報、背景・経緯などの情報をまとめ、ユーザーの理解や行動をサポート。遷移先に掲載するものは、パートナーから配信されるコンテンツのほか、官公庁のサイトや各メディア、民間団体のサイトなど、一定の信頼性があるページを中心に選択。 **■LINE NEWS** ・ 「LINE アプリのニュースタブ」ではタブ上部を中心に「いま話題になっている重要ニュース」を中心に掲載。トップニュースとは LINE アプリの下方にある「ニュース」のタブを選択した際に表示される「主要ニュース」部分で、編集部が見出し。トップニュースは「いま話題になっている重要ニュース」を中心に、記事を選び、ページ上部に掲載。 ・ トップページにおける並べ方は、ニュースの「公共性」と「社会的関心」のバランスを加味。災害や有事に関する情報や国政選挙など、ユーザーの生命・財産や生活に大きく関わるもの、あるいはオリンピックといった著しく社会的関心が高まる国民的なイベント時は、枠内で同じ事象のニュースを複数掲載も。 ・ 大規模災害などが発生した際、パートナーから配信された最新情報の記事を複数掲載するほか、各自治体・官公庁などによる、ユーザーの生命財産・生活に関わる情報(避難・各種インフラ情報、災害用伝言板など)をトップニュースやトップニュースからの掲載先の記事ページなどで掲載。 **■ドワンゴ** ・ 新型コロナウイルスの流行に伴い、政府関連機関が発信する情報が広く伝わるよう、コンテンツ視聴ページに誘

導表示を実施。新型コロナウイルス流行と同等レベルの、国内情勢に甚大な影響を与える非常事態が発生した場合には、再度類似の取り組みを行う可能性あり。

■はてな

・ 現在のところ一律に優先表示はしていない。公的機関が優先表示しやすい形で情報を発信している場合は関連トピックに優先的に枠を作るなど、今後の取組としては検討。

・ はてなブックマークでは、新聞社のサイトなど伝統メディアに対するブックマークが多く、自然に上位に表示される状況。ただ、昨今では情報の主要部分が有料であることも多く、システム上で伝統メディアを上位表示することは考えていない。

・ 一律に各サービスで信頼性の高い情報を優先表示できる仕組みといった法整備についてはポジティブ。

■サイバーエージェント

・ 取り組みは実施していない。必要な状況が発生した場合は個別対応。今後も同様に個別に対応する予定。

■YouTube

・ YouTube の検索結果、おすすめ動画、情報パネルにおいて、信頼できる情報源からの高品質な情報の表示順位を上げ、ユーザーが正確で有益な情報を見つけられるように対応。その一環として、例えば、偽・誤情報が生まれやすい話題やニュースに関連するコンテンツについて説明する情報パネルや、ニュースメディアが発信する情報を提供。さらに、YouTube のニュース速報セクションや最新ニュースセクションなど、信頼できるニュースの情報源となる関連動画を紹介する専用のサービス機能も用意。

・ YouTube のポリシーに違反してはいないものの、有害な方法でユーザーを誤解させる可能性のあるコンテンツの拡散を制限。これには、陰謀論の動画や、許容範囲の限界事例となるようなコンテンツが含まれる。このようなコンテンツは YouTube に残るとしても、YouTube がユーザーに推奨したり、検索結果の上位に表示することもない。

・ YouTube は収益化のより高い基準を満たすクリエイターに還元。YouTube は質の低いコンテンツの収益化を防ぎ、視聴者に常に信頼性の高い高品質のコンテンツを提供するクリエイターに収益化の機会を提供。

・ 信頼できる質の高いニュース提供元を識別してニュース機能で取り上げるために、YouTube のシステムはチャンネルの品質や、最近のニュースや関連事件の報道範囲など、さまざまなシグナルを使用。また、チャンネルは Google 検索結果の機能に関するポリシーと Google ニュースのコンテンツポリシーを遵守している必要。Google は機械学習技術を使用して、ニュース機能のニュース提供元を特定するためのシグナルを生成。サードパーティ

	の評価担当者からのフィードバックにより、これらのシグナルを継続的に強化、評価、改善。 ■TikTok Japan ・ 信頼できる情報（例えば省庁が発表している情報など）に飛べる、あるいは選挙に関する情報に飛べるラベルが存在。 ・ まずは正しく、今何が起こっているのかという情報を TikTok 内で見られる環境を作るのが大切。そこで、ウェザーニュースと連携し、災害発生時の信頼性の高い情報の発信源として、ウェザーニュースのライブをずっと見られるような形。例えば日本国内で震度 5 弱以上の地震が観測された場合、TikTok の日本ユーザーに対し、自動的に緊急地震速報をプッシュ通知。この通知をタップすると、TikTok のライブストリーミング機能に遷移し、そこでウェザーニュースのライブが見られる。 ■Microsoft ・ サードパーティのニュースガード偽情報インデックス指標というものを用いており、そういった情報を基に、信頼性の高いものをランクを上げて表示。緊急危機時、災害時や投票のときだけでなく、一般的に実施。ヨーロッパでは、ウクライナの戦争、選挙などがあったため、そういった指標を基に、情報源からの情報を基に実施。 ■X ・ 政府関係、公共の政府、あるいは、警察、消防等、災害の場合については、特別なグレーバッジを提供するという形で閲覧。 ・ 能登半島地震後に Yahoo/日本財団等に提供したような緊急告知をサポートするために無償広告枠の活用等は今後も検討可能。
５　偽・誤情報の発信者（投稿者）の表現の自由等への配慮	○　ほぼ全ての国内事業者及び全ての国外事業者は、**コンテンツモデレーション等の対象となった発信者（投稿者）への表現の自由等への配慮**について、**発信者（投稿者）のマイページへの訪問**、**登録メールアドレスへの通知**、**アプリ内通知やカスタマーサービスへの問い合わせ**等により、**投稿したコンテンツに違法性があると法的請求を受けたこと**、**削除されたコンテンツ**、**投稿の削除等になったことや違反したポリシー等削除理由等の確認ができる**と回答。 ●　全ての国内事業者及びほぼ全ての国外事業者は、**投稿の削除等以外のコンテンツモデレーション等について、発信者（投稿者）への実施の事実等の通知等に関する明確な回答なし**。 【発信者（投稿者）に対するモデレーション等の実施の事実や理由の通知等状況】（参考資料22-1-1 及び22-1-2　Q4－1）

	○　全ての国外事業者は、**苦情等受付窓口**について、**再審査請求含め発信者（投稿者）からの日本語による専用の窓口を整備・公表**と回答。 ●　全ての国内事業者は、**一般的な相談等と共通の対応窓口にて対応**しており、専用の窓口は整備していない。 ○　ほぼ全ての国内事業者は、苦情等対応受付窓口について、**日本語対応が可能な人数を回答**。 ●　全ての国外事業者からは、**苦情等対応受付窓口における日本語通報に対応可能な人数について、随時変動しうることや非公表とする等、明確な回答なし**。 【苦情等受付窓口において、日本語通報対応が可能な人数】(参考資料22-1-1 及び22-1-2　Q4－3(1)) ○　全ての国内事業者及び一部の国外事業者は、苦情等受付対応について、**ＡＩその他の機械的手段は利用していない**と回答。 ●　一部の国外事業者は、**苦情等受付対応におけるＡＩその他の機械的手段の利用について、当該手段の概要や利用手順（どのようなケースで用いるのか等）に関する明確な回答なし**。 【苦情等受付対応における AI 等の利用状況】(参考資料22-1-1 及び22-1-2　Q4-3(2)) ○　ほぼ全ての国内事業者は、苦情等受付窓口を通じた日本語による苦情等をどの程度の期間で処理しているかについて、**２４時間以内や２営業日以内等の目標期間**を回答。 ●　全ての国外事業者は、**処理の目標期間について、非公表とする等、明確な回答なし**。 【苦情等受付における処理期間の目標の設定状況】(参考資料22-1-1 及び22-1-2　Q4－3(5)) ○　ほぼ全ての国内事業者及び一部の国外事業者は、**苦情等受付対応結果を発信者（投稿者）に通知等**していると回答。 ○　一部の国内事業者及び一部の国外事業者は、**発信者（投稿者）からの再審査に対応**していると回答。 ●　多くの国外事業者は、**苦情等受付対応結果を発信者（投稿者）に通知等しているかどうかに関する明確な回答なし**。 ●　ほぼ全ての国内事業者及び多くの国外事業者は、**発信者（投稿者）からの再審査に対応しているかどうかに関する明確な回答なし**。

	【苦情等受付対応結果の発信者(投稿者)への通知等状況】(参考資料22-1-1 及び22-1-2 Q4－3－(6)) 【苦情等受付対応結果に関する発信者(投稿者)からの再審査制度の状況】(参考資料22-1-1 及び22-1-2 Q5－2－(7)) ○ 一部の国内事業者及び国外事業者は、**政府機関やNGO/NPOなど特定の発信者（投稿者）からの苦情等の優先的取扱いの実施**を回答。 ● ほぼ全ての国内事業者及び一部の国外事業者は、**政府機関やNGO/NPOなど特定の発信者（投稿者）からの苦情等の優先的取扱いは実施していない**ことを回答。 ● 全ての国内事業者及び国外事業者は、**偽・誤情報ポリシー等に違反する偽・誤情報について、特定の発信者（投稿者）からの苦情等の優先的取扱いを実施しているかどうかに関する明確な回答なし**。 【特定の発信者(投稿者)からの苦情等の優先的取り扱い】(参考資料22-1-1 及び22-1-2 Q4－3－(8))
6 レコメンデーションやモデレーション等に関する透明性・アカウンタビリティ確保に向けた取組状況	○ 多くの国内事業者及び国外事業者は、**ＡＩ等の機械的手段の利用によるコンテンツモデレーション等の自動的な対応**について、**不適切コメント対策、ポリシー等違反可能性コンテンツの検出や同違反コンテンツの特定、ポリシー違反ではないが潜在的に有害な誤情報等のボーダーラインコンテンツのおすすめ制限、わいせつ・不快・不適切な画像やコメントの抽出・判定**等において**考慮する要素や用いられている主なパラメータやアルゴリズム**を回答。 ○ 一部の国内事業者及び国外事業者は、ＡＩ等の機械的手段によってコンテンツモデレーション等を自動的に対応する際に考慮する要素や用いられている主なパラメータやアルゴリズムについて、**透明性レポートや自社ウェブページ等で公開**と回答。 ● 多くの国内事業者及び国外事業者は、**コンテンツモデレーション等のアルゴリズムやパラメータの重み付け等の詳細な内容について、非公開等、明確な回答なし**。 【コンテンツモデレーション等を自動的に決定している場合の主なパラメータ等の状況】(参考資料22-1-1 及び22-1-2 Q8－1～Q8－2) ○ 全ての国内事業者及び多くの国外事業者は、**ＡＩ等の機械的手段の利用によるレコメンデーションの自動的な対応**について、**建設的・注目コメントや信頼できる情報源からの高品質な情報の優先表示、ユーザの属性情報等による投稿者・使用者双方が満足する掲出マッチング、多様なコメント表示、ニュースコンテンツのランキング**等、ＡＩ等の機械的手段によってレコメンデーションを自動的に対応する際に**考慮する要素や用いられている主なパラメータやアルゴリズム**を回答。

	○　全ての国内事業者及び多くの国外事業者は、ＡＩ等の機械的手段によってレコメンデーションを自動的に対応する際に考慮する要素や用いられている主なパラメータやアルゴリズムについて、**自社ウェブページや利用規約等で公開**と回答。また、一部の国外事業者は、と回答。 ●　多くの国内事業者及び国外事業者は、**レコメンデーションのアルゴリズムやパラメータの重み付け等の詳細な内容について、非公開等、明確な回答なし**。 【レコメンデーションを自動的に決定している場合の主なパラメータ等の状況】(参考資料22-1-1 及び22-1-2　Q8－3～Q8－4) ○　半数の国内事業者は、**レコメンデーションに使用している要素の公開可能性**や**必要に応じたアルゴリズムの開示可能性**を回答。 ○　一部の国外事業者は、プラットサービスやコンテンツモデレーション等のアルゴリズムについて、**秘密保持契約を締結した上で限られた研究者にはコードの一部公開や、研究者を含む第三者におすすめのタイムラインのアルゴリズムを Github で開示**を実施していると回答。 ●　全ての国内事業者は、**コンテンツモデレーション等やレコメンデーションのアルゴリズムについて、研究機関等ふくめ第三者に開示していない**と回答。また、ほぼ全ての国外事業者は、**米国・欧州で連携している一部の研究者のみへの研究者向けＡＰＩの公開**を除き、**研究者等の特定の第三者への開示に関する明確な回答なし**。 【アルゴリズムの特定の第三者への開示状況】(参考資料22-1-1 及び22-1-2　Q8－5、Q8－6)
7　令和６年能登半島地震関連の偽・誤情報の流通・拡散への対応状況	○　一部の国内事業者は、令和６年度能登半島地震に関連して、**投稿の削除（削除基準としては、例えば、明示的に虚偽情報を拡散させようとする投稿や、その説を強く信じ込み、他者に対してもそれを信じさせようとする意図が感じられる投稿）・非表示（例えば、人工地震と断定し流布するような投稿や募金を募る行為について利用規約等に照らして違反行為と認められた場合）やアカウント停止等を実施した件数**を回答。 ●　ほぼ全ての国内事業者及び全ての国外事業者は、令和６年能登半島地震に関連して、**投稿の削除等のコンテンツモデレーション等を行った日本国内における全体の件数について、明確な回答なし**。 【能登半島地震におけるモデレーション等の取組状況】(参考資料22-1-1 及び22-1-2　Q7－1) ○　一部の国内事業者は、令和６年度能登半島地震に関連して、**投稿の削除対象とする根拠として、ファクトチェック機関により明確に誤りとされていることを根拠として採用**したことと**削除を実施した件数**を回答。 ●　ほぼ全ての国内事業者及び全ての国外事業者は、令和６年能登半島地震に関連して、**投稿の削除等におけ**

	るファクトチェック機関との連携や削除等を実施した件数について、明確な回答なし。 【能登半島地震に関連するコンテンツにおけるファクトチェック状況】(参考資料22-1-1 及び22-1-2 Q7－2) ○　一部の国内事業者及び国外事業者は、能登半島地震に際し、**チーム編成や特別な対策等による震災関連投稿のパトロール・モニタリングの強化（ポリシー等違反コンテンツの一斉確認、危険なキーワード・ハッシュタグのモニタリング、偽・誤情報が含まれていないか注意を払うよう審査員チームに対する指導、審査の精度向上のためのガイドラインの共有等）**、**危機管理プロトコルによる対応**、**最新情報まとめページや特設ページの作成**、**警鐘を鳴らすトピックや図解の掲載**、**地震関連のデマの打ち消しのトピックスの作成**、**専門家やジャーナリストによる注意喚起や記事の紹介**、**キー局のライブ配信、ローカル局や地方新聞が運営するチャンネル等の信頼できる情報を見つけやすくする施策**、**偽情報に関する注意喚起の掲載・若年層向けの啓発動画キャンペーンの紹介**、**自治体や政府機関との連携**など、偽・誤情報の流通への対応を強化したことを回答。 ●　多くの国内事業者及び一部の国外事業者は、**能登半島地震における対応体制について、既存人数で対応可能等、対応を強化せず、また明確な回答なし**。 【能登半島地震におけるコンテンツモデレーション等に関する体制強化の状況】(参考資料22-1-1 及び22-1-2 Q7－3) ○　一部の国内事業者及び国外事業者は、能登半島地震に際し、**業界団体（SMAJ）において他事業者と連携した注意喚起（実際の被害と異なる救助要請といった虚偽情報の拡散や、送金を促す詐欺行為）**、**ファクトチェック機関との連携**、**民間の気象関連機関との連携による情報発信**、**内閣府・総務省・警察庁や地方自治体との連携**等を実施したことを回答。 ●　多くの国内事業者及び国外事業者は、**関係機関等との連携状況について明確な回答なし**。 【能登半島地震における関係機関等との情報共有等の状況】(参考資料22-1-1 及び22-1-2 Q7－4) ○　一部の国内事業者及び多くの国外事業者は、今後の地震等災害において、**ユーザへの注意喚起・リテラシー向上**、**政府の公表情報やファクトチェック機関によるファクトチェック記事の活用**、**信頼できる情報空間の在り方に関する情報発信**、**技術・ユーザーの報告・専門家やファクトチェック機関からのトレンドレポートによる誤情報の検出**、**地域の誤情報トレンドの積極的な監視や新たな誤情報に類似するコンテンツの積極**

	的なチェック等モニタリング強化やＮＰＯや行政機関との連携による緊急コミュニケーションのサポート等の取組を実施することを回答。 ●　多くの国内事業者及び一部の国外事業者は、今後の対応について、強化する予定はないこと等明確な回答なし。 【災害に関連する偽･誤情報への今後の対応】(参考資料22-1-1 及び22-1-2 Q7－5) ○　一部の国内事業者及び国外事業者は、災害時における偽・誤情報への対応における支障について、緊急時においては一見して明らかな虚偽であるか信頼できる機関からの情報提供がないと真偽の判断が困難であること、ファクトチェック能力を有していないことや投稿の件数が膨大であることから偽・誤情報であることを理由に記事を削除することが難しいこと、政府としての公式情報をワンストップで確認できるページがあれば正しい情報への誘導がしやすくなること、情報は常に変化し専門家や当局の見解が一致しないこともあり、正確で権威ある情報を得ることが難しい場合があること等を回答。 ●　多くの国内事業者及び国外事業者は、災害時における偽・誤情報への対応における支障について、特に回答なし。 【災害に関連する偽･誤情報への対応における支障】(参考資料22-1-1 及び22-1-2 Q7－6)
８　選挙時の偽・誤情報の流通・拡散への対応状況	○　一部の国内事業者は、選挙関連ポリシーの日本国内での策定・運用状況（直近の国政選挙に際しての認知件数、通報件数、対応件数等）や日本国内で選挙時に実施した対応及びその効果について、選挙管理委員会からの指摘に基づいて公職選挙法に抵触する行為が利用規約では禁止事項とされていることへの対応、信頼できる質の高い情報を上にあげることを回答。 ●　ほぼ全ての国内事業者及び全ての国外事業者は、選挙関連ポリシーの日本国内での策定・運用状況や日本国内で選挙時に実施した対応及びその効果について明確な回答なし。なお、今後の更なる取組として、米国では選挙に関する信頼できる情報を整理したページを政府や NPO が作成しておりラベルからリンクさせやすく、日本でも選挙情報に関し、特に信頼できる情報が集まった場所を、政府や関連機関でまとめる仕組みがあると、プラットフォーム事業者でも、それにユーザーがアクセスできるような仕組みを作り上げていくという連携ができるのではないかと回答。 【選挙関連ポリシーの日本国内での運用状況】(参考資料22-2-1 及び22-2-2 Q3) 【日本国内で選挙時に実施した対応及びその効果】(参考資料22-2-1 及び22-2-2 Q21) **【取組［例］】**

・ 信頼できる質の高い情報を上に上げ、コンテンツモデレーションを行い信頼できない情報をダウングレード。
・ Thread Analysis Group と連携し、国内外の協調的な選挙介入に関する情報を収集し、業界内他社や法執行機関と共有。
・ YouTube にこうしたリスクを判定するためのチームがあり、ユーザーエクスペイエンスやコミュニティにとって有害なものを分析・研究。最近は生成 AI に特化。
・ 選挙に関しても選挙に特化したコンテンツモデレーターがモデレーションを行い、ローカルランゲージでのサポートも実施。モデレーションのための機械学習にも投資。
・ 虚偽の選挙関連の情報が特定された場合には、削除を実施。そういったアクション等は、透明性報告書（Transparency Report）の中で透明性を持って開示。選挙の候補者や選挙当局から何か通報があってアクションを取った場合、その対処についても透明性報告書の中で開示を実施。
・ 今年は世界中で多くの選挙が予定されているため、そういった活動をしっかりと報告するとともに、LinkedIn では、既に Transparency Report を開示。
・ ファクトチェックの結果なども開示。選挙に限ったことではなく、一般的に誤情報・偽情報があった場合の内容なども、こちらのレポートの中で報告。
・ 選挙期間については、通常とは違う事態・状況であり、特別な対策を基本的には切り分けて実施。3Tierという3層に分かれた対策を実施。①人為的な対応、②ポリシーの設定、③プロダクト上の対応。①は、選挙等に関しては、具体的に専任のチームを編成し、24時間、常に無休で対応し、マシンラーニングのモデルの活用も含めて稼働し、十分な対策を取れるような体制を実施。②は、選挙等が発生したときに対応する専用のポリシーやルールを策定。具体的には、例えば、選挙については、選挙関連、あるいは政治的な関連と思われる広告については、こういったルールを適用。ここについても、ポリシーに対応する、運用するための人材を配置し、マシンラーニングなどのテクノロジーも活用して対応。③も、改変し強化をすることによって、機能を強化。例えば、マシンラーニングなど、十分に具体的にどのようなコンテンツがどのようなものなのかということの判定も含め、対応できるようなプロダクト上の改定なども鋭意実施。

● **全ての国外事業者は、「選挙におけるＡＩの不正利用に対抗するための技術協定（ミュンヘン協定」について日本で具体的にどのように対応するかの明確な回答なし。**

【「選挙におけるAIの不正利用に対抗するための技術協定（ミュンヘンアコード）」の日本国内で実施予定の取組状況】（参考資料22-

	2-1 及び22-2-2 Q5) **【取組［例］】** ・ ①有用な情報を有権者に伝える（質の高い信頼できる情報を検索結果の上位に上げる）、②プラットフォームを不正から守る（一連のポリシーや様々なツールの利用）、③AIで生成されているか否かを識別する（例えばYouTubeでは、改変又は合成された動画であるというラベル付けをクリエイターに求めている）。 ・ Geminiが回答できる選挙関連のクエリを制限。 ・ 業界横断の取組に積極的に関与（①AI で生成されたコンテンツであることをユーザーに知らせる、②C2PA（コンテンツの来歴及び信頼性のための標準化団体））。 ・ ツールやトレンド情報などの情報（ディープフェイクのデータセット、不正コンテンツのハッシュ値など）を共有し、それによって悪意あるアクターがAIを使って選挙に不正に介入しようとして偽・誤情報を打ち出すのを防止。 ・ ①コミュニティガイドラインを厳正に執行していく、②サードパーティファクトチェッカーとのパートナーシップに基づいてファクトチェックを厳正に行っていく、③利用者のリテラシー向上のための努力を引き続き推進していく、④生成 AI によって生成されたコンテンツにラベルを付けていく（ラベル付けの技術をより洗練させ、またオープンソース化することによって他の企業も使えるような形にしていく）。 ・ コンテンツの来歴。デジタルウォーターマーキング、または、サインをすることにより、ユーザーが情報やビデオがどのような情報源から来ているのか、それが権威ある情報源なのかどうかを理解できるようにする。このコンテンツの来歴を使うことで、例えば選挙に関係する当局が、この情報やコンテンツに署名をすることにより、しっかりとした情報源から展開されているということをユーザーが理解をし、例えば何か疑いのあるもの、もしくは誤解を生むような情報があった際に、その情報源に対して直接何か掛け合うことができるようにする取組を実施。 ・ 検知の機能の強化として、例えば様々な業界、また、様々な市場において、AIが生成した選挙に関わる虚偽の情報があった際に、それをレポートができる、それを検知することができるようにしている。 ・ 候補者や担当が、例えばAIが生成した情報が画像、ビデオ、音声であれ、候補者に関して生成されたものが何か誤解を生むものであるとか、虚偽の情報であるとか、そういうことが生じた際に、マイクロソフトのプラットフォーム上にレポートをするポータルを開設し、そのような情報を共有することが可能。それによって、必要な是正措置、検証を行い、必要があれば、そのような情報を削除。 ・ 権威ある選挙に関する情報、選挙に関する当局から直接展開されている選挙に関する情報を、Bingサーチまたは Bingチャットで提供。その信憑性もマイクロソフトにおいて検証・担保し、投票する際に、しっかりとした信頼でき

	る情報に基づいて判断することができるように取り組んでいる。 ・ 選挙に関する当局と連携することで、ユーザーに対してAIがもたらし得る影響、インパクトについても啓蒙活動を行い、AIがもたらし得るリスクを理解し、それにしっかりと対処できるように教育プログラムを提供。
9　なりすましへの対応状況 **10　広告の質の確保への対応状況**	**○　全ての国内事業者及びほぼ全ての国外事業者は、<u>なりすまし対策</u>について、<u>投稿等のコンテンツや広告に対するポリシーや、削除やアカウント停止等による対応</u>を回答。** **●　一部の国外事業者は、<u>なりすまし対策について明確な回答なし</u>。** 【なりすまし対策（削除等対応の具体的な条件など）状況】（参考資料22-2-1 及び22-2-2 Q8） 【なりすましアカウントやなりすまし広告への対応状況】（参考資料22-5-1 及び22-5-2 Q5-1） **【取組［例］】** ・ LINE ヤフー共通利用規約において、「15.当社サービス利用にあたっての順守事項」として、「（13）当社または第三者になりすます行為または意図的に虚偽の情報を流布させる行為」を規定。 ・ 各サービスにおいて、禁止行為として、「ユーザー本人以外の人物や集団、組織などを詐称する、もしくは関係性を詐称する行為」、「（Yahoo!ニュースコメント欄）、「禁止事項9：なりすまし行為や自作自演他人になりすましたり、他人に誤認を与える内容の投稿、自作自演行為によって、質問、回答を繰り返す投稿」（Yahoo!知恵袋）、「他人になりすましたり、意図的にアイコンや表示名を他のユーザーに似せたり同じにし、不適切な投稿をする行為」（Yahoo!ファイナンス掲示板）、「著名人や企業（法人・個人事業主いずれも含みます。）などの公式運営と誤認させる、又はそのおそれのあるプロフィールをトークルームに設定する行為」（LINE オープンチャット）等を規定。 ・ LINE 公式アカウントについては、ポリシーとして、LINE 公式アカウント利用規約、LINE 公式アカウントガイドラインにてなりすまし行為を禁止し、認証済アカウント開設時の審査と、全ての LINE 公式アカウントに関する監視（ユーザーから通報によるものも含む）を実施。なお、上記利用規約とガイドラインについて、分かり易く伝えるためのウェブサイト）などを使用して、LINE 公式アカウントオーナーへの啓発を実施。 ・ LINE 広告・Yahoo!広告については、なりすまし行為を直接的に禁止する広告掲載停止・ アカウント停止の基準はなく、なりすましによって名誉毀損や肖像権侵害の侵害になったり、商標権等の権利を侵害する表現の広告を広告掲載基準で禁止。また、権利侵害の恐れがある内容が運営サイトに含まれている場合は、広告アカウントを開設不可。 ・ なりすまし広告による投資詐欺への対策として、「未認証の LINE 公式アカウント」や「個人の LINE アカウント」の

	友だち登録へ誘導する場合は、広告の掲載を停止。このような投資詐欺広告については、「投機心、射幸心をあおる表現」「サービス、商品の内容が不明確なもの」「不適切と判断したもの」等の基準で掲載停止となることがある。 ・ 一般のアカウントやコンテンツについて、なりすまし・無断使用（権利侵害行為）の対象となった当事者自身からの要請があり、該当行為の事実が確認された場合には対応を実施。「有名人」と判断できる人物を対象としたものではなく、サービス利用者に他者がなりすましているといったケースでも対応を検討。当事者からの要請がない状態では、実際になりすまし・無断使用（権利侵害行為）であるか否かを当社が明確に判断できないため、ポリシーでの対応は定めていない。なお、「ニコニコ」サービスでは、作成したひとつのアカウントにて、複数のサービスの利用や商品購入が可能となっているため、違反行為への基本的な対処は、アカウントの削除ではなく、違反があったサービスごとに実施。 ・ 広告について、弊社媒体に対して配信している広告は、弊社が広告主となる企業もしくは広告代理店に提案を行い広告案件を獲得している「純広告」と、Google や Amazon 等の媒体社向け広告サービスを介して配信される「ネットワーク広告」の 2 種類が存在。前者の純広告に関し、クリエイティブに使用される素材に関する権利に関しては、広告主向けに展開しているガイドライン上で、広告主側で責任を負うことを明記。 ・ はてな利用規約では「他者になりすましてサービスを利用したり、情報を改ざんする行為」を禁止し、なりすまし行為は利用停止相当の違反行為。ただし、名称やロゴを利用していてもなりすましの意図があるか否かについては判断が難しいケースも多数。（利用者が以前より偶然同じ名称を使用していたり、ある企業のファンがその企業のロゴマークをアイコンとして利用するなど）明確に不正行為を意図していると判断できない場合には、なりすまされた当事者からの申立によりプロバイダ責任制限法に基づき対応。その場合は、なりすましによる情報発信が、商標権や肖像権、パブリシティ権などの権利を侵害する情報であると定義しての対応。 ・ なりすましと判断できる場合は、ブログページ全体を非表示にする対応を実施。 ・ 不実表示のポリシーに基づき、当事者が指示していないにもかかわらず、その当事者と関連している、又はその当事者による指示を受けているといったことを誤って提供するような広告やコンテンツは削除。繰り返し行われているようであればアカウント自体を削除。 ・ より強固なポリシーとして、許可されないビジネス手法に関してのポリシーに基づき、具体的に金銭を奪う詐欺を行う目的で公的な人物との関連性やその人の指示を謳っているものに対しては、即座にアカウントが Google Ads

	から停止され、二度とGoogle Adsに広告を出稿できない。 ・ ポリシーとして、金銭の搾取を目的としたコンテンツは削除。有名人の画像、偽の画像などを活用した広告も対象。同ポリシーは、生成AIを使ったものか否かにかかわらず適用。そのコンテンツがどうやって作られたかということは関係ない。ミスインフォメーションそのものだけでなく、虚偽の情報を伝えるニュアンス等をも対象としている。なりすまし行為をしたアカウントの削除も実施。 ・ 広告に関しては広告宣伝ポリシーの中で、有名人の画像はその本人の許可なく掲載してはならないとしている。 ・ 著名人のなりすまし、ロゴ等を無断で使用した広告は、広告ポリシーの禁止コンテンツに該当し、そもそも表示される前に広告審査ではじかれている。なりすまし等の被害に遭った方から通報いただける窓口も準備。見つけた瞬間にアプリ内から通報することも、ウェブ上のフォームで通報することも可。 ・ 広告を利用いただく際に認証プロセスを設け、必ず事前にベリフィケーションプロセスを経ないと広告を実施することができない。完全ではないものの、ペイメント情報が必ずひもづく形になるので、非常に重要なプロセスの一つ。直近3月22日から始まったが、特定カテゴリーにおける広告運用に関しては、事前に目視での広告審査を強化し、幾つかのカテゴリーに関連するところで既に開始。 ・ そもそもポリシーで禁じており、具体的には、他者の身元を示す2つ以上の要素を使用した場合は、これはポリシー違反。極論、Xのアカウントを持っていない方が、誰かに自分の名前と写真ともろもろを使ってアカウントを作られて嘘の投稿をされていることがあった場合、こういったものもポリシー上、ポリシー違反として対応することが可能。 ○ 一部の国内事業者は、**投稿等のコンテンツに関するアカウント開設時**において、**SMS等による認証やオフィシャルブロガーへの与信審査**の実施を回答。 ● 全ての国内事業者は、**投稿等のコンテンツに関するアカウント開設時の事前審査**について、**アカウントが他サービスの利用にあたっても必要となること、投稿時においてはサービスごとに表示名の設定が可能であること等から、なりすましに関しては、アカウント開設時ではなく、アカウント開設後に個別サービスにおける禁止事項等への違反等があった場合に必要な範囲でアカウント停止等の措置を行うなど、事前審査を実施していない**と回答。 ● ほぼ全ての国外事業者は、**投稿等のコンテンツに関するアカウント開設時の日本国内における事前審査に**

	ついて明確な回答なし。 ○　半数の国内事業者及び一部の国外事業者は、**配信される広告に対する日本国内における事前審査**について、**アカウント審査基準による審査**、**一部に対するアカウント開設時による本人確認**、**掲載開始前の目視による審査**、**広告ポリシーへの準拠を確認するための広告主のアカウントとコンテンツに対する承認プロセス**等を実施と回答。うち一部の国内事業者は、**本人確認で否認されたアカウント数**と**問題あった件数全体における本人確認で否認された件数の割合**を回答。 ●　ほぼ全ての国外事業者は、**配信される広告に対する日本国内における事前審査について明確な回答なし。** 【なりすましアカウントやなりすまし広告への対応状況】(参考資料22-5-1Q5-2(1)(2)及び22-5-2Q7-1(1)(2)) 【広告サービス状況】(参考資料22-5-2 Q6) **【取組［例］】** ・ Yahoo! JAPAN ID 及び LINE ID の取得に当たってはそれぞれ SMS 等による認証を実施。これらの ID は他サービスを利用するにあたっても必要となること、投稿時においてはサービスごとに表示名の設定が可能であること等から、なりすましに関しては、アカウント開設時ではなく、アカウント開設後に個別サービスにおける禁止事項等への違反等があった場合に必要な範囲でアカウント停止等の措置を実施。 ・ オフィシャルブロガーには与信審査、トップブロガーにおいても所定の確認を実施。 ・ 以下のような場合には、サービスの使用を許可せず、アカウントや他のエンティティ(ページ、グループ、イベントなど)を制限または無効。 – スクリプトやその他の不正な手段でアカウントを作成または使用する場合、 – 以前の削除を回避するためにアカウントやエンティティを作成または再利用する場合、 – 他の人を誤解させたり欺いたりするために意図的に身元を偽るアカウントを作成または使用する場合 ・ 誤解を招く身元の偽装を評価する際には、次のような要素を考慮。 – 名前や年齢などの身元情報の繰り返しまたは重大な変更 – 誤解を招くような自己紹介内容や位置情報などプロフィール情報 – ストック画像や盗まれた写真の使用 – その他の関連するアカウント活動 ・ LINE 広告・Yahoo!広告について、アカウント開設にあたり、アカウント審査基準に基づく審査(「アカウント開設に関する基準」として、「アカウント申込時の登録情報から不正な広告出稿の懸念があると判断した場合は、アカウン

	トを開設できません。」、「開設後のアカウントに関する基準」として、「アカウントの登録情報から不正な広告出稿の懸念があると判断した場合」は「アカウントを停止します。」)と、アカウントを申込みいただいた広告主が本人かを確認するため、一部のお客様に対しお申込み時に「本人確認」を実施。 ・ 純広告については、掲載ガイドラインを含む弊社規約に同意頂いたことになる旨の申込確認書を締結しているため、全てのクリエイティブは広告主の責任の下で許諾されていることが前提。配信される全ての広告に対して、掲載開始前に弊社側で目視による審査を実施。 ・ 偽広告防止のための基準(掲載ガイドライン)としては、一般的に偽広告が多く見られる、健康食品やサプリメント、投資サービスといったカテゴリの広告に関しては別途レギュレーションを規定。 ・ 提携している DSP 事業者からの広告が配信。基本的には広告クリエイティブが入稿された際に審査が行われていると理解。基準については各事業者のポリシーに準拠されていますが、基本的には JIAA の定めるガイドラインに従っていると理解。 ・ Google の不実表示に関するポリシーでは、広告主による「広告主のビジネス、商品、サービスに関する情報について隠蔽または虚偽記載を行ってユーザーを欺く」広告や「広告のリンク先で『フィッシング』の手口を使ってユーザー情報を収集する」広告の運用は許可されない。 ・ 今年に入ってから、著名な人物、ブランド、組織の支持を受けている、またはそれらと提携関係にあると虚偽の宣伝をしている悪質な広告主のアカウントを素早く永久停止できるようにするポリシーの更新を発表。 ・ 新たに広告配信の制限ポリシーを施行。新たな広告主には「会社確認」を行う期間を設定。この期間中は、広告インプレッションが制限される可能性。(当面このポリシーは、広告キャンペーンでブランド名を対象にしている広告主にのみ適用。つまり、広告主とその広告で言及されているブランドの関係性が不明確な場合や、その広告が一般的な内容である場合に、このポリシーが適用。) ・ 広告が Facebook または Instagram で表示される前に、広告基準に基づいてレビュー。これは、広告が表示される前に自動的に実施。また、個別のレビュアーが自動化技術を改善およびトレーニングするために働いており、場合によってはマニュアルで広告をレビュー。以前に承認された広告が再レビューの対象となることがあり、その理由は、利用者がその広告を非表示にしたり、ブロックしたり、報告したり、その他の否定的なフィードバックを行った場合等で、これらのフィードバックは、レビュー過程で見逃したものを示す手助け。既に表示されている広告が弊社のポリシーに違反している場合、当該広告は取り下げられ、広告の配信は停止。

	・ 現在アクティブなすべての広告は広告ライブラリ（https://www.facebook.com/ads/library/）で閲覧可能。社会問題、選挙、政治に関する広告は、アクティブでなくなっても 7 年間広告ライブラリで閲覧可能。利用者は、ポリシーに違反していると思われる広告を広告ライブラリから直接調査し報告可能。 ・ 独立したサードパーティのファクトチェッカーによってファクトチェックされ、評価された広告を拒否。 ・ 広告主が X 広告を使ってコンテンツを宣伝する場合、そのアカウントとコンテンツは承認プロセスの対象。このプロセスにより、X は広告主が広告ポリシーを遵守しているかを確認。 ○ ほぼ全ての国内事業者は、**なりすましアカウント開設後や偽広告の配信後の日本国内における対応件数**について、**悪質ななりすましユーザが利用するオープンチャットの削除件数**、**なりすましアカウントによるブロック件数と停止全体における割合**、**広告アカウントにおける本人確認で否認されたアカウント数と問題あった件数全体における本人確認で否認された件数の割合**、**検索広告・ディスプレイ広告・LINE 広告における「未認証の LINE 公式アカウント」や「個人の LINE アカウント」の友だち登録へ誘導する広告の否認数とその全否認数に占める割合**等を回答。 ○ ほぼ全ての国内事業者は、**なりすましアカウント開設後の日本国内における対応の契機**について、**カスタマーサポート等への入信**、**ユーザーによるアプリ通報**、**なりすまされた本人からの報告**を回答するとともに、**なりすまされた本人からの報告に対しては、本人からの報告の可能性があると判断し本人確認・被害状況の確認を行った件数や報告者からの返答があり行為状況を確認できたため実際に対応を行った件数**等を回答。 ○ 一部の国内事業者は、**偽広告の配信後の日本国内における対応の契機**について、**全て自動システムによる検知のため、なりすまされた本人やその他消費者団体等外部からの報告はない**と回答。 ● ほぼ全ての国外事業者は、**なりすましアカウント開設後や偽広告の配信後の日本国内における対応件数、なりすましアカウント開設後や偽広告の配信後の日本国内における対応の契機等について明確な回答なし。** ● ほぼ全ての国外事業者は、**明らかにポリシー違反の広告が日本において多数残ってしまっている主な原因について明確な回答なし。** ● 一部の国内事業者及びほぼ全ての国外事業者は、**生成 AI により作成される有名人の偽動画等を活用した偽広告に対する日本国内における対策や効果**について、ポリシー施行における最先端の大規模言語モデル（LLM）の活用を除き、**AI によって生成されたメディアやコンテンツの共有自体が違反ではない**ことや、**生**

	成 AI に特化した対策は実施しておらず、偽広告に係るアカウントや広告が審査基準に抵触する場合には、アカウント開設の不可や広告掲載の停止を実施と回答。 【なりすましアカウントやなりすまし広告への対応状況】(参考資料22-5-1Q5-2(3)～(6)及び22-5-2Q7-1(3)～(6)、Q7-2、Q7-3) ○　多くの国外事業者は、**政治広告の日本国内における制限**について、**選挙キャンペーンに紐ついた広告や政治広告等の禁止**等を回答。 ●　一部の国外事業者は、**政治広告の日本国内における制限について明確な回答なし**。 【政治広告の制限(制限の内容や具体的な条件など)状況】(参考資料22-2-1 及び22-2-2 Q9) **【取組［例］】** ・ 偽・誤情報に関するポリシーや不実表示に関するポリシーがあり、広告も含めたプロダクト全般に当てはまり、例えば、危険な広告、他者の名誉を傷つけるような広告、選挙等の民主的なプロセスを毀損するような広告、市民社会に混乱を来すような広告などを禁止。 ・ 選挙キャンペーンに紐付いた広告は日本では不許可。政治的な信条に基づいたターゲティングを広告で行うことも不許可。 ・ 政治広告(政治に関する問題提起をするような広告)は禁止。 ・ 政治的なメッセージであったり、実はなりすましであるとか、あるいは、実際には関係のない政党との関係者というような偽ったIDを持っていろんなものを投稿してくるということも含めて、何らかの政治的な意味を持ったものについては、①国家的なレベルで、例えば、トロールファームを使ってフェイクアカウントを大量に作成して投稿してくるものが検知された場合は、トロールファームの関連のアカウントは停止しており、あるいは、これが明らかになりすましであって、本人ではないということが分かったような場合についても、このアカウントについて削除を実施。②実際にはなりすましではないが、何らかの実在するような政党であったりグループと関係のないアカウント、投稿者であるにもかかわらず、いかにもそれをかたるような形で何かを投稿しているような場合についても、ディセプティブID、欺瞞的なID、これを検知した場合についても、削除を実施。③その時々の、特に地政学的なトレンド等に関連したものにも、リアルタイムで対応するように全力。一例としては、10月７日を起点として中東での情勢が緊迫した中で、ハマスのいろんな行動に端を発したものにつき、それ以降、実際のアカウントとして対応したものが80万アカウント。 ・ グレーバッジという国の機関からのアカウントに関しては、グレーのバッジをつける形で、なりすましをされにくくす

	るというような新しい対応策等も実施。
11 広告配信先の質の確保への対応状況	○ 一部の国内事業者は、**広告配信先のパブリッシャーが運営するメディア（広告媒体）の日本国内における事前審査**について、**広告配信ガイドラインや広告ネットワーク利用規約に基づく審査**、**訓練された担当者や専門部署による審査**、**広告配信後における審査やモニタリングの実施**等を回答。 ● 一部の国外事業者は、**広告配信先のパブリッシャーが運営するメディア（広告媒体）の日本国内における事前審査について明確な回答なし**。 【広告サービス状況】（参考資料22-5-1Q4及び22-5-2Q6） **【取組［例］】** ・ 配信前の審査として、掲載先パートナー選定の際に、主に広告配信ガイドラインに基づく審査を実施。Yahoo!広告の広告配信ネットワークには、メディアとしての知名度・実績があり、掲載されている情報に信頼性のある多数のパートナーが参加。サイトのテーマやコンテンツの傾向、サイト内に運営者情報を明記されているかなど、さまざまなポイントについて、ガイドラインに基づき審査を行い、広告の配信に適したパートナーであるかを検討。審査はトレーニングを積んだ担当者によって行われ、ガイドラインに抵触しないサイトに対してのみ広告が配信。配信前の審査を通ったサイトでも、コンテンツ内容は流動的なものも多く含まれるため、広告配信が開始した後もガイドラインに違反している内容がないかを審査。 ・ LINE 広告（LINE 広告ネットワーク）では、配信前の審査として、広告配信ネットワークのパートナー選定時に「LINE 広告ネットワーク利用規約」に基づく審査を実施。審査基準には、登録情報の一致、アプリストアでの掲載状況、コンテンツの適正性などが含まれ、これにより信頼性と安全性を確保。専門部署によって行われる事前審査では、各アプリが「LINE 広告ネットワーク利用規約」に適合しているかを確認。問題があると判断されたアプリは配信対象外。事前審査を通過したアプリに対しては、外部機関からの情報を活用し、定期的な監視やクローリング技術を活用したモニタリングを実施。特定のリスク評価基準に基づき、適宜審査を行い、問題が発見された場合には迅速に対応を行うことで、広告の品質を維持。 ・ Google パブリッシャー向けポリシー（https://support.google.com/adsense/answer/10502938?hl=ja&ref_topic=1250104&sjid=9707478738802213653-AP）に違反すると、コンテンツに基づいた広告の表示がブロックされたり、アカウントが停止または閉鎖されたりする場合がある。本ポリシーは、Google パブリッシャー向けサービスのご利用に関するポリシーの附則として適用。次のカテゴリに分類：コンテンツポリシー、行動ポリシー、動画ポリシー、プライバシーに関するポリシー、要件とそ

	の他の基準。 ○　一部の国内事業者は、**広告配信先のパブリッシャーが運営するメディア（広告媒体）の日本国内における事前審査の件数**について、**配信先に掲載されている情報の真偽を確認する審査は実施していないが配信先に掲載されている情報に懸念があり不適切と判断した件数とその総否認数に占める割合**、**肖像権・パブリシティ権・名誉・信用毀損等の確認により問題と判断した件数とその審査基準違反件数全体に占める割合**を回答。 ●　一部の国外事業者は、広告配信先のパブリッシャーが運営するメディア（広告媒体）の日本国内における事前審査の件数について明確な回答なし。 【広告サービス状況】(参考資料22-5-1 及び22-5-2 Q4(2)) ○　一部の国内事業者は、**メディア（広告媒体）が悪質なサイトであった場合に広告主が広告掲載停止を求める通報窓口の設置等の日本における状況**について、**広告配信ガイドライン違反が疑われる場合や担当営業を通じ広告主が特定の配信先への広告掲載の停止を依頼する窓口の設置**、**広告主自身が広告管理ツールから掲載先サイトの情報を確認しいつでも自由に配信停止を行うことが可能であること**を回答。 ●　一部の国外事業者は、**メディア（広告媒体）が悪質なサイトであった場合に広告主が広告掲載停止を求める通報窓口の設置等の日本における状況について明確な回答なし**。 【広告サービス状況】(参考資料22-5-1 及び22-5-2 Q4(3)(4))
12　発信者への広告収入分配等の状況	○　ほぼ全ての国内事業者及び一部の国外事業者は、**発信者への広告収入の分配プログラム**について、**記事の読者ユーザによる評価、視聴者のエンゲージメント（いいね、コメント、シェア等を含む）**、**作品の盛り上がり（作品の閲覧数やコメント数等によって算出）**や**良質な情報を提供する信頼できるメディアやブロガーとの支払・パートナー契約の締結**等により、**チャンネル登録者・フォロワー数、インプレッション数やコンテンツの再生時間等のいわゆるインプレッション稼ぎにより発信者に広告収入が分配されないような仕組みで対応**と回答。 ●　一部の国内事業者及び多くの国外事業者は、**発信者への広告収入の分配プログラム**について、**チャンネル登録者・フォロワー数、インプレッション数やコンテンツの再生時間等のいわゆるインプレッション稼ぎにより発信者に広告収入を分配**と回答。 【発信者への広告収入分配における基準】(参考資料22-1-1 及び22-1-2 Q9) 【発信者の収益化の停止・無効化に関するポリシーの日本国内での運用状況】(参考資料22-2-1 及び22-2-2 Q2)

【発信者への収益還元状況】(参考資料22-5-1Q3及び22-5-2Q5)

【取組［例］】

■Yahoo!ニュース「課題解決バリュープログラム」 https://news.yahoo.co.jp/info/news-operation-policy#section2

・ Yahoo!ニュースでは、PV だけでは測れない良質なコンテンツをユーザーに届けるため、パートナー(記事配信元)とのエコシステムを強化。具体的には、既存の情報提供料に加え、PV に連動しない情報提供料を支払いする取組(「課題解決バリュープログラム」)を実施。基本的にはパートナーである媒体者に対し、PV(ページビュー)をベースに支払い。それとは違った形で、ユーザーに最後まで読んでいただいて各記事について評価をしていただき、そうした評価に応じた支払いの仕組みを設定。

■LINE NEWS

・ LINE NEWS においては、アライアンスチームが信頼できる各種メディアと契約し、情報の正確さ・信頼性、その裏付けとなる取材体制について一定の担保がなされているほか、それらの良質な情報を提供するメディアに対して双方合意に基づいた最適な支払い契約を締結。契約後においても、配信記事のガイドラインに対するモニタリング、ファクトチェックも含む校正校閲等を実施し、良質な情報が配信されるよう実施。

■LINE VOOM Creator Program (LVCP)

・ LINE VOOM 側が用意した原資について、投稿者における投稿パフォーマンスに応じてプログラム参加者に分配する仕組み。2021 年からクローズドで運用を開始し、2024 年 5 月に一般開放を予定。他 SNS のフォロワー5000 以上であること、及び、投稿内容に違反がないことが収益化の条件。

■ドワンゴクリエイター奨励プログラム「作品収入」 https://qa.nicovideo.jp/faq/show/19563?site_domain=default

・ 投稿コンテンツによって収益を得る方法として、「作品収入」という機能を用意。これは、プレミアム会員(サービスの月額会員)あるいは、本人確認書類の提出による本人確認手続きを行った一般会員が、投稿した各作品について「作品収入申請」を行い、第三者の権利を侵害していないか等のポリシーに則った審査を通った作品に対して奨励金を付与するシステム。本システムではプレミアム会員収入や企業広告収入等を原資にしており、作品の盛り上がり(作品の閲覧数やコメント数等によって算出)や、対象作品で贈られたギフト・ニコニ広告に応じて算出された金額を奨励金として分配。審査の中で「第三者権利物の利用と思わしき内容」や「弊社の審査ポリシーに反すると思わしき内容」を確認した場合には、その作品の収益化は認められないが、投稿者からの異議申立により問題がないと判断した場合には収益化が可能。なお、「掲載された情報の真偽」を判断するのは困難なため、「コンテンツに偽・誤情報が含まれるか否か」は、作品収入の審査基準に含まれていない。

■サイバーエージェント

・　トップブロガーや公式ブロガーに関しては、報酬還元や支払いの仕組みを導入。基本的には個々のブロガーと全員契約を締結。ブロガーの選別自体を厳しくし、信頼できるブロガーの方とパートナー契約を締結。PV に基いて収益分配するブロガーは基準による審査したオフィシャルブロガー・トップブロガーに限定。規約に基づき違反をしたブロガーは収益配分権利を剥奪。アフィリエイトの場合は、本人情報を登録いただいた上で、基準により審査した登録者に限定。PV 収益分配と同様、規約に基づき違反をしたブロガーは収益権利を剥奪。

■YouTube チャンネル収益化ポリシー　https://support.google.com/youtube/answer/1311392?hl=ja

・　YouTube のすべてのチャンネルは、コミュニティガイドラインを遵守する必要がある一方、クリエイターの YouTube パートナープログラムを通じた収益化（例えば、広告収入については、①チャンネル登録者数 1,000 人及び②次のいずれかを満たしている:公開されている長尺動画の過去 365 日間における総再生時間が 4,000 時間以上、又は、公開されているショート動画の過去 90 日間の視聴回数が 1,000 万回以上）に対しては、さらに高いハードルを設定。収益化するには、Google の広告掲載に適したコンテンツのガイドラインを含む YouTube のチャンネル収益化ポリシーを遵守する必要。YouTube のチャンネル収益化ポリシーに違反した場合、収益化が停止または完全に無効になる可能性。

■ TikTok クリエイター向け収益化プログラム「Creator Rewards Program」
https://prtimes.jp/main/html/rd/p/000001002.000030435.html

・　Creator Rewards Program について、コミュニティガイドライン違反のコンテンツはそもそも対象外。偽・誤情報に関するコンテンツには収益が還元されていくことがない形で運営。TikTok では他のプラットフォームと異なり、投稿される個々の動画と、広告として投稿される動画は、直接には紐ついていない。システムの仕組み上、ある投稿者の動画の再生中に、広告動画が差し込まれて再生されることもない。このようなシステムの特徴から、より良いコンテンツの投稿を促したりクリエイターを支援する方法としては、現時点では「Creator Rewards Program」などのプロジェクト型のクリエイター支援を実施。

■Xクリエイター広告レベニューシェア（Creator Ads Revenue Sharing）
https://help.x.com/en/using-x/creator-ads-revenue-sharing

・　X の広告レベニューシェアでは、X 上で投稿したコンテンツへの返信に表示される広告の認証済みユーザーのオーガニックインプレッションから収益を共有。人々が X を通じて直接生計を立てる手助けをするための取り組み

の一環。資格を得る方法として、クリエイター広告収益共有の対象となるには、以下の条件を満たす必要。①X プレミアムまたは認証組織に加入していること、②過去 3 ヶ月間の投稿の累計で少なくとも 500 万のオーガニックインプレッションを有していること、③少なくとも 500 人のフォロワーを有していること。また、コミュニティノートがついた投稿は、レベニューシェアの対象外。

○　全ての国内事業者は、**発信者への広告収入の分配プログラムへの参加に関する日本における事前審査**について、**審査基準・フロー**、**本人確認の実施**、**審査により参加が否認された件数とその申請件数に占める割合**等を回答。

●　ほぼ全ての国外事業者は、**発信者への広告収入の分配プログラムへの参加に関する日本における事前審査**について、**審査基準・フローや本人確認の実施に関する明確な回答がなく**、全ての国外事業者は、**審査により参加が否認された件数とその申請件数に占める割合は非公開や把握困難と回答**。

【発信者への広告収入分配における基準】(参考資料22-1-1 及び22-1-2 Q9)
【発信者の収益化の停止・無効化に関するポリシーの日本国内での運用状況】(参考資料22-2-1 及び22-2-2 Q2)
【発信者への収益還元状況】(参考資料22-5-1Q3及び22-5-2Q5)

【取組［例］】

■LINE VOOM Creator Program（LVCP）

・　審査フローは、契約 MCN から毎月 20 日までに新規追加希望のクリエイターリスト(クリエイター名・クリエイターの他 SNS 情報等)を提出、他 SNS の投稿内容を見て審査担当者が審査を実施(具体的な確認事項は、過度に暴力的な表現、露骨な性的表現、児童ポルノ・児童虐待に相当する表現、人種、国籍、信条、性別、社会的身分、門地等による差別につながる表現、自殺、自傷行為、薬物乱用を誘引または助長する表現、その他反社会的な内容を含み他人に不快感を与える表現がないか、宣伝目的の投稿でないか、著作侵害のコンテンツがないか、その他、社内審査担当チーム内で不適切と判断したものがないか)、審査結果を各 MCN 担当が最終チェックを行い承認。

■ドワンゴクリエイター奨励プログラム「作品収入」　https://qa.nicovideo.jp/faq/show/19563?site_domain=default

・　本人確認:プレミアム会員(弊社サービスの月額会員)であること、あるいは、本人確認書類の提出による本人確認手続きが完了したことをもって本人確認。この条件を満たしていないと、「作品収入」の申請自体が不可能。

・　基準:「作品収入」の申請が行われたコンテンツに対しては、収益化の可否を判断するために、審査を実施。審

	査内で「第三者権利物の利用と思わしき内容」や「弊社の審査ポリシーに反すると思わしき内容」を確認した場合には、その作品の収益化は否認。ただし、投稿者からの異議申立により問題がないと判断した場合には収益化が可能。 **■サイバーエージェント** ・ 2 点を審査。①Ameba の利用規約に同意しており、規約に準拠したサービス利用を行っていること、②18 歳以上であること(18 歳以上 20 歳未満の未成年者の場合は、親権者などの法定代理人の同意を取得。18 歳未満の場合は、法定代理人の同意があってもお申し込み不可能)。おまかせ広告の利用権限は Ameba アカウントに紐付。Ameba ブログ自体の通常監視が事前の悪質なアカウントの 1 次フィルタリングになっており、ブログの監視により利用停止の対象となったブログアカウントは同時におまかせ広告の利用権利を喪失。本人確認は本審査プロセスの中には含まれていないが、発生した報酬を実際に受け取るためのフローの中で本人確認が実施。 **■YouTube チャンネル収益化ポリシー** https://support.google.com/youtube/answer/1311392?hl=ja ・ YouTube のクリエイターが収益化のために YouTube パートナープログラム(YPP)への登録申請を行う際、コミュニティガイドライン、ポリシー等を全て遵守していることをレビュー・確認。YPP に登録したクリエイターへは、広告がビデオ内で再生された場合に、そのネットレベニューの 55%を支払。YPP に登録されたチャンネルにおいて、広告掲載に適したコンテンツのガイドラインに違反するものがあった場合、掲載される広告の数が減少。場合によっては YPP の資格自体を喪失(広告宣伝からのレベニューシェアだけでなく、他の収益化機会も喪失)。例外的な場合、コミュニティガイドラインへの繰り返しの違反や悪質なコンテンツに関しては、チャンネル自体を削除する場合もある。 **■ TikTok クリエイター向け収益化プログラム「 Creator Rewards Program 」** https://prtimes.jp/main/html/rd/p/000001002.000030435.html ・ 「Creator Rewards Program」に参加するには、18 歳以上であること、コミュニティガイドラインを遵守したコンテンツであること等が必要。その上で、当プログラムにおいては主に以下の 4 つの指標に焦点を当て、報酬が支払われる。①動画のオリジナリティ(クリエイター自身が制作したオリジナル動画コンテンツであること。クリエイター独自の視点や創造的な思考プロセスを表現しているもの)、②再生時間(再生時間には動画の視聴時間と視聴完了率の両方を含む)、③視聴者のエンゲージメント(いいね、コメント、シェア等を含む)、④検索価値(人気の検索ワードに基づいてコンテンツに割り当てられる指標)。

	■Ｘクリエイター広告レベニューシェア(Creator Ads Revenue Sharing) https://help.x.com/en/using-x/creator-ads-revenue-sharing ・　Ｘの広告レベニューシェアでは、X 上で投稿したコンテンツへの返信に表示される広告の認証済みユーザーのオーガニックインプレッションから収益を共有。人々が X を通じて直接生計を立てる手助けをするための取り組みの一環。資格を得る方法として、クリエイター広告収益共有の対象となるには、以下の条件を満たす必要。①X プレミアムまたは認証組織に加入していること、②過去 3 ヶ月間の投稿の累計で少なくとも 500 万のオーガニックインプレッションを有していること、③少なくとも 500 人のフォロワーを有していること。また、コミュニティノートがついた投稿は、レベニューシェアの対象外。 ○　全ての国内事業者は、**発信者への収益化の停止等に関するポリシーの日本における運用状況**について、**収益化対象に限定されていないが偽・誤情報ポリシー等違反による削除率（全削除数に占める割合）、収益化可能件数のうち収益化不可能となった件数・割合、利用停止（広告の利用権利の喪失）の対象となった件数等を回答**。 ●　全ての国外事業者は、**発信者への収益化の停止等に関するポリシーの日本における運用状況**について、**日本においてポリシー違反により収益化を停止等した件数と世界全体における当該件数における日本国内の件数の割合は非公開や把握困難と回答**。 【発信者への広告収入分配における基準】(参考資料22-1-1 及び22-1-2 Q9) 【発信者の収益化の停止・無効化に関するポリシーの日本国内での運用状況】(参考資料22-2-1 及び22-2-2 Q2) 【発信者への収益還元状況】(参考資料22-5-1Q3及び22-5-2Q5)
13　AI・ディープフェイク技術への対応状況	○　全ての国内事業者及びほぼ全ての国外事業者は、提供するサービスにおける AI の利用等状況ついて回答。 ●　一部の国外事業者は、**提供するサービスにおける AI の利用等状況ついて明確な回答なし**。 【サービスにおける AI の利用等状況】(参考資料22-1-1 及び22-1-2 Q10-1) **【取組［例］】** ■LINE ヤフー ・　Yahoo!ニュースコメント欄においては、生成 AI を利用したコメント欄の要約機能を試験提供。 ・　LINE オープンチャットでは、生成 AI を利用したトーク内容の要約機能を試験提供。 ・　Yahoo!知恵袋では、生成 AI(ChatGPT-4)を利用した回答機能を提供。当該機能においては、生成 AI によるハルシネーション等のリスクがあることを想定し、正確性について注意喚起をしているほか、法律・医療などの専門

	的知識が必要とされるコンテンツへの提供はしていない。 ■ドワンゴ ・　ニコニコ動画、ニコニコ生放送では、コメント監視に機械学習（ディープラーニング）のシステムを導入。投稿されたコメントに対し、不適切な内容と思われるものの抽出、対応を自動で行ったり、人力目視を行う際のサポートを実施。 ・　ニコニコ生放送では、上記システムによるコメント出し分けの強弱設定をコンテンツ配信者が任意に選択できる「コメントフィルター」という機能をユーザーに提供。 ■はてな ・　ブログ記事の内容に基づいたタイトルを AI で生成する機能を提供。 ■サイバーエージェント ・　ヘルプページチャットボット（質問への自動回答）家事ラク AI（献立レシピ提案）。 ■Google ・　Gemini Advanced：Google の生成 AI チャットボット Gemini（旧名：Bard）は、日本語を含む 40 以上の言語に対応しておりウェブ上で利用可能。昨年の公開以来、世界中の人々が、面接の準備やコードのデバッグ、新しいビジネスアイデアのブレインストーミングなど、生産性を高める新しい方法で AI とコラボレーションするために Gemini（旧 Bard）を活用。Gemini モデルは、Google Workspace や Google Cloud などの人々や企業が毎日使う製品にも導入される予定。 ・　Workspace: 2 月 21 日、Google Workspace は Gemini for Google Workspace と名称を変更し、Google の最も有能な AI モデルにアクセス可能。このアップデートの一環として、Gemini は、何百万人ものお客様が毎日使用している Workspace アプリに組み込み。Gemini for Workspace は、消費者向けやチーム向け等、あらゆる規模の組織向けに提供され、例えば誕生日パーティーの企画からマーケティングキャンペーンの立案、新規事業のビジネスプランの作成などの作業を支援。Gemini for Workspace には、大企業向けのデータ保護を備えた、Gemini とチャットするための新しいスタンドアロンエクスペリエンスも含む。 ・　Google Cloud: Google Cloud の顧客に対しては、Duet AI も今後数週間で Gemini 。Gemini は、企業が生産性を向上させ、開発者がより速くコードを書き、組織がサイバー攻撃から身を守るのを支援し、そのメリットは多岐。

	・ Youtube：Dream Track は、YouTube ショートの試験運用版の楽曲生成ツール。協賛アーティストの音声をもとにして、クリエイター独自の 30 秒間のサウンドトラックを作成可能。現在、この機能は米国の一部のクリエイターのみを対象に、特定のモバイルデバイス限定で提供。世界中のユーザーが、このサウンドトラックをそのままリミックスして自分のショート動画に取り込むことが可能。Dream Screen は、 YouTube Short の画像や動画の背景を生成する AI ツール。AI Insights は、クリエイターのインスピレーションを掻き立て、彼らが次の動画の内容を決めるのに役立つツール。 ■Meta ・ 他の企業が恩恵を受けることができるさまざまなオープンソースツールを開発。例えば、他の企業が自社のプラットフォーム上でテロリストのコンテンツをよりよく検出し、拡散を阻止できるよう、弊社は無料のオープンソースソフトウェアツール、Hasher-Matcher-Actioner (HMA)を公開。HMA は、画像や動画のコピーを特定し、それらに対して一斉にアクションを実施。HMA は、弊社が以前から提供しているオープンソースの画像・動画マッチングソフトウェアをベースにしており、あらゆるタイプの違反コンテンツに使用することが可能。 ・ オープンソースの大規模言語モデルイニシアティブの一環として、弊社は研究および商業利用のための安全で責任ある AI 開発のための Purple Llama を導入。このプロジェクトは、開発者が責任を持って AI モデルを構築できるよう、オープンな信頼性と安全性のツールと評価を特徴。まず Purple Llama には、サイバーセキュリティと入出力セーフガードのためのツールと評価が含まれる。 ■TikTok Japan ・ テキストを入力すると自動で動画の背景を作れる AI グリーンスクリーンなど多様なエフェクト機能や、聴覚障害のある方でも動画の音声を楽しんでいただける、自動で動画の音声を認識して字幕をつける機能のように、AI を活用した多様なサービスを提供。 ■Microsoft ・ マイクロソフトは 2023 年 2 月、AI で強化された Web 検索エクスペリエンスである新しい Bing をリリース。新しい Bing は、Web の検索結果を要約し、チャット エクスペリエンスを提供することで、ユーザーを支援。ユーザーは詩やジョーク、物語といったクリエイティブなコンテンツを生成したり、Bing Image Creator で画像を生成可能。AI で強化された新しい Bing は、マイクロソフトと OpenAI 社のさまざまな先進テクノロジを活用。これらテクノロジには、最先端の大規模言語モデル (LLM) である GPT-4 、自然言語記述からデジタル画像を生成す

	るディープ ラーニング モデルの DALL-E など（共に OpenAI 社の技術）が含まれる。マイクロソフトは、両モデルが一般公開される前の数か月間で、この最先端の AI テクノロジと新しい Bing の Web 検索機能を連携させるための一連のカスタマイズされた機能とテクノロジを開発。2023 年 11 月、マイクロソフトは新しい Bing:の名称を Copilot in Bing に変更。 ・ マイクロソフトは、責任ある AI へのコミットメントに真摯に取り組み。Copilot in Bing のエクスペリエンスは、マイクロソフトの AI の原則とマイクロソフトの責任ある AI の基準に従って開発。開発には、マイクロソフトの責任ある AI オフィス、エンジニアリング チーム、Microsoft Research、Aether をはじめとする全社の責任ある AI のエキスパートたちが協力。 ・ Copilot in Bing では、情報を求めているユーザーに対応する際、Web 検索結果に根拠を置いている。つまり、ユーザーのクエリやプロンプトに対して応答を提供するときは、Web 上の上位のコンテンツを中心。そして、ユーザーが詳細を知ることができるように Web サイトへのリンクも提供。Bing は、関連性、品質および信頼性、鮮度といった特徴に強い重み付けをして Web 検索コンテンツをランク付け。Copilot in Bing からの応答は、出力文がクエリまたはプロンプトによる Web 検索結果や、Bing のファクトチェック済み情報のナレッジ ベース、あるいはチャットの最近の会話履歴（チャット エクスペリエンスの場合）といった入力ソースに含まれる情報で裏付けられていることにより、根拠のある応答と見なされる。根拠のない応答とは、出力文がそのような入力ソースに根拠付けられていない応答のこと。 ・ LinkedIn が現在ユーザーに提供している AI 機能のほとんどは、Microsoft が運用する Azure AI サービスを通じて OpenAI GPT モデルを利用。たとえば、LinkedIn では AI を活用したテイクアウェイとアドバイスを提供。これは、Azure AI サービスによる生成 AI を使用してメンバーにパーソナライズされたアドバイスと洞察を提供し、情報を入手し、スキルを伸ばし、適切な仕事をより簡単かつ効率的に獲得できるようにする一連の機能。もう 1 つの例として、LinkedIn Learning サービスに AI を活用したコーチング機能。これは、チャットボットであり、Azure AI サービスによる生成 AI を使用して、LinkedIn Learning のコンテンツ ライブラリの中を案内し、リーダーシップやマネジメントなどの重要なスキルを身に付けることを支援。 ・ LinkedIn では InBart という GAI モデルを開発しており、LinkedIn 上のリクルーターが求職者候補への InMail メッセージを書くのに役立っている。

○　一部の国内事業者及びほぼ全ての国外事業者は、「AI 事業者ガイドライン」を踏まえた対策状況について、**注意喚起の実施**、**専門的知識が必要なコンテンツの対象外化**、**生成 AI を利用して投稿を行う場合にその旨を投稿上明記するようガイドラインの制定**、**公開前・公開後の安全対策**、**生成 AI コンテンツのウォーターマーク・来歴管理・検知・リテラシー**等を回答。

●　ほぼ全ての国内事業者及び一部の国外事業者は、**「AI 事業者ガイドライン」を踏まえた対策状況について明確な回答なし**。

【AI事業者ガイドラインを踏まえた対策状況】(参考資料22-1-1 及び22-1-2 Q10-2)

【取組［例］】

■LINE ヤフー

- 正確性の担保について注意喚起を実施(ニュースコメント欄・オープンチャット・知恵袋)。
- 法律・医療などの専門的知識が必要とされるコンテンツにおいて生成 AI 活用の対象から除外(ニュースコメント欄・知恵袋)。
- Yahoo!ニュースコメント欄においては、要約結果をユーザーが報告し、事後的に要約機能を停止できる仕組みを用意。
- オープンチャット・知恵袋においては、違反申告の仕組みを用意し、パトロールによる削除の対象。
- ユーザーが生成 AI を利用して投稿を行う場合は、その旨を投稿上に明記するようガイドラインを制定。(知恵袋)。

■Google

- 新しい生成 AI モデルを開発する上で、より大きな規模で品質と安全性の高い基準を満たすよう取り組み。ヘイトスピーチ、誤った情報、嫌がらせなどの分野でシステムの訓練に役立つポリシーを制定。
- 専門家やテスターからの外部のフィードバックと社内のテストを組み合わせて、これらの新製品を安全で有用なものにし、この分野での能力を継続的に向上。
- 公開前の安全対策:AI 原則に基いて製品に最初から安全性を組み込み、製品が危害をもたらすことがないよう厳正なテストを実施。
- 公開後の安全対策:新しいサービスや製品を公開する際には、ユーザーのフィードバックを参考にしながら慎重かつ段階的なアプローチを取って不正使用を防御。
- エコシステムの安全対策: ツールや知識を共有し、ユーザーを教育し、開発者が責任を持って構築できるよう

	支援。 ■Meta ・　国際的なレベルで AI に関する調和されたルールを支持し、その実現に向けた取り組みを支援。例えば、G7 の広島プロセスを注意深く監視し、安全・安心と産業界との協力を促進することを目的とした原則に基づくアプローチを支持。 ■TikTok Japan ・　AI の透明性と責任あるイノベーションのためのフレームワークである「Partnership on AI」の Responsible Practices for Synthetic Media のローンチパートナー。TikTok では、業界のベストプラクティスとなるこの新しい規範に沿って行動することを宣言しており、各システムの開発段階から、透明性と責任ある AI の観点で行動。 ・　例えば AI グリーンスクリーンにおいては暴力的あるいは性的な映像は生成できないように設計するなど、開発段階から安全性の観点を取り入れ。 ■Microsoft ・　悪意ある人物が Bing の生成 AI エクスペリエンスを使用して虚偽の情報を作り出すリスクを最小限に抑えるための保護策を措置。こうした取り組みとしては、分類（Classifiers）やメタプロントの使用、来歴ツール、報告機能の強化、堅牢な運用とインシデント対応など。また、Bing の生成 AI 機能は、不正な情報や虚偽の情報、または誤解を招く情報を作成または共有するためにこれらのサービスを使用することを禁止。 ・　生成 AI モデルが生み出すコンテンツの品質が向上しているため、AI が生成したコンテンツの出所に関する透明性を高める必要性。現在、Azure OpenAI サービスが提供するすべての AI 生成画像には、改ざんを防止するためにコンテンツの出所と来歴を開示するコンテンツ クレデンシャルが含まれる。コンテンツ クレデンシャルは、Joint Development Foundation のプロジェクトである Coalition for Content Provenance and Authenticity (C2PA) のオープン技術仕様に基づく。 ・　ディープフェイクに対する、決め手となるような解決策は存在しません。最も有効なアプローチは、従来の徹底的なセキュリティ計画を実行。具体的には、ディープフェイクの脅威の大部分を軽減することが期待される 4 つの軽減策。①生成 AI コンテンツのウォーターマーク：マイクロソフトが Bing Image Creator で行っているように、生成 AI 画像の作成者を明らかにし、生成 AI により作成されたコンテンツであることを明示することが極めて重要。②来歴：C2PA 基準またはマイクロソフトの新しい“サービスとしてのコンテンツ クレデンシャル”製品を通じ

て提供される来歴により、個人または組織は、どのコンテンツが自らの出所から提供されているのかを主張可能。③検知：画像や動画が AI によって作成または操作されているかどうかを判断できることには大きな価値。どちらのテクノロジも進化し続けるため、検知機能を最終的なソリューションと見なすべきではないが、多くの状況において有益なツール。④·リテラシー：消費者に対する啓発は、AI によるメディアの操作に惑わされないようにするために不可欠。人々がコンテンツ クレデンシャルおよびメディアの明確なラベル付けを要求し、これらに対応していないメディアに疑問を呈することが、より強靭な、そうした策略に負けない社会を実現。

・ LinkedIn では Azure AI サービスが展開する安全対策を導入。また、LinkedIn のすべての AI 製品およびサービスについて「責任ある AI の原則」に取り組み。違法なコンテンツや有害なコンテンツへの対応の問題に特化した「Embrace Accountability」(アカウンタビリティを受け入れる) という原則は、LinkedIn が潜在的な有害性や目的への適合性を評価し、それに対処することや、人間による監督とアカウンタビリティを確保することを含め、堅牢な AI ガバナンスを展開することを意味。LinkedIn は、AI のベスト プラクティス、規範、法律が変化する中で、他者から学び、他者を支援することに取り組み。

○ 半数の国内事業者及びほぼ全ての国外事業者は、AI で生成されたコンテンツの投稿への対応について、ユーザにおいてその旨を投稿上に明記することを求めるガイドライン、実在する人物の映像や音声を含む AI 生成コンテンツを禁止するガイドライン、AI で生成したコンテンツを投稿する際の AI 生成ラベルの義務付け、コンテンツが大幅に改変されたり合成して生成されたものである場合に視聴者に知らせるパネルの表示等を回答。

● 半数の国内事業者及び一部の国外事業者は、AI で生成コンテンツの投稿への対応について明確な回答なし。

【AI 生成コンテンツの投稿に対する検知ツールや投稿時ラベリングの投稿者への義務づけ等状況】(参考資料22-1-1 及び22-1-2 Q10-3)

【AI 生成コンテンツへの対策状況】(参考資料22-2-1 及び22-2-2 Q4)

【取組［例］】

■LINE ヤフー

・ ユーザーが生成 AI を利用して投稿を行う場合は、その旨を投稿上に明記するようガイドラインを制定。(知恵袋)。

■サイバーエージェント

・　AI活用したり、アルゴリズムを組んで、画像に関しても危ない画像を定義しているため、そこを抽出して干渉するというシステムを 20 年前から組んでおり、ノウハウもたまっている。

■Google

・　不実表示、プライバシー、なりすましといった様々な対策や、AI 以前から実施しているディープフェイク等の対策を組み合わせ。

・　Google DeepMind が開発した最先端の透かし技術 SynthID を使用して、Google のモデルが生成する画像に目に見えない電子透かしを挿入。最近発表されたこの技術は、Gemini、ImageFX、その他のエクスペリエンスの最新の画像モデルで生成されるすべてのコンテンツに電子透かしを挿入。

・　ユーザーに追加のコンテキストを提供：Google 検索の About this image は、ユーザーがウェブ上で見つけた画像の信頼性とコンテキストを評価可能。Gemini の回答を再確認する機能は、Gemini の回答を裏付けるコンテンツがウェブ上にあるかどうかを評価できる機能で、現在さまざまな地域や言語で展開。

・　YouTube は、視聴しているコンテンツが大幅に改変されていたり、合成して生成されたものである場合に視聴者に知らせるパネルを表示。2024 年 2 月には、YouTube の Dream Track ツールで生成されたコンテンツに付加するパネルの第一弾を導入。このパネルは動画の説明欄に情報を追加して、そのコンテンツが AI 技術で生成されたことを視聴者に知らせる。

・　AI 生成コンテンツへのラベル付けに関する第 1 ステップとして、2024 年 2 月に、YouTube の AI ツールを使ったものに関する自動的なラベル付与を開始。第 2 ステップとして、同年 3 月に、合成又は改変されたデータが実物に近い場合にクリエイターにラベル付けを求める取組を開始。継続的にラベル付けをしないユーザーに関しては、例えばコンテンツの削除、YPP の資格停止その他のペナルティを検討。これらのペナルティはまだ一切科していない。導入したばかりなので、クリエイターにもある程度経験してもらって新しいプロセスに慣れていただく時間が必要。また、生成 AI を使うことが悪いことといったメッセージは出したくない。クリエイターが自主的に開示することを期待するが、選挙、ニュース、金融、健康などのセンシティブなトピックにまつわるコンテンツの場合は、強制的なラベル付与も実施。クリエイターが合成又は改変されたコンテンツであることを開示できるよう、テクノロジーにも注力。

・　今後数週間のうちに、次のステップを開始予定。クリエイターがアップロードする動画に改変または合成素材が

	含まれていることの開示を求める、Creator Studio の情報開示要件。これらの開示情報を示す新しいパネルは動画の説明欄に追加され、健康、ニュース、選挙、金融などのセンシティブなトピックに関するコンテンツについては、パネルを動画プレーヤーにも表示。有害性のリスクを軽減するにはパネルだけでは不十分な領域がいくつかあり、一部の合成されたコンテンツは、コミュニティガイドラインに違反する場合、パネルの有無にかかわらずプラットフォームから削除。たとえば、合成により作成された動画が、視聴者に衝撃や不快感を与えることを目的として現実的な暴力を描写している場合は、削除の対象となることがある。 ・ クリエイターが改変したコンテンツや合成したコンテンツを動画に使用した場合には、自らそれを開示することが期待されるが、上記のようなセンシティブなトピックについて議論しているコンテンツにおいて、このような開示が行われていない場合、一部の動画に対してパネルを表示させることがある。 ・ 今後数か月のうちに、特定可能な個人の顔や声などを模倣している生成 AI コンテンツや合成あるいは改変されたコンテンツの削除を、プライバシー侵害の申し立ての手続きを通じて要請できるようにする。要請された全てのコンテンツが YouTube から削除されるわけではなく、さまざまな要因を考慮して判断。たとえば、コンテンツがパロディや風刺であるか、削除要請を行った人物が本人として特定可能か、あるいは公人や著名人（この場合は、より高い基準が適用される可能性があります）を取り上げているかといったことが判断材料。 ■Meta ・ フォトリアリスティックなコンテンツが、AI を使用して作成されたものであることを理解してもらうことが重要。Meta AI 機能を使って作成されたフォトリアリスティックな画像に「Imagined with AI」というラベルを貼ることでこれを実現。他社のツールで作成されたコンテンツでもこれができるようにしたい。そのため、業界のパートナーと協力し、コンテンツが AI を使用して作成されたことを示す共通の技術標準に取り組み。これらのシグナルを検出できるようになれば、利用者が Facebook や Instagram に投稿する AI 生成画像にラベルを付けることが可能。現在、この機能を構築中で、今後数ヶ月のうちに、各アプリがサポートするすべての言語でラベルの適用を開始する予定。 ・ Meta AI 機能を使ってフォトリアリスティックな画像が作成された場合、弊社は AI が関与していることを利用者に知ってもらうために、画像上に目に見えるマーカーを付けたり、画像ファイル内に見えない透かしやメタデータを埋め込む。このように目に見えない透かしとメタデータの両方を使用することで、これらの目に見えないマーカーの堅牢性が向上し、他のプラットフォームが識別しやすくなる。これは、生成 AI 機能を構築するために私たちが取っている責任あるアプローチの重要な部分。

■TikTok Japan

・　AI 生成コンテンツの制限について、コミュニティガイドラインにおいて、実在する人物の映像または音声を含む AI 生成コンテンツを禁止。

・　AI ラベルの義務付けについて、TikTok では、AI で生成したコンテンツであることをユーザー自身が動画に表示できる「AI 生成ラベル」を開発。コミュニティガイドラインにおいて、AI で生成したコンテンツを投稿する際には「AI 生成ラベル」をつけることを義務付け。

・　検知ツールについて、TikTok は、AI 生成コンテンツやテクノロジーが進化するのに合わせて、検知機能を進化させ続けている。そのための専門家との緊密な連携も継続。

・　コンテンツの出所と信頼性に関する標準化団体 Coalition for Content Provenance and Authenticity (C2PA)と連携し、他のプラットフォームで作成された AI 生成コンテンツ にも自動的にラベル付けを開始。この取り組みは、C2PA の技術である「コンテンツクレデンシャル」機能を動画プラット フォームとして初めて実装するもの。

・　AI を使ったエフェクト機能も出しているが、これを使ったコンテンツには自動的にその旨表示。

■Microsoft

・　AIが生成した、もしくはAIによって修正されたコンテンツについて、マイクロソフトの製品を使用した場合、例えば、AIイメージクリエーターやペイントといったような製品を使うと、コンテンツの来歴というものがラベルとして貼られる。つまり、AIで生成されたものなのか、AIで修正されたものなのかといったラベルが貼られるため、それによって、どういった状況かというのが確認可能。

■X

・　合成メディアや操作されたメディアなど今後予測されるリスクに対抗するためのポリシーを保持。

・　メディアマッチングによりコミュニティノートがはるかに多くの投稿に表示。最近では、「メディアに対するノート」を導入・強化しており、写真やビデオにノートが追加されると、マッチングするメディアを含む他の投稿に自動的に表示。場合によっては、個別のノートが数百または数千の投稿と一致。

○　多くの国外事業者は、他の AI 関連事業者との間の連携・協力に向けた取組状況について、「選挙における AI の不正利用に対抗するための技術協定（ミュンヘンアコード）」への参加に加え、C2PA（Coalition for Content Provenance and Authenticity）への参加、動画や音声を含む AI コンテンツを識別するための共通

	の技術標準の策定や**ディープフェイクの検出・帰属の研究**等を回答。 ●　全ての国内事業者は、**他の AI 関連事業者との間の連携・協力に向けた取組状況について明確な回答なし**。 【他の AI 関連事業者との間の連携・協力に向けた取組状況】(参考資料22-1-1 及び22-1-2 Q10-4) **【取組［例］】** ■Google ・　C2PA(Coalition for Content Provenance and Authenticity) に運営委員会メンバーとして参加し、デジタルコンテンツの信頼性と透明性を強化し、ウェブ上の誤った情報に対抗するためのコンテンツクレデンシャルの開発と導入を支援。 ・　進化した C2PA の標準(バージョン 2.0)は、コンテンツがオンラインで拡散される際、そのコンテンツの耐用期間中、出所と加えられた編集を確実に追跡する改ざん耐性のある方法を提供。カメラで撮影された写真や動画、生成 AI モデルを使用して合成・編集されたコンテンツ、ソフトウェアで作成されたデジタルアートなど、あらゆる形態のデジタルコンテンツに等しく適用できるため、コンテンツがどのように作られ、時間をかけて編集されるのかについて、より透明性を高めることが可能。 ■Meta ・　動画や音声を含む AI コンテンツを識別するための共通の技術標準について、業界のパートナーと取り組み。2019 年には「Deep Fake Detection Challenge」を立ち上げ、世界中の人々がディープフェイクを検出するためのより多くの研究やオープンソースのツールを生み出すことに拍車。1000 万米ドルの助成金で支援されたこのプロジェクトには、Partnership on AI、コーネル大学、カリフォルニア大学バークレー校、MIT、WITNESS、マイクロソフト、BBC、AWS など、市民社会やテクノロジー、メディア、学術のコミュニティに所属する複数の組織の横断的な連合が参加。 ・　世界最大のマルチメディア・ニュースプロバイダーであるロイターと提携し、無料のオンライントレーニングコースを通じて、世界中のニュースルームがディープフェイクや操作されたメディアを識別できるよう支援。報道機関は、大量の画像や動画をサードパーティに依存することが増えており、操作されたビジュアルを識別することは重要な課題。本プログラムは、この作業を行おうとするニュースルームを支援することを目的。 ・　2021 年、ミシガン州立大学(MSU)と共同で、ディープフェイクの検出・帰属の研究手法を新たに発表。これは、AI が生成した 1 枚の画像から、その画像を生成するために使用した生成モデルをリバースエンジニアリングするもの。この方法により、ディープフェイク画像そのものが検出器の唯一の情報であることが多い実世界でのディー

	プフェイク検出とトレースが容易。
14　ファクトチェックの推進に向けた取組状況	○　一部の国内事業者及び多くの国外事業者は、**ファクトチェック関連団体との連携その他のファクトチェックの推進の観点からの取組**について、**ファクトチェック推進団体との定期的な意見交換・寄付・法人会員としての支援**、**ファクトチェック機関への資金提供**、**提供するサービスにおけるファクトチェック記事の掲載**、**第三者ファクトチェック・プログラムの提供**、**ファクトチェックツールの提供**、**ファクトチェッカー育成とリテラシー向上のための講座開催**等を回答。また、一部の国内事業者より、**具体的な情報がないとモデレーションできず、またモデレーションしたとしても事後的にファクトチェック結果の解釈が正しかったのかということも問われうる中、あるファクトチェック結果に基づき、どういった範囲の投稿が削除の対象となり得るかは、プラットフォーム事業者で解釈がまちまちになるという状態もあまり望ましくはなく、ある種の相場感みたいなものを形成していくことも重要であること**、**ファクトチェック機関との具体的な連携方法について情報がなくどのように取り組めば良いのかが不明であること**、**連携に当たっての具体的な方法やファクトチェックの団体自体の信頼性がちゃんと検証されているのか等でまだ不安があること**や、**体制が少人数のため協力内容によっては担当者を手配しづらいこと**等を回答。 ●　ほぼ全ての国内事業者及び一部の国外事業者は、**ファクトチェック関連団体との連携その他のファクトチェックの推進の観点からの取組について明確な回答なし**。 【ファクトチェックの推進に向けた取組状況】（参考資料22-1-1 及び22-1-2 Q11、同22-2-1 及び22-2-2Q1、同22-5-1Q8 及び22-5-2Q12） **【取組［例］】** ・ 連携相談や活動内容の共有などについて FIJ と定期的な MTG を実施。 ・ ファクトチェック支援団体である FIJ の活動に賛同し、寄付の実施や法人会員として支援連携。 ・ Disinformation 対策フォーラムの議論に参画し、産官学民の連携を実践。 ・ 日本ファクトチェックセンターの設立にあたって資金提供を行い、22 年 11 月からは制作されたファクトチェックコンテンツをヤフーニュースへ掲載。 ・ 日本ファクトチェックセンターのファクトチェック結果に基づいて削除対応を実施。 ・ 2022 年 2 月に、ユーザーが偽情報や誤情報などの情報に惑わされず、ニュースを正しく理解するための学習コンテンツ「Yahoo!ニュース健診」を公開。 ・ Yahoo!ニュースとして日本ファクトチェックセンターと情報提供契約を締結。2022 年 11 月からは制作されたファク

	トチェックコンテンツをパートナー社として記事配信、配信コンテンツについては一覧として公開。 ・ ファクトチェック機関「日本ファクトチェックセンター」の設立：2022 年 9 月に Google.org によるセーファーインターネット協会に対する 150 万米ドルの支援。国際ファクトチェックネットワークの基準に沿ったファクトチェックを実施し、各地域で悪影響を受けているコミュニティのメディアリテラシーの知見を高めることに焦点。 ・ 日本の情報空間における偽情報・誤情報の動向、特にインターネット上の偽情報・誤情報のパターンや手法を分析するための調査を実施。 ・ メディアリテラシー研修として、誤った情報に騙される被害を防ぐため、重要なテーマに関する基礎知識を普及、偽情報・誤情報の手法や典型的なパターンについて啓発するコンテンツを発信、ファクトチェッカー育成とリテラシー向上のための講座を開催。 ・ その他幅広い偽情報・誤情報対策の推進として、リテラシー教育とオンライン上の安全性とその重要性に向けた啓発活動を推進、シンポジウムの開催など。 ・ ファクトチェックのラベルは Google 検索及び Google ニュース結果に表示されることもあり、ユーザーがコンテンツのどの部分がポジティブにファクトとしてチェックされているかが確認でき、より詳細を調べたい場合も簡単により情報を探すことが可能。これらの機能は、第三者パブリッシャーのオープンなネットワークに依存。 ・ Google は、ファクトチェックエクスプローラー及びファクトチェックマークアップツールの2つのツールで構成されるファクトチェックツールを提供。これらのツールはファクトチェックを行う人、ジャーナリスト及び研究者の作業を容易にすることを目指す。例えば、ファクトチェックマークアップツールの目標は、記事そのものには手を加えず、シンプルなウェブフォームを通してマークアップに提出できるようにすることにより、クレームレビューマークアップを作る過程を容易にする。2023 年 6 月、ファクトチェックエクスプローラーにあらゆる画像の URL をアップロードまたはコピーして、その画像が既存のファクトチェックで取り上げられているかどうかを確認できるグローバルベータ版を公開。このバージョンでは、さらに、画像に関連するさまざまなコンテキストと、時間とともに加わった変化についても概観することが可能。 ・ Google は、偽情報対策に欠かせないユースケースである画像の出所の確認ができる、オープンソースでモバイル向けに最適化された Storyful 社の Source の開発もサポート。Source は、Google Cloud の技術を用いて画像の潜在的履歴を迅速に分析、オブジェクトや言語を認識し、自動翻訳を提供。このツールは、ファクトチェッカー、ジャーナリスト、偽情報について研究する学者に利用。

	・ ファクトチェッカーが改変された画像のオリジナルを見つけるために、Google の画像検索が利用されているほか、Google マップや Google Earth などの地形情報も情報検証に欠かせないツールとして利用。また、セミナーを開催したり、こうしたツールを使ってファクトチェックを行う方法の解説動画も提供。 ・ 質の高い信頼できる情報を優先的に表示する一方、偽・誤情報であり得るものにはフラグをするため、ファクトチェック組織・コミュニティをサポート。 ・ YouTube でビデオファクトチェックを容易にできるよう投資。 ・ 2023 年 6 月、世界中のファクトチェッカーを招待し、ファクトチェックビデオの作り方など共有。 ・ 2022 年後半、ポインターインスティチュートに 1300 万米ドルの助成金支出。 ・ アジア太平洋で様々なファクトチェック組織が連携を取れるよう支援 ・ 2024 年に第三者ファクトチェック・プログラムを日本に拡大する予定。International Fact-Checking Network (IFCN)の認定を受けた日本におけるファクト・チェック団体を対象に、プログラムへの参加を呼びかけ。プログラムの仕組みとして、独立したファクトチェック団体に対して、一次情報源へのインタビュー、公開データの参照、写真やビデオを含むメディアの分析を含む独自の取材を通じて、記事の正確性を検証し、評価することを委任。ファクトチェック団体が、送信したコンテンツの一部を評価した場合、そのコンテンツの配信を削減し、ラベルを付け、それを見た可能性のある他の利用者に通知。 ・ ファクトチェック団体が新しいスキルを開発し、イノベーションを追求し、オンライン上の誤情報に対処するための取り組みを拡大できるよう支援することで、業界をサポート。 ・ あらゆるプラットフォームの中で最大のグローバル・ファクトチェック・ネットワークを構築し、2016 年以来、ファクトチェックの取り組みを支援するプログラムに 1 億ドル以上を拠出。これには、スポンサーシップ、フェローシップ、助成金プログラムなどの業界イニシアチブだけでなく、弊社のプラットフォームで活動するファクトチェッカーへの直接支援。また、法的支援基金の支援など、危機的状況にあるファクトチェック団体を支援するために多額のリソースを投入。 ・ ファクトチェックに関わる審査員のトレーニング ・ 誤情報関連のコンテンツトレンドの定期的なチェック ・ 現在連携している米国ベースのファクトチェック機関(Lead Stories)との詳細のやり取りについては非公表。ただ、日本のコンテンツモデレーションにおいても Lead Stories に確認するなど実施。

	・　日本ファクトチェックセンターやセイファーインターネット協会等と、双方の取り組みや問題意識の共有等を目的とした意見交換を実施。たとえば、日本での持続可能なファクトチェック　エコシステムの構築方　、テクノロジ企業の貢献方法、ファクトチェック機関との連携方法等について話し合い。 ・　コミュニティノートの提供。 ○　一部の国外事業者は、**第三者からの通報対応におけるファクトチェック機関の関与**について、**第三者ファクトチェックパートナーを導入している国ではパートナーに審査のために送るコンテンツの種類を決定するためのシグナルとして活用**していること、**第三者通報への対応でファクトチェックが必要な場合はファクトチェック機関と連携**していることを回答。 ●　多くの国外事業者は、**第三者からの通報対応におけるファクトチェック機関の関与について実施していない又は明確な回答なし**。 ●　全ての国内事業者は、**モデレーション等の対象となった発信者からの苦情等対応におけるファクトチェック機関の関与や第三者からの通報対応におけるファクトチェック機関の関与について実施していない旨を回答**。 【モデレーション等の対象となった発信者からの苦情等対応や第三者からの通報対応におけるファクトチェック機関の関与状況】(参考資料22-1-1 及び22-1-2　Q4-3(3)、Q5-2(3)) ●　全ての国内事業者及び国外事業者は、**2022 年中及び 2023 年中に実施した偽・誤情報に対するモデレーション等件数のうちファクトチェック機関の意見等を反映し実施した件数**について、**実施していない旨、該当がない旨や非公開である旨を回答**。 【2022　年中及び　2023　年中に実施した偽･誤情報に対するモデレーション等のうちファクトチェック機関の意見等を反映して実施した状況】(参考資料22-1-1 及び22-1-2　Q4-3(3)、Q6-1(5)) ○　一部の国外事業者は、**令和 6 年能登半島地震関連の偽・誤情報への対応におけるファクトチェック機関との連携**について、**ファクトチェック機関と連携**していた旨を回答。 ●　全ての国内事業者及びほぼ全ての国外事業者は、**令和 6 年能登半島地震関連の偽・誤情報への対応におけるファクトチェック機関との連携**について、**実施していない旨や明確な回答なし**。 【令和 6 年能登半島地震関連の偽･誤情報への対応におけるファクトチェック機関との連携状況】(参考資料22-1-1 及び22-1-2Q7-4)
15　マスメディア	○　一部の国内事業者及び国外事業者は、**マスメディア（新聞・放送）との日本国内における連携**について、

（新聞・放送）との連携状況	**ニュースサービスにおける信頼できる質の高い情報として連携体制の構築、メディア関係の業界団体や学会への参加・意見交換、記者向けファクトチェック・デジタルツール講習会の開催**等を回答。 ● ほぼ全ての国内事業者及び国外事業者は、**マスメディア（新聞・放送）との日本国内における連携**について、**実施していない旨や明確な回答なし**。 【マスメディア（新聞・放送）との連携状況】（参考資料22-1-1 及び22-1-2Q12） **【取組［例］】** ・ アライアンスを専門に行う組織を設置し、日頃から各提携媒体と密に連携。特に影響の大きなメディアとは定期的に情報交換をし、情報の信頼性に関する話題を取り上げ。アライアンスチームが信頼できる各種メディアと契約し、情報の正確さ・信頼性、その裏付けとなる取材体制について一定の担保がなされているほか、契約後においても、配信記事の当社ガイドラインに対するモニタリング、ファクトチェックも含む校正校閲等を実施しており、偽情報の掲載自体を抑止するよう取り組み。 ・ 信頼できる質の高い情報を増やしていくため、伝統的メディアやネットメディア等と連携体制を構築。質の高い情報を増やすための施策として、適宜媒体社の配信内容に関する審査を行ない、ガイドライン等に照らして改善いただきたい点についてお伝え、公共性・公益性が高いテーマや社会課題について媒体社とともに記事制作を実施、ユーザーの課題解決に資するため公共性の高い情報をサービスの最も目立つ場所である Yahoo!ニュースト ピックスやトピックス詳細ページに掲載。 ・ 偽情報等の課題に限らないがメディアの業界団体が一堂に会する「マスコミ倫理懇談会全国協議会」（2022 年 9 月）に参加し、「インターネット上の報道をめぐる諸課題を考える分科会に発表者として参加、偽情報等の課題に限らないがメディア学会の秋季大会（2022 年 11 月）のワークショップに参加し信頼される情報空間をつくるためのヤフーの取り組みを発表し意見交換、マスコミ倫理懇談会等において引き続き偽情報を含め近時の課題と対策等の情報共有・議論。 ・ 日本経済新聞社との取り組みで、中高生のリテラシー底上げを支援する「日経電子版 for Education」に協賛。 ・ 複数の報道機関等（読売新聞社、共同通信社、TBS 及び JNN 系列のテレビ各社、日本記者クラブ、朝日新聞社）に対して個別に記者向けファクトチェック・デジタルツール講習会を実施。 ● 全ての国内事業者及び国外事業者は、**第三者からの通報対応におけるマスメディア（新聞・放送）の関与について実施していない又は明確な回答なし**。なお、一部の国外事業者は、**報道機関が自らアカウントを開**

	設した場合、報道の自由・表現の自由の関係で意図的にタッチしないという整理と回答。 【モデレーション等の対象となった発信者からの苦情等対応や第三者からの通報対応におけるマスメディア(新聞・放送)の関与状況】(参考資料22-1-1 及び22-1-2 Q4-3(3)、Q5-2(3)、同22-2-1 及び22-2-2Q1) ●　全ての国内事業者及び国外事業者は、**2022 年中及び 2023 年中に実施した偽・誤情報に対するモデレーション等件数のうちマスメディア（新聞・放送）の意見等を反映し実施した件数**について、**実施していない旨、該当がない旨や非公開である旨を回答**。 【2022 年中及び 2023 年中に実施した偽・誤情報に対するモデレーション等のうちマスメディア(新聞・放送)の意見等を反映して実施した状況】(参考資料22-1-1 及び22-1-2 Q6-1(6))
16　利用者の ICT リテラシー向上等に向けた取組状況	○　多くの国内事業者及び国外事業者は、**小中学校や大学等の教育機関、安心なインターネット環境づくり等に関する普及啓発機関等と連携・協力した取組を実施**と回答。また、一部の国外事業者より、インターネットを取り巻く環境は日々変化しており、その変化に対応するために必要なクリティカルシンキングやスキルを身につける機会を提供することが重要である一方、教育現場でそのようなスキルを教えられる教員や保護者の数は限られており、このような取り組みを行う一部のファクトチェック団体を含む非営利団体は、財政的に不安定な傾向があることから、このような取り組みには、指導者のための研修の機会や財政的な支援が必要との回答。 ●　全ての国内事業者及び国外事業者は、**消費者団体・利用者団体と連携・協力した取組に関する明確な回答なし**。 【利用者の ICT リテラシー向上に向けた取組状況】(参考資料22-1-1 及び22-1-2 Q13－1～13－3) **【取組［例］】** ■教育機関との連携 ・　「インターネット上でのコミュニケーション」と「対面のコミュニケーション」の違いを子どもたちに学んでもらうためのオリジナル情報モラル教材を開発するとともに、全国の学校や自治体に LINE が講師を派遣するワークショップ授業・講演活動等を 2012 年より全国で開始。 ・　GIGA スクール構想の展開にあわせ、「情報モラル」と「情報活用」の育成や向上を図るため、2022 年 7 月、新たな活用型情報モラル教材「GIGA ワークブック」の汎用版を開発し、同年 9 月から、全国の小中学校で活用いただけるよう無償提供を開始。教材を導入いただいた自治体と連携し、学校現場での教材活用のサポートとして教員の方々へのオンライ ン研修(無償)も開始。

	・　LINE みらい財団におけるユーザ啓発活動（GIGA ワークブック）として、啓発教材（GIGA ワークブック）を導入することとした自治体（教育委員会）を通じて、地域内の全ての公立学校（小学校・中学校・高校）へ教材導入の推進を案内。 ・　大学と連携し、リテラシー教育授業を実施。 ・　中学生・高校生にオンライン・リテラシー・カリキュラムを提供。このカリキュラムでは、ファクトチェックを含む主要なオンラインリテラシーのトピックを扱っています。これまでに約 10 万人の中高生に教材を提供。 ・　中高生 15,557 人と、中高生の教員 119 名を対象にインターネット利用について調査し、この結果を「中高生インターネット利用白書 2021 」として公開。調査では、利用時間や目的、インターネット利用で感じるメリットやデメリット、そして実際に経験したトラブル等について調査。 ・　YouTube では、誤情報・フェイクニュースに関する取り組みとして、総務省並びに国際大学 GLOCOM のご協力の元、2023 年 4 月に「ほんとかな？があなたを守る」というテーマのキャンペーンを実施。ユーザーに向けて、フェイクニュースが自分の日常に潜む問題であると気付くきっかけを作ること、そして、情報との向き合い方について考える機会を提供することを目指し、若者層に人気の高い 9 組の YouTube クリエイターの協力を得て、3 つのメッセージ（フェイクニュースは身近に存在すること」、「ファクトチェックが重要であること」、「安易な拡散が人に迷惑をかけてしまうリスクに繋がりかねないこと」）を伝えるショート動画を作成。 ・　学校・教育委員会や、自治体が行う青少年向けのカンファレンス、財団が行う研究事業などにおいて、パネリストや講演者として参加し、ICT リテラシー向上に向けた啓発活動。 ・　マイクロソフトは Future Learning Lab プロジェクトのメンバーとして GLOCOM（国際大学グローバル・コミュニケーション・センター）とも連携し、FuLLのイベント等にスピーカーを派遣するなど、関係者との意見交換・連携強化に向けたディスカッションに参加。また、FuLLの取り組みとして、児童（小学校中学年・高学年）がオンライン上で自己のデータを適切に管理するために注意すべきこと等を学べる学習コンテンツを作成。作成されたコンテンツは、経済産業省のSTEM Libraryに掲載。 **■普及啓発機関との連携** ・　「安心ネットづくり促進協議会」への参加。 ・　高校生 ICT カンファレンスに参加し、各地の高校において、ICT リテラシー向上に関する講演。 **■その他**

	・　ウェブサイト上に LINE Safety Center −LINE の安心安全ガイド−を掲載。 ・　LINE みらい財団において、教育工学や授業デザインを専門とする研究者と共同で、独自の情報モラル教育教材の開発を行い、ウェブサイトで公開。 ・　LINE みらい財団において、情報モラル・情報リテラシーの啓発活動の強化やネッ トトラブル防止を目的に、地方公共団体や専門家と協力しながら調査研究・教材作成等を実施。 ・　一般社団法人セーファーインターネット協会主催の「Disinformation 対策フォーラム」へ 参加し、有効な対策について多様なプレイヤーと議論。 ・　媒体社と連携し、ファクトチェック記事の配信を拡充する取り組みを実施。 ・　有識者とフェイクニュース対策について議論、対策コンテンツ制作の助言を受ける活動を実施。 ・　総務省がセーファーインターネットデーに合わせて公開したリテラシーコンテンツ紹介サイト「ネット＆SNS よりよくつかって 未来をつくろう～ICT 活用リテラシー向上プロジェクト～」に協力。 ・　Grow with Google という私たちの無料のデジタルスキルトレーニングを提供するプログラムを通じて、ユーザーのインターネットリテラシー教育を強化するために、無料のオンライン講座を提供。家族で実践安心・安全なインターネット利用のためにできることでは、家族を中心としたユーザーを対象に、情報の正確さを確認するためのヒントを含むオンライントレーニングを提供。あわせて個人で実践編も公開。 ・　こどもたちがテクノロジーを安全に、かつ最大限に活用できるようになるために、お子様向けインターネットリテラシープログラム Be Internet Awesome（日本語版）を公開。子どもが楽しみながら自発的に学べるオンラインゲーム「インターランド」のほか、保護者と教師向けのさまざまなリソースを含む。 ・　有識者が主催する全国こどもネットフォーラム 2023 を後援し、子どもがインターネットの利用について自ら議論し学ぶ場を設けることを支援。 ・　Google が主幹事となり、あらゆる方のスキル開発を支援する取り組み「日本リスキリングコンソーシアム」を立ち上げ。総務省、経済産業省の協力のもと、200 を超えるパートナーとともに様々なリスキリングプログラムを提供。 ・　Google では、外部専門家の知見を取り入れる形で Grow with Google 「はじめてのメディアリテラシー− 情報と向き合うとき、子どもも大人もすべきこと」として 10 本のトレーニング動画を公開している他、5 組の YouTube クリエイターの協力も得てより広い層に訴求する啓発動画を制作。 ・　2020 年より、NPO 法人企業と教育協会（ACE）と共同でデジタルリテラシー教育プログラム「みんなのデジタル教

	室#wethinkdigital」を実施。「フェイクニュースの見抜き方」「デジタルアイデンティティを考える」の 2 つのモジュールで、これまでに全国 24000 人以上の中高生を対象に授業を実施。 ・ 2023 年 7 月、弊社は既存の 2 モジュールのリニューアルを発表。誤情報に関する総務省のリテラシー資産を統合し、新モジュール「デジタルシチズンシップと情報発信」を導入。このプログラムは総務省情報通信局の支援。 ・ 総務省と共同で政策討議: 2022 年以降の ICT 利活用のためのリテラシー向上に関する研究会。ICT 利活用のためのリテラシー向上プロジェクトに貢献。 ・ 専門家やクリエイターと連携して、多様な啓発活動を実施。例えば、サイバーセキュリティ月間、若年層の性暴力被害予防月間、自殺予防週間、感染症ワクチン啓発、研究者と連携した「TikTok クリエイター向けメンタルヘルス講習会」、TikTok 悩み相談「性の悩み」「不登校」「親との関係」、選挙に関するデジタルリテラシーキャンペーン、親子向け啓発イベント、ペアレンタルコントロール利用の手引きの公開、クリエイターと連携した、安心安全のための啓発動画の制作。 ・ Microsoft は、育て上げネットなどの国内外の非営利団体と協力して、女性やサービスの行き届いていないコミュニティの若者など、日本社会のメンバーに責任ある AI とサイバーセキュリティを促進するためのコンテンツとスキルを提供。 ・ LinkedIn は、一般社団法人 1mm Innovation と協力して、女性が自信を取り戻し、雇用の障害を乗り越えて、経済的な安全と自立を確保できるよう支援するプログラムを策定。 ・ マイクロソフトは、ウェブサイトでリテラシー学習教材を提供。たとえば、"デジタル セーフティ" に関する専用サイトを公開し、教師が授業で使える教材を提供。"ゲーム安全性ツールキット" も提供しており、安全にゲームをする方法や Xbox の安全性に関する機能の使い方について親子で学ぶことが可能。 ・ マイクロソフトは、自社のプラットフォームやパートナーシップを活用し、信頼できる情報を検索、利用、共有する方法について消費者に情報を提供。検索の練習と検索コーチは Teams for Education に組み込まれている無料のツールで、効果的な検索条件を指定し、信頼できる情報源を特定して確認する方法について、学生にリアルタイムのコーチングを実施。The Investigators は教育版マインクラフトの新しい世界で、ゲームを利用した学習を通じて学生が情報リテラシーとメディア リテラシーを身に付けられるよう支援。このゲームと補助教材は英語で提供が開始されており、2024 年には (日本語を含む) 28 か国語で公開され、世界中で数百万人の学生と教師が利用できるようになる予定。

	・　生成 AI を責任をもって安全に活用するためのリテラシー習得を目的として、中高生を対象とした「AI Classroom ツールキット」や、教師や教育者等を対象とした「AI Toolkit for Trainers」などを提供。 ・　LinkedIn サイバーセキュリティの基礎知識 by Microsoft x LinkedIn：サイバーセキュリティの基礎知識 by Microsoft x LinkedIn を通じたリテラシー教材も提供。
17　研究開発の推進等に向けた取組状況	**○　ほぼ全ての国内事業者及び全ての国外事業者は、研究開発の推進に向けた取組について、偽情報等の流通状況や拡散の仕組みに係る分析、日本におけるフェイクニュースの実態と対処策の研究、海外でのフェイクニュース事例・諸問題などの分析、ジャーナリズムにおけるデジタル技術の活用方法に関する研修やフェイクニュースに対するリテラシー教育授業の実施等を回答。** **●　一部の国内事業者は、研究開発の推進に向けた取組に関する明確な回答なし。** 【研究開発の推進に向けた取組状況】（参考資料22-1-1 及び22-1-2 Q14-1、14-2、14-6） **【取組［例］】** ・　研究者・官公庁・業界団体等とともに、偽情報等の流通状況や拡散の仕組みに係る分析や、プラットフォーム事業者に求められる偽情報等への対応に係る検討を実施。 ・　一般社団法人ソーシャルメディア利用環境整備（SMAJ）において、偽誤情報を含むソーシャルメディア上の諸課題に関する検討に参加。 ・　2022 年 12 月、一般社団法人マスコミ倫理懇談会全国協議会の「ネット空間における倫理研究会」において、LINE NEWS における取組み（釣り見出し、コタツ記事、自殺報道）を発表。 ・　海外でのフェイクニュース事例・諸問題など外部シンクタンクに協力いただき分析調査を実施し、専門家よりインプットを取得。 ・　「デジタル時代における民主主義を考える有識者会議」を開催し、デジタルプラットフォームが民主主義に及ぼす影響やデジタルプラットフォーム事業者に期待される役割等について、フェイクニュースなどの例をもとに検討を進める有識者会議を開催。構成員として、マスメディアの方から学者まで、幅広い分野の専門家をお招きし検討を実施。 ・　有識者とフェイクニュース対策について議論、対策コンテンツ制作の助言を受ける活動を実施。 ・　大学と連携し、フェイクニュースに対するリテラシー教育授業を実施。22 年参院選に向けたリテラシー特集を制作。デマに惑わされないよう情報摂取の注意喚起コンテンツを提供。有識者による解説動画を 3 本制作し公開。

	・ 産業技術総合研究所メディアインタラクション研究グループが、人々の音楽の聴き方をより豊かにすることを目的に、「Songrium」の研究開発を行っているところ、お知らせの掲出などにより、こういった取組のユーザー認知向上に協力。 ・ 2023 年、従業員の知人（経済産業省に所属しており、IPA に出向中の方）からの要請を受け、「web プラットフォームと肖像権」をテーマとした論文作成にあたり、「権利侵害による削除申請件数」等の情報を提供。論文は東京大学先端科学技術研究センターに提出予定。 ・ 現状では配信内容についての情報収集をしていないが、将来的に、配信番組の音声認識や視聴者コメントの内容から、キーワードや主張内容を抽出し、その内容に対して適切な関連情報（内部データベースもしくは信頼できるドメインを対象にしたインターネット検索等による情報）を取得し、参考情報として提示するようなことができるのではないかと検討。元々は「放送者への話題の自動提供（ネタ振り）」を目的として考えていましたが、信頼できる情報の提示を行うことができれば、偽・誤情報の拡散防止にも繋がるのではないかと考え中。 ・ ピグパーティの場合は、「未成年事案に対するメディアの健全化」の取り組みを実施。専用の体制を立てて研究機関と連携して未然防止策の実施や、ユーザーの啓蒙などを実施。 ・ 国際大学 GLOCOM の研究プロジェクト Innovation Nippon を支援。Innovation Nippon では、2019 年以降、4 年連続で日本におけるフェイクニュースの実態と対処策について研究。 ・ Google News Lab は、日本の News Lab フェローの協力のもと、ジャーナリストや大学生を対象に、偽情報や誤った情報の識別、検索の有効活用、地図や衛星画像の活用など、ジャーナリズムにおけるデジタル技術の活用方法に関する研修を実施。 ・ Facebook や Instagram のようなソーシャルメディアアプリが世界に与える影響を理解するためには、厳密で独立した調査をサポートすることが重要。米国 2020 年調査のような公益調査をサポートするツールを提供するなど、オープンかつプライバシー保護された調査アプローチに長年取り組み。 ・ 大学と連携して、「日本の SNS を起因とした児童の性犯罪に至るオンライン上でのコミュニケーションとプロセスに関する研究」を実施。 ・ 2023 年 8 月、マイクロソフトと東京大学は、日本の社会と未来に貢献する学生の育成を目的として、グリーントランスフォーメーション、ダイバーシティ & インクルージョン、人工知能の各分野の研究推進に向けて連携するための基本合意書を締結。

	○　半数の国内事業者及び一部の国外事業者は、研究機関や研究者等向けの日本国内におけるデータ提供について、**API による提供ではない**が、**投稿自体（記事、掲示版、ブログ、質問、回答、コメント）**や**メタデータ（動画、コンテンツ視聴ログ）**等を回答。 ○　多くの国外事業者は、**研究機関や研究者等向けの日本国内におけるデータ提供**について、**グローバルな動画メタデータへの API による大規模かつ幅広いアクセス**、**SNS のページ・投稿・グループ・イベント・クリエイターやビジネスアカウント・ほぼリアルタイムで公開されているコンテンツ・リアクション数・シェア数・コメント数・投稿の閲覧数等への API 等によるアクセス**や、**無料・ベーシック・プロ・エンタープライズを含む API による 4 つのアクセスレベル**の提供と回答。 ●　半数の国内事業者及び一部の国外事業者は、**研究機関や研究者等向けの日本国内におけるデータ提供**について、**実施していない**旨や、**欧米では非常に限られた研究者のみに API により公開しているが他地域では公開していない**旨を回答。なお、**研究機関や研究者等向けの日本国内におけるデータ提供における支障となっている点や課題**として、**個人情報**、**著作権**、**提供する目的やデータの位置づけ**、**データの更新**、**API システムの維持管理**等を回答。 ●　一部の国外事業者は、**日本国内の研究者への API によるデータ提供の可能化に係る具体的な条件について明確な回答なし**。 ●　一部の国外事業者は、**日本国内の研究者への API 公開の提供先等の実績**について、**研究者との契約情報は秘密保持契約により共有できない**旨を回答。 【研究機関や研究者等向けの実データのAPI等による提供状況】(参考資料22-1-1 及び22-1-2 Q14-3～14-5、同22-2-1 及び22-2-2 Q20、同22-5-1Q7、同22-5-2Q10) **【取組［例］】** ・ 偽情報対策を含めた研究用データとして、知恵袋のデータベースからランダムサンプリングにより抽出した解決済みの質問(約 247 万件)と、それら各質問に対するすべての回答約(約 649 万件)について、投稿者の Yahoo! JAPAN ID を暗号化するなど、個人を特定することができない情報に処理したうえで国立情報学研究所(NII)を通じて研究者に対し提供。 ・ 国立情報学研究所情報研究データリポジトリにて、ニコニコデータセットとしてニコニコ動画のコメントデータと動画メタデータ、ニコニコ大百科の記事データと掲示板データを研究目的として研究者向けに無償で提供。

	・ 産業技術総合研究所の技術・サービス等とニコニコが持つデータ・サービス等を活用した新機能の実用化や商用利用の実証実験のため、コンテンツ視聴などの匿名化されたログデータ等を提供。これにより、コンテンツの推薦 API 等が作成。 ・ ブログ上のコンテンツは全て web 上に公開されておりアクセス可能になっており、研究目的で活用いただいているケースは存在。 ・ YouTube Researcher Program は、認定された高等教育機関に所属する学術研究者に対し、データ API を通じて、公開されている YouTube コーパス全体のグローバルな動画メタデータへの大規模かつ幅広いアクセスを提供。資格がある研究者は、YouTube データへのアクセスを申請して、偽・誤情報をはじめ各自の専門分野に関するトピックについて研究可能。 ・ 透明性を維持するための取り組みの一環として、Google で受領した個々の法的通知について、その写しが Lumen プロジェクトに送付され公開。Lumen とは、ハーバード大学法科大学院バークマンセンター（Berkman Klein Center for Internet & Society）が運営する独立調査プロジェクト。Lumen には研究者向けのツールがあり、さまざまな研究活動に従事する研究者によって利用。 ・ Google Trends も研究者にとって貴重なツールであり、Google News Initiative を通じて、Trends データの活用、理解、視覚化に役立つレッスンを自由に利用可能。 ・ ClaimReview マークアップを使用しているファクトチェックの記事を誰でも閲覧できる Google ファクトチェックエクスプローラーという製品も提供。Google FactCheck Claim Search API も提供しており、ファクトチェックエクスプローラーで入手できるものと同じ結果を照会可能。 ・ データセット、調査、地図、API を構築することで、研究者がプラットフォームの政治的、経済的、社会的影響を研究可能。ツールやプロセスは、研究者が研究をサポートするための情報や分析機能にアクセスできるよう支援。Meta コンテンツライブラリーと API ツールは、Facebook のページ、投稿、グループ、イベント、そして Instagram のクリエイターやビジネスアカウントから、ほぼリアルタイムで公開されているコンテンツへのアクセスを提供。リアクション数、シェア数、コメント数、そして今回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・ユーザー・インターフェース（UI）またはプログラム API の両方で、コンテンツを検索、探索、フィルタリング可能。これらのツールを組み合わせることで、Facebook と Instagram で公開されているコンテンツへの最も包括的なアクセスが可能。

	・ 科学的あるいは公益的な研究テーマを追求する資格のある機関の個人は、ミシガン大学の Inter-university Consortium for Political and Social Research を皮切りに、研究のための安全なデータ共有に深い専門知識を持つパートナーを通じて、これらのツールへのアクセスを申請することが可能。これは、研究者が ICPSR の Social Media Archive（SOMAR）の Virtual Data Enclave の API からデータを分析することを可能にする初めてのパートナーシップ。 ・ Raj Chetty とハーバードの Opportunity Insights Program と共同で、フェイスブック上の 210 億人の友人関係からの情報を使って、米国における経済的流動性の促進要因を測定する画期的なソーシャル・キャピタル研究を発表。社会経済や学校に関する一般に入手可能なデータだけでなく、ソーシャルネットワークの力学に関するプラットフォームからの洞察を使用することにより、世界中の経済的移動の要因をよりよく理解するために、ハーバード大学とこの研究プログラムを拡張。Behavioural Insights Team、Royal Society of Arts、Stripe Partners、Neighbourly Lab の専門家と共同で、イギリス全土の階級を超えた友人関係を調査する予定。さらに多くの国に拡大するだけでなく、事業創出、大学進学、就職など、社会的つながりが経済的機会に果たす役割についてもさらに調査を進める予定。社会的なつながりが人々にどのような利益をもたらすかを調べる研究を土台に、社会的ネットワークがどのようにコミュニティが危機から立ち直り、避難民や移民を助けるかを引き続き研究。 ・ 透明性･説明責任情報公開センターでの活動を通じて、研究者への積極的な情報公開を実施。 ・ API は欧州・米州で先行して公開しているが、現時点では非常に限られた研究者のみという公開になっており、日本も含むそれ以外の地域に関しては現在検討中。 ・ API は研究者を含む誰でも申し込むことが可能。過去 1 年間で利用可能なサービスの範囲を拡大し、無料、ベーシック、プロ、エンタープライズを含む 4 つのアクセスレベルを提供。 ・ サービスの特定の側面をオープンソースにし、レビュー可能。例えば、X はおすすめのタイムラインのアルゴリズムを公開。また、コミュニティノートのアルゴリズムとデータも公開。コミュニティノートのデータは、アーカイブという形で、研究者にも公開。活用いただくことが可能。アルゴリズムのソースコードについても GitHub から、ダウンロードしてアルゴリズムを利用可能。アルゴリズムで、こういう形でダウンロードしていただいて利用可能という体制にしているのは、恐らく我々だけではないのかと自負。 **【課題等［例］】**

	・ 投稿データは個人情報に該当することから外部の第三者にデータを提供することが困難。 ・ 投稿データ以外のデータについては著作権者から契約上の許諾を得て掲載しているものも多数含まれることから、それら著作権者ごとに許諾の取得が必要。 ・ データの開示に関しては、それが透明性・説明責任の問題なのか、あるいは政策的に検討を行うに当たってある種の公共財的な形でデータを提供してほしいということなのか、これは恐らく個別具体の項目によってそれぞれ異なるところ、それらを峻別しながら丁寧に議論していくことが必要。 ・ 実務の立場からは被調査者側の負担にも配慮いただき、個別の項目ごとに問題意識を明らかにしていただくほうが、結果としてはより充実したデータを得られることになるのではないか。 ・ 研究者、専門家が御覧になるとまた違った視点もあるかもしれず、また、ほかのサービスも提供していることから、具体的な要望についてもし相談があれば、それを踏まえて検討。 ・ データがどのように位置づけられるのかがもう少しはっきりしてくれば、データの提供に向けてよりよい環境が整ってくる。 ・ 各研究者へのデータ配布はデータリポジトリ側で行う形式としているため、データの更新を迅速に反映させることが困難。 ・ 更新対応を高頻度で依頼することは現実的ではないため、自前システムに研究用 API を用意し直接データをダウンロード可能とする方が迅速に反映可能だが、その場合、API を含めたシステムの維持管理が必要。 ・ 研究の内容として公益に資するものになっているか、個人情報を除いたうえで有益な情報を提供できるか、ブログ文章などは内容から本人との紐づけや類推が可能なためより慎重さが必要。 ・ 作業量や、コンピューティングリソースの消費量のような課題感もあり、API として広く公開するよりは、研究用途や必要な情報について相談いただき、個別に対応するのが現実的。 ・ ユーザーのプライバシー権を保護し、データへのアクセスが悪質な行為を行う者によってユーザーに損害を与えるような方法で不正利用されないようにするための十分な保護措置が含まれるべき。これをどのように実践するかについての合意は限定的にしか得られていない。
18　サイバーセキュリティ関係機関等との連携状	○　一部の国内事業者及び国外事業者は、**偽サイトや偽アカウントに関する情報共有等**において、**日本国内のサイバーセキュリティ関係機関等と連携・協力した取組を実施**と回答。 ●　全ての国内事業者及び国外事業者は、**偽・誤情報の流通・拡散への対応という観点において、日本国内の**

況	**サイバーセキュリティ関係機関等と連携・協力した取組に関する明確な回答なし。** 【サイバーセキュリティ関係機関等との連携状況】(参考資料22-1-1 及び22-1-2 Q15－1～15－3) **【取組［例］】** ・ 関係団体(日本サイバー犯罪対策センター(JC3)やフィッシング対策協議会)へ会員として参加。対象サービスに限らず、当社に関わるフィッシングサイト情報や、偽サイトを含む脅威情報の提供を受領。また、当社で発見したフィッシングサイト情報については協議会経由で JPCERT/CC へ共有を行い、サイト閉鎖に向けた対応。 ・ 社内での調査、社内外からの CSIRT 窓口への通報により、当社に関連する SNS 等の偽アカウントの発見することもあり、偽・誤情報の流通・拡散への対応のため、停止に向けた措置を実施。 ・ IPA から脆弱性情報の共有(偽情報対策という面において特に連携はしていない) ・ 偽・誤情報の流通・拡散への対応に特化した取り組みではないが、CyberAgent CSIRT の活動として日本シーサート協議会、 FIRST (Forum of Incident Response and Security Teams)に加入して情報収集。 ・ 外部のインテリジェンス専門機関と、サイバーセキュリティに関するモニタリング及び情報提供において連携。この連携には、TikTok のコミュニティを欺いたりする可能性のあるアカウント(例:選挙の信頼性に悪影響を及ぼす可能性のあるアカウント)を探知するための協力などを含む。 ・ Tech Against Terrorism のメンバーとして、暴力的過激主義者がプラットフォームを利用して害を及ぼすことを防ぐ取り組みを実施。 ・ マイクロソフトはさまざまな省庁・団体と協力し、オンラインのセキュリティと安全性の確保に努力。日本サイバー犯罪対策センター (JC3)、JPCERT、ICT-ISAC、FISCACAS、Software ISAC、フィッシング対策協議会、日本サイバーセキュリティ・イノベーション委員会 (JCIC) 等。 ・ 専用フォームを通じて報告が可能な日本政府や法執行機関と協力しており、これには日本の警察も含む。違法・有害情報相談センターは X に報告することができ、関連する問題やトピックについて情報提供や議論を行うために連携。
19 行政機関や地方公共団体等との連携状況	**○ 半数の国内事業者及び全ての国外事業者は、偽・誤情報の流通・拡散への対応という観点にかぎらないものも含め、偽情報等の流通状況や拡散の仕組みに係る分析、事業者に求められる偽情報等への対応に係る検討、事業者の事例の発表・共有、インターネット悪用事例に関する情報交換、違法・有害情報相談センターに基づく対応、若年層向けの啓発動画キャンペーン、ICT 利活用のためのリテラシー向上プロジェクト、総**

	務省のデジタルリテラシーキャンペーン、**観光・広報・文化・伝統芸能・中小事業者支援・防災啓発・平和教育・気候変動啓発など多岐にわたる政策テーマにおいて広報啓発・プロモーション**、**オンラインにおけるセキュリティと安全性の確保**、警**察庁や各都道府県警等との情報共有やサービス・製品の説明**、**防災目的のために公共機関のアカウントに無料の公共サービス API を提供**等、日本の行政機関や地方公共団体等と連携・協力した取組を実施と回答。 ● 半数の国内事業者は、**偽・誤情報の流通・拡散への対応という観点において、日本の行政機関や地方公共団体等と連携・協力した取組に関する明確な回答なし**。なお、支障となっている点として、**連携に関する具体的な情報不足**、**担当者の確保が難しい**ことを回答。 【行政機関や地方公共団体等との連携状況】(参考資料22-1-1 及び22-1-2 Q16－1～16－2) **【取組［例］】** ・ 研究者・官公庁・業界団体等とともに、偽情報等の流通状況や拡散の仕組みに係る分析や、プラットフォーム事業者に求められる偽情報等への対応に係る検討を実施。 ・ 総務省「ICT 活用のためのリテラシー向上に関する検討会」にオブザーバー参加し、事業者の事例を発表・共有。 ・ テロ等に関連するインターネット悪用事例に関する情報交換(GIFCT ワークショップ参加)。 ・ 違法・有害情報相談センターの「青少年案件情報提供スキーム」に基づく対応。 ・ 常日頃から総務省を含む様々な省庁の関係者と対話をし、また数多くの研究会にも委員ないしはオブザーバーとして参画。今回のようにヒアリングの要請にも対応。 ・ 総務省「ICT 活用のためのリテラシー向上に関する検討会」(座長:山本龍彦 慶應義塾大学大学院法務研究科教授)の取り組みの一環として、セーファーインターネットデー(2月6日)に合わせて「ネット&SNS よりよくつかって 未来をつくろう～ICT 活用リテラシー向上プロジェクト～」が公開された際、コンテンツの提供者として協力。 ・ 昨年4月には、GLOCOM 国際大学が主催しましたイベント「フェイクニュースと日本―私たちにできること・社会としてできること―(G7 デジタル・技術大臣会合関連イベント)」を支援。そのイベントに合わせて、YouTube では、総務省並びに国際大学 GLOCOM ご協力の元、フェイクニュースに惑わされないための若年層向けの啓発動画キャンペーンをローンチ。本キャンペーンは、若者に人気のある 9 名の YouTube クリエイターの協力を得て作成された、「フェイクニュースは身近に存在すること」「ファクトチェックが重要であること」そして、「安易な拡散が人に

	迷惑をかけてしまうリスクに繋がりかねないこと」この三つのメッセージを伝えるショート動画。YouTube で配信された各クリエイターのこれらの動画は、合計で 1500 万回以上再生（2024 年 3 月現在）。YouTube は、「情報に対するリテラシーを高める」ことも大切と考えており、今後も真摯に取り組む。 ・ 2020 年より、NPO 法人企業と教育協会（ACE）と共同でデジタルリテラシー教育プログラム「みんなのデジタル教室#wethinkdigital」を実施。「フェイクニュースの見抜き方」「デジタルアイデンティティを考える」の 2 つのモジュールで、これまでに全国 24000 人以上の中高生を対象に授業を実施。2023 年 7 月、弊社は既存の 2 モジュールのリニューアルを発表。誤情報に関する総務省のリテラシー資産を統合し、新モジュール「デジタルシチズンシップと情報発信」を導入。このプログラムは総務省情報通信局の支援。 ・ 総務省と共同で政策討議： 2022 年以降の ICT 利活用のためのリテラシー向上に関する研究会。ICT 利活用のためのリテラシー向上プロジェクトに貢献。 ・ 2024 年 2 月、「セーファーインターネットデー」に向けて、透明性ツールや、広告を含むパーソナライゼーションがプラットフォーム上でどのように機能するかについての利用者教育キャンペーンを開始。このキャンペーンを IG のクリエイターと共に作成し、また同日に開始された総務省のデジタルリテラシーキャンペーンとも連携。キャンペーン・ランディングページは総務省のウェブサイトでも紹介。 ・ 多様化する行政課題の解決を支援するため、地方公共団体や行政機関と連携のもと、観光、広報、文化・伝統芸能、中小事業者支援、防災啓発、平和教育、気候変動啓発など多岐にわたる政策テーマにおいて広報啓発・プロモーション等の取り組みを推進。 ・ 令和 5 年 7 月九州北部豪雨災害や能登半島地震を支援するチャリティーLIVE 開催および寄付贈呈等の復興支援プロジェクト。 ・ 広島県との「県政コミュニケーションに係る連携及び協力に関する連携協定」の締結および毎年 8 月 6 日に開催する「平和記念式典」の LIVE 配信。 ・ 札幌市と「魅力発信の取り組みに関する連携協定」を締結。 ・ 選挙ドットコムと連携した統一地方選挙に合わせた選挙教育プロジェクト。 ・ 各省庁とも連携したプロジェクトも多く実施。 ・ マイクロソフトは各省庁と協力し、オンラインにおけるセキュリティと安全性の確保に努力。例えば、･総務省リードによるリテラシーキャンペーン、テロ等に関連するインターネット サービスの悪用に関する勉強会、マイクロソフト

	が有する脅威情報を各省庁に提供（GSP: 政府セキュリティ プログラム）、各省庁との月 1 回のセキュリティ勉強会、専門知識の提供や各種セキュリティ ガイドラインの策定支援（内閣サイバーセキュリティセンター（NISC）クラウド設定ガイドライン、経済産業省クラウド セキュリティ ガイドライン、金融情報システムセンター（FISC）金融システム監査基準など）、地方自治体と提携し、工業高校でのリテラシー習得支援事業を実施。 ・ 警察庁や各都道府県警、またその他省庁と随時連携。特定の事項に関する情報共有や、私たちのポリシー、製品（コミュニティノートを含む）についての説明など。 ・ X は防災目的のために公共機関のアカウントに無料の公共サービス API を提供。
20 国際機関等との連携状況	○ 一部の国内事業者及び全ての国外事業者は、偽・誤情報の流通・拡散への対応という観点にかぎらないものも含め、**G7 デジタル・技術大臣会合関連イベント**、**G7 内務・安全担当大臣会合への対面での参加やプレゼンテーション**、**International Fact-Checking Network（IFCN）の支援**、**IFCN 主催の世界ファクトチェック会議 Global Fact 10 のスポンサー及びセッション開催**、**Global Network Initiative（GNI）・Global Internet Forum to Counter Terrorism（GIFCT）・Digital Trust and Safety Partnership（DTSP）・Tech Coalition 等の創設**、**「2024 年選挙における AI の欺瞞的使用に対抗するための技術協定」を 2 月 16 日に業界各社と締結**、**テクノロジー企業のための自主的な Election Integrity Guidelines に署名**、**コンテンツが AI を使用して作成されたことを示す共通の技術標準への協力**、**UNDP と連携したハラスメント及びヘイトスピーチ防止キャンペーン**、**ダボス会議等におけるプラットフォームの安全性について各国の政府関係者と意見交換**、**GIFCT・Rome Call for AI Ethics（人工知能の倫理的ガイドライン）・Paris Call for Trust and Security in Cyberspace（サイバー空間の信頼性と安全性のためのパリ・コール）・2024 年選挙における人工知能の欺まん的使用に対抗するための技術協定等マルチステークホルダーアプローチを主導**等、国際機関等と連携・協力した取組を実施と回答。 ● ほぼ全ての国内事業者は、**偽・誤情報の流通・拡散への対応という観点において、国際機関等と連携・協力した取組に関する明確な回答なし**。なお、支障となっている点として、**連携に関する具体的な情報不足**、**担当者の確保が難しい**こと、また、一部の国外事業者は、**関連する技術の堅牢性は業界全体の課題で、可能な限り取り組む必要があるが、何がどこまで可能なのかを見極める必要もあること**を回答。 【国際機関等との連携状況】（参考資料22-1-1 及び22-1-2 Q17－1～17－3）

	【取組［例］】 ▪ 旧 LINE および旧ヤフーは、G7 デジタル・技術大臣会合関連イベント「フェイクニュースと日本―私たちにできること・社会としてできること―」（国際大学 GLOCOM 主催、総務省およびグーグル合同会社後援、2023 年 4 月開催）に協力参加。 ▪ デジタルやメディア・リテラシープログラムやユーザーだけでなく、ジャーナリスト及び研究者向けトレーニング等の活動を通して、Google はプロダクトを超えて健全なジャーナリズムのエコシステムをサポートし、市民団体や研究者とパートナーシップを組み、将来リスクの一歩先を行き、このエコシステムの一部として、世界中の事実確認の専門家やネットワークとの協力し、各国のファクトチェック団体の連合組織である International Fact-Checking Network （IFCN）のような組織を支援。 ▪ Global Network Initiative (GNI)、Global Internet Forum to Counter Terrorism (GIFCT)、Digital Trust and Safety Partnership (DTSP)、Tech Coalition など、多くの国際組織やイニシアティブの創設メンバーまたは参加メンバー。 ▪ ミュンヘン・セキュリティ・カンファレンスにおいて、「2024 年選挙における AI の欺瞞的使用に対抗するための技術協定」を 2 月 16 日に業界各社と締結し、欺瞞的 AI 選挙コンテンツに関連するリスクを軽減するための 8 つの具体的な約束。Adobe、Amazon、Google、IBM、Meta、Microsoft、OpenAI、TikTok、X を含む 20 のテクノロジー企業がこの協定に署名。 ▪ Google、Microsoft、The Rockefeller Foundation、Snap、TikTok とともに、International Foundation for Electoral Systems (IFES)が運営するテクノロジー企業のための自主的な Election Integrity Guidelines にも署名。このガイドラインは、企業と選挙管理当局が選挙の完全性を推進し、情報エコシステムに対する信頼を高めるために期待されること、および実践すべきことを共有。ガイドラインは、韓国ソウルで開催された民主主義サミットで発表。 ▪ コンテンツが AI を使用し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。これらのシグナルを検出できるようになれば、利用者が Facebook や Instagram に投稿する AI 生成画像にラベルを付けることが可能。現在この機能を構築中で、今後数ヶ月のうちに、各アプリがサポートするすべての言語でラベルの適用を開始する予定。弊社は、利用者がどのように AI コンテンツを作成し共有しているのか、利用者がどのような透明性に最も価値を見出すのか、そしてこれらの技術がどのように進化していくのかについて、さらに多くのことを学べることを期待。業界のベストプラクティス、そして弊社自身の今後のアプローチに反映。s ▪ 「みんなのデジタル教室」で一般の人々に対しても認識向上の活動。

	▪ 国際会議のスポンサーとして参加し、国際ファクトチェックネットワーク（IFCN）主催の世界ファクトチェック会議 Global Fact 10 のスポンサー及びセッション開催。 ▪ 国際機関とのメディアリテラシーキャンペーンのコラボとして、UNDP クリエイター・フォワード：UNDP と連携したハラスメント及びヘイトスピーチ防止キャンペーン。 ▪ 米国国立児童行方不明センター（NCMEC）との連携として、児童の性被害防止のために、24 時間体制で、疑わしい児童搾取コンテンツを積極的に米国国立児童行方不明センター（NCMEC）に報告。その後、NCMEC はこれらの情報を警察庁と共有。また、コンテンツに緊急性があるか、または深刻な場合は、該当チームが迅速に情報を提供するため警察庁に、サイバーチップを共有。 ▪ 世界経済フォーラム年次総会（ダボス会議）等に出席し、プラットフォームの安全性について各国の閣僚はじめ政府関係者とも意見交換。 ▪ デジタル関連の問題に対処する際、マイクロソフトはマルチステークホルダーアプローチを主導して奨励するよう努力。インターネット上の問題に効果的に対処するには、その性質上、世界規模の対策が必要になることが多い。例えば、マイクロソフトは GIFCT（テロ対策に関するグローバル インターネット フォーラム）、Rome Call for AI Ethics（人工知能の倫理的ガイドライン）、Paris Call for Trust and Security in Cyberspace（サイバー空間の信頼性と安全性のためのパリ・コール）、Tech Accord to Combat Deceptive Use of AI in 2024 Elections（2024 年選挙における人工知能の欺まん的使用に対抗するための技術協定）など、複数の利害関係者による国際的な枠組みを主導。マイクロソフトは、日本をはじめとする世界中の関連する政府、企業、学術機関、非営利組織（NPO）を支持し、支援。 ▪ 2023 年 12 月の G7 内務・安全担当大臣会合への対面での参加やプレゼンテーション、また GIFCT のメンバーとしての活動など。 ▪ テロ関連の違反コンテンツを特定するため、業界が共有するハッシュデータベース（GIFCT が支援）を使用したり、報告がある前に PhotoDNA のような業界共有ハッシュを利用するなど、さまざまな方法を実施。オンラインでのテロリストや暴力的過激派のコンテンツをより効果的に検出し削除するための技術投資を続けることを約束。これには、デジタルフィンガープリント技術や AI 技術の発展や導入。クライストチャーチ・コール、GIFCT、EU インターネットフォーラム（EUIF）など、様々な団体との協力を通じて、テロリストや過激派がインターネットをどのように利用しているかの新たな動向を把握。

	▪　CSE（子供への性的搾取）やテロリストコンテンツを撲滅するための取り組みを実施。
21　その他（上記以外のステークホルダーとの連携状況、児童その他脆弱な主体の保護）	○　ほぼ全ての国内事業者及び全ての国外事業者は、偽・誤情報の流通・拡散への対応という観点にかぎらないものも含め、**様々な分野の専門家・ジャーナリスト・クリエイター等と連携した情報発信**、**クリエイターへの情報・ノウハウの共有・エデュケーションプログラム・人材交流等支援**、**（一社）VRM コンソーシアム等への参加による他のプラットフォーム事業者やメタバース関連事業者との連携・協力**、**（一社）コンピュータエンターテインメント協会に参加**、**（一社）セーファーインターネット協会・（一社）安心協・モバイルコンテンツフォーラム等業界団体を通じての取り組み**、**アジアインターネット日本連盟（AICJ）や一般社団法人ソーシャルメディア利用環境整備機構（SMAJ）等業界団体における取り組み**、**ゼロレーティングの享受を通じた電気通信事業者との連携**、**10 代や保護者向けの啓発活動**、**新型コロナ感染症ワクチンに係る誤情報に関する啓発活動への協力**、**不登校生動画選手権の共催**、**プラットフォームにおける性被害の防止のための啓発**等、その他のステークホルダーと連携・協力した取組を実施と回答。 ●　一部の国内事業者及び国外事業者は、その他のステークホルダーとの連携において支障となっている点として、**行われているイベントや勉強会に関する情報共有の仕組みがないこと**や、**誤情報や偽情報のような問題には業界を超えた協力と社会全体からのアプローチが必要であること**を回答。 【その他のステークホルダーとの連携状況】（参考資料22-1-1 及び22-1-2 Q18－1～18－4） **【取組［例］】** ▪　2023 年8月、さまざまな分野の専門家やジャーナリスト、クリエイターが自らの知見をもとにユーザーの新しい気づきや考えるヒント、行動につながる情報を発信するプラットフォーム（Yahoo!ニュース エキスパート）を新たに開始。記事や、コメントで各分野のエキスパートからの信頼性の高い情報を提供。クリエイター」のほか、Yahoo!ニュース個人の書き手である「オーサー」「コメンテーター」総勢約 2600 名が「Yahoo!ニュース エキスパート」に参加。 ▪　多くのコンテンツ制作主体者に協賛・ご協力いただき、多種多様なイベントを開催。日本ネットクリエイター協会等、クリエイターの方々の支援に繋がる組織への参加。 ▪　一般社団法人 VRM コンソーシアム等への参加により、他のプラットフォーム事業者様やメタバース関連事業者との連携・協力。 ▪　一般社団法人コンピュータエンターテインメント協会に参加し、「CS品質向上委員会」に所属。各メーカーのカスタマーサポートに寄せられた課題・取組事例等の共有、消費者トラブルを未然に防ぐ取組。

	・　セーファーインターネット協会、安心協、モバイルコンテンツフォーラムなど、業界団体を通じての取り組みや、有志企業担当者との勉強会は随時開催。 ・　アジアインターネット日本連盟（AICJ）や一般社団法人ソーシャルメディア利用環境整備機構（SMAJ)などの業界団体における取り組みを通じ、日本における革新的なビジネス及びインターネット産業の健全な成長向け連携・協力体制を構築。 ・　Instagram は 2020 年より、クリエイターと連携し、若年層ユーザーと一緒にインスタグラムの安全な使い方を考えるプロジェクト「#インスタアンゼン会議」を立ち上げ、10 代や保護者向けの啓発活動を実施。 ・　他の事業者との業界連携に積極的に貢献。SIA が主催した 2021 年の新型コロナ感染症ワクチンに係る誤情報に関する啓発活動に積極的に協力。 ・「不登校新聞」と連携し、『不登校生動画選手権』を 2023 年 5 月に共催。今年もこの取り組みは継続予定。 ・　日本におけるゼロレーティングの享受を通じた連携として、Softbank: ウルトラギガモンスタープラスプラン（2020 年3月 11 日をもち新規受付終了）、NuroMobile：Neo プラン、Povo2.0 PrePaid プラン：期間限定のゼロレーティングを含むオプション。SNS データ使い放題(3/20-4/20, 2023)、SNS＋動画データ使い放題(4/28-5.28, 2023)。 ・　SNS等のプラットフォームを運営する事業者等から構成される一般社団法人ソーシャルメディア利用環境整備機構（SMAJ）に理事会社として参加し、会員の事業者と、偽・誤情報の課題検討等、各種委員会・ワーキンググループにおいて意見交換やベストプラクティスの共有。 ・　他のプラットフォームと連携した、安全のための啓発活動。2021 年にはプラットフォームにおける性被害の防止を目的として、専門家や NPO に加えて旧 Twitter Japan 株式会社、旧 LINE 株式会社と連携したオンラインセミナーを開催し、LIVE で配信することで啓発メッセージを発信。 **○　ほぼ全ての国内事業者及び全ての国外事業者は、児童その他脆弱な主体の保護について、一定年齢以上のみの利用、年齢や健康状態等によるターゲティング広告の禁止、色彩多様性への配慮、摂食障害や自殺を助長または美化するようなコンテンツ等の禁止や基本的な考え方の制定等を回答。** 【児童その他脆弱な主体の保護状況】(参考資料22-2-1 及び22-2-1Q16、同22-5-1Q6 及び22-5-2Q9) **【取組［例］】** ・　Yahoo!広告において、一定の年齢以上の方のみの利用を推奨したり、酒類など一定の商品に関する広告に関し

	て未成年への配慮を求めるなど、サービス等の性質に応じ多岐にわたる配慮を実施。プロファイリングに基づく広告規制に関しては、Yahoo!広告において、13 歳未満のユーザーデータの利用及び 18 歳未満のユーザーに対する年齢ターゲティング及び興味関心に基づいたターゲティングを禁止。また、人格形成の未熟な子どもに対して悪影響を及ぼさないようにするため、以下のような表現を禁止。 (1)下劣、卑わい、暴力的、また危険性を伴う表現で幼児や児童がまねしやすい表現 (2)いじめを助長する表現 (3)懸賞などによって必要以上に児童の射幸心をあおるような表現 (4)幼児や児童がその商品を持っていなければ仲間はずれにされるような表現 ・ Yahoo!広告において、健康状態に関連するデータ(病気や障がいなど)を利用したターゲティングを禁止 ・ LINE 広告において、広告審査ガイドラインにて、「子どもに対する配慮」として、以下のような表現を禁止。 ・暴力的、性的、不快感のある下品な表現 ・危険性を伴う表現で子どもが模倣するおそれのある表現 ・いじめを助長したり、暴力を肯定したりする表現 ・子どもの劣等感や優越感を利用する表現 ・過度に子どもの購買欲を煽るような表現 ・商品を買ってもらえるよう大人に働きかけることを促す表現 ・ 保護者およびお子さんに向けて、子どもたちのプライバシー保護について説明するページを公開し、子ども向けポータルサイト「Yahoo!きっず」の利用や、弊社サービスにおけるお子さんのプライバシーの保護施策を紹介。 ・ さまざまなサービスにおいて、色覚多様性の方にも情報をわかりやすくお届けできるよう工夫。特に Yahoo!天気・災害では、毎日の生活に関わるからこそ、あらゆる人に使いやすいサービスが提供できるように日々デザインの改善を実施。Yahoo!天気・災害で使っている天気図と地震画像は、色覚多様性の人にとってもわかりやすいデザインになるように検証を行ったほか、天気図などのデータ提供元の意見も踏まえ、2018 年に改善。 ・ 視聴者としての児童への配慮として、アカウントに登録された生年月日から 18 歳未満のユーザーと判断した場合、投稿ユーザー自身が「R-18」ジャンルを設定しているコンテンツについて、トップページからの導線を表示しなくするなどの制限を実施。投稿ユーザー自身が設定したジャンルに関わらず、性的な表現、暴力的な表現を含むなど、未成年が視聴するのは不適切と判断したコンテンツについては、ランキング・検索等への表示を制限。

	・ 被写体としての児童への配慮として、児童自身が公開するコンテンツであるか否かに関わらず、児童ポルノ・児童虐待にあたる可能性があると判断したコンテンツについては、削除対応を実施。映像内容から児童虐待を受けている可能性があり、アカウント情報等からおよその地域の判断がつくようなケースでは、児童相談所への通報を実施。 ・ 広告表示に関する配慮として、未成年に対する配信が好ましくないアルコール等の広告に関しては 20 歳以上の年齢ターゲティングを必須として販売。 ・ 人権保護については充分に留意して対応。当事者や専門機関（本人・代理人弁護士・自治体・法務局等）から、人権を侵害する行為があったと通報があった場合は、本人確認の上、慎重に対応を検討。 ・ 過去、児童の利用が多いサービスを運営していた際には、登録時に国籍を確認し、その国に合わせた児童保護ポリシーに基づく対応（保護者確認など）、投稿情報のうち特に実写画像を含むものは即時公開を行わず、チェックをしてから公開するなど、通常サービスより厳しい監視。児童に対するプライバシー侵害、性被害等を防ぐため、対象が児童であると見られる場合には、特に厳しい基準での利用停止、公開停止措置を実施。具体的には、対象が成人である場合には意見聴取や注意勧告を経て自主的な改善を促すところ、対象が児童であれば即時に利用停止、公開停止等。これらのポリシーについては、利用規約にて青少年保護規定を設けているほか、はてな情報削除ガイドラインに具体的な基準を記載。児童に限らず、保護を要する脆弱な主体のサービス利用が想定される場合には、あらかじめ約款にて厳しい基準での対応を行う旨を定めておくことが望ましく、特にインターネットサービスに一旦公開された情報の拡散速度は非常に早いため、早期に対応ができる体制を、監視体制なども含めて整備しておく必要。 ・ プライバシーポリシーに記載の通り「15 歳未満の子供から法定代理人（親権者等）の同意なく個人に関する情報をみだりに収集しないよう留意します。」に従ってサービスの運営。広告配信においては、会員情報、またはプロファイリングにより未成年と思われるユーザーに対して煙草等の業種の広告は表示しないように配信制御。一般的にセンシティブなテーマや扇動的なテーマに関する広告は Ameba 上では配信されないようにカテゴリブロックを運用。 ・ 摂食障害や自殺を助長または美化するようなコンテンツや、危険行為に挑戦するコンテンツを禁止。また、YouTube Kids（13 歳以下）や、もう少し年齢の高い子どもの保護者向け管理機能を通じて、年齢に応じたコンテンツライブラリを利用できるようにしているほか、子どもの YouTube 利用時間の管理に役立つ各種機能を提供。

	18 歳未満のユーザー向けに、休憩およびおやすみ時間のリマインダーや、若年層にはふさわしいとは言えないコンテンツに年齢制限を設けるなど、さまざまな安全機能を提供。 ・ YouTube では、個人を対象とした身体的特徴または保護対象グループへの所属（年齢、障がい、民族、性別、性的指向、人種など）に基づく長期に及ぶ侮辱または中傷を含むコンテンツは不許可。その他の有害な行為（脅迫や晒し行為など）も不許可。未成年を対象とするコンテンツに対しては、より厳しく対処。 ・ YouTube が運営上の指針としている 5 つの基本的な考え方を紹介。①子どもと青少年のプライバシー、身体的安全、メンタルヘルス、ウェルビーイングを守るために、オンラインにおいて特別な対策が必要。②特に 12 歳以下の子どもがいる家庭内でのオンライン利用ルールを定めるにあたり、保護者は重要な役割を担う。③すべての子どもは、個人的な興味やニーズに応じて、年齢にふさわしい質の高いコンテンツを無料で視聴する機会を与えられるべきである。④子どもの発達上のニーズは青少年とは大きく異なり、オンライン体験に反映させる必要がある。⑤適切な安全保護対策があれば、革新的なテクノロジーは子どもや青少年に恩恵をもたらすことができる。 ・ 2023 年 10 月に「子どもと青少年をオンライン上で保護するための法的枠組みへの提言」を発表。 ・ プラットフォームを若者にとって安全で年齢に適したものにするために、青年の発達、心理学、および精神衛生の専門家と定期的に相談し、どの種類のコンテンツが若年層の利用者にとって不適切である可能性があるかを理解するために努力。ポリシーに違反するコンテンツを削除することに加えて、若年層の利用者が年齢に不適切なコンテンツを見にくくすることも目指す。年齢に不適切なコンテンツとは、生物学的、社会感情的、認知的な発達の違いから若年層の利用者に特有のリスクをもたらす可能性があるコンテンツ。若年層の利用者にとって、潜在的にセンシティブなコンテンツを見つけにくくしている。本人がフォローしている人が共有した場合でも。 ・ 若年層の利用者に対して年齢に適した保護を提供する一方で、彼らを含む利用者が自分に合った方法でソーシャルメディアの体験を管理し、ウェルビーイングをサポートできるようにすることも重要と認識。これが、利用時間の管理、望まないやり取りの防止、見たいコンテンツやアカウントの制御ができる機能を導入した理由。また、いじめを防ぐためのツールも開発し、利用者がアカウントを管理していじめに遭遇しないようにしている。 ・ 年齢に応じて表示されるコンテンツを制限。例えば成人向けのコンテンツは 13 歳から 17 歳のアカウントを持つユーザーは視聴不可。16 歳未満のユーザーが作成したコンテンツはおすすめフィードの対象外。それ以外にも、ペアレンタル機能や青少年保護のための啓発活動・イベント等、様々な取組。 ・ 児童が有害なコンテンツに触れないように保護していただけるように、家族向けに Bing セーフサーチと

	Microsoft ファミリー セーフティ設定を用意。Bing のセーフサーチ機能は、露骨な画像やテキストを検索結果から除外。セーフサーチは、ほとんどの市場では既定で「標準」に設定されており、成人向けの画像が検索結果から除外。「高レベル」モードでは、露骨な表現を含むテキストと画像の両方が検索結果から除外。ファミリー セーフティ設定も、保護者が Microsoft アカウント、デバイス、またはローカルネットワークで、子どもの検索エクスペリエンスを特定のセーフサーチ設定に設定できるようにするもの。さらに、これらの機能を使用してデバイスまたはネットワーク上で Bing の生成 AI 機能を無効にすることも可能。 ▪ ユーザーは 13 歳以上でなければなりません。13 歳未満の人がアカウントを作成しようとすると、登録不可。未成年者は X のユーザーベースの一部を占めるだけですが、オンラインでより脆弱な対象であるこのグループの保護に全力。私たちは、プラットフォーム上の若年層のユーザーを保護するためにさまざまなツールとポリシーを導入。 ▪ 年齢保証プロセスは、自己申告による年齢確認と追加の技術的措置を組み合わせて、アカウント保持者の年齢が本物であり、子供を保護するための適切なコントロールが設置されていることを確認。13 歳から 17 歳のユーザーは、デフォルトでプライバシー、安全、セキュリティの設定が強化。例えば、彼らはグラフィックコンテンツ、成人の裸体、性的行動を含む敏感なメディアを見ることができず、ダイレクトメッセージが制限され、位置情報もオフ。広告主はこの年齢層をターゲットにすることが禁止。親や他の人からアカウント保持者が 13 歳未満であるとの確認が得られた場合、私たちのポリシーはそのアカウントを永久に停止。さらに、ユーザーが 18 歳未満であることを示す生年月日を入力すると、新しい生年月日を再入力することができなくなる。 ▪ 敏感なメディア設定：未成年者が知っているアカウントやプロフィールに生年月日を含めていないユーザーに対して、成人向けコンテンツなど特定の敏感なメディアの視聴をインタースティシャルを通じて制限。 ▪ コンテンツ通知： 通知やインターステーショナルの背後に敏感なメディアを隠す。これには未成年者が成人向けコンテンツを閲覧するのを制限する年齢制限コンテンツポリシーが含まれる。 ▪ セーフサーチ： 自動的に、未成年者や生年月日が記載されていないアカウントの検索結果から、潜在的に敏感なメディアコンテンツ（ユーザーがミュートまたはブロックしたアカウントも含む）を除外。 ▪ 広告： 未成年者向けに禁止されているコンテンツを含む広告は「ファミリーセーフではない」とタグ付けされ、21 歳以上のユーザーにのみ表示。この基準は現時点で全世界で一貫。 ▪ アカウントのデフォルト保護設定： 未成年者のアカウントはデフォルトで「保護された投稿」に設定されます。これは、新しい人がフォローを希望した際に未成年者がリクエストを受け（承認または拒否が可能）、彼らの投稿がフォ

	ロワーにのみ表示され、彼らとそのフォロワーにのみ検索可能であることを意味（つまり、公開検索には表示されない）。 ・ ダイレクトメッセージのデフォルト保護設定： 未成年者のアカウントはデフォルトでフォローしているアカウントからのダイレクトメッセージのみを受け取るよう制限。 ・ 年齢ロック： 新しいユーザーが 18 歳未満であるとする生年月日を入力した場合、そのアカウントで新たに生年月日を再入力することはできなくなる。 ・ 自殺および自傷行為の援助： ユーザーが自殺や自傷行為を促進または奨励することを禁じるポリシーを開発。自殺または自傷に関連する用語を検索すると、最上位の検索結果は助けを求めることを奨励する通知が表示。 ・ 人々が X 上で虐待を経験すると、自己表現の能力が危険にさらされることがある。研究によると、特定のグループの人々がオンラインで過剰に虐待の対象。複数の代表性の低いグループに属する人々にとって、虐待はより一般的で、性質がより重く、より有害である可能性。特に過去に疎外されてきた人々の声を封じ込めようとする虐待に対抗することに尽力。この理由から、保護されたカテゴリーに属すると見なされる個人またはグループを標的にした虐待行為を禁止。 ・ 自殺と自傷行為に関するポリシー。報告を繊細に扱うために、アプリ内報告は、自傷行為を行う意図を表現している可能性のある人々と、自傷行為や自殺を励ますまたは促進するコンテンツのために、別々のオプションを提供。