

目次

1. ファクトチェック団体や他のメディアとの連携の在り方・課題について	・ ・ ・ P. 1	21. 日本国内で選挙や自然災害時に実施した対応及びその効果	・ ・ ・ P. 17
2. 情報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（件数、具体的な事例など）について	・ ・ ・ P. 1	22. 選挙期間中にコンテンツの削除の判断をするために必要な期間について	・ ・ ・ P. 18
3. 選挙関連ポリシーの日本国内での運用状況直近の国政選挙に際しての認知件数、通報件数、対応件数など）について	・ ・ ・ P. 2	23. 検索サービスでのモデレーションに関する日本の状況について	・ ・ ・ P. 18
4. AI 生成コンテンツへの対策（選挙時の対策、自動ラベル付け、ポリシーによる対応など）及びその運用状況（ポリシー違反の件数など）について	・ ・ ・ P. 3	24. ニュースフィードのパーソナライゼーションについて	・ ・ ・ P. 19
5. ミュンヘン安全保障会議に際しての「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）」締結を受けて日本国内で実施予定の取組について	・ ・ ・ P. 4	25. 検索エンジンにおける、情報の信頼性の確保について	・ ・ ・ P. 20
6. 電子透かしやフィンガープリント技術を、全てのユーザーやデータが識別・認識されているわけではないインターネットの世界にどのように拡張していくのかについて	・ ・ ・ P. 5	26. 外部からの情報提供の重要性、日本における取組への影響について	・ ・ ・ P. 21
7. 日本で偽・誤情報対策に関するガイドラインや行動規範を作る必要性・妥当性について	・ ・ ・ P. 6	27. コンテンツモデレーションの機能の有効性を社内で評価する取組について	・ ・ ・ P. 21
8. なりすまし対策（削除等対応の具体的な条件など）について	・ ・ ・ P. 8	28. コミュニティノートを実際に導入した前後で、偽・誤情報の流通・拡散状況に何らかの影響	・ ・ ・ P. 22
9. 政治広告の制限（制限の内容や具体的な条件など）について	・ ・ ・ P. 9	29. コミュニティノート等、ボランティアベースの取組の限界について	・ ・ ・ P. 23
10. 利用者からの通報受付体制について	・ ・ ・ P. 10	30. スпамと分類された投稿の中に、誤情報あるいは虚偽情報に該当する件数について	・ ・ ・ P. 24
11. IT 業界における共同の取組を成功させ、継続させていくために重要なことは何かについて	・ ・ ・ P. 10	31. コミュニティノートの運用の透明性について	・ ・ ・ P. 24
12. 日本語のコンテンツへの対応状況の集計・開示の可否について	・ ・ ・ P. 11	32. コードを公開することによる一般利用者へのリーチについて	・ ・ ・ P. 25
13. 日本語を母語とする／専門とするコンテンツレビュワーの人数について	・ ・ ・ P. 11		
14. 監督委員会への評価について	・ ・ ・ P. 12		
15. 自動モデレーション技術をすり抜けてコンテンツが公開され、問題となった事例について	・ ・ ・ P. 13		
16. 児童の保護に関する特別な配慮（プロファイリング規制等）の有無について	・ ・ ・ P. 14		
17. コンテンツモデレーションの段階ごとの線引きについて	・ ・ ・ P. 14		
18. 警告ラベルの種類について	・ ・ ・ P. 15		
19. 信頼できる情報を目立たせる（いわゆるプロミネンス）取組について	・ ・ ・ P. 15		
20. 国内の研究者への API 等を用いたデータ開示に係る課題・悩み・連携可能性等について	・ ・ ・ P. 15		

(設問)	Google	Meta	TikTok Japan	Microsoft	X
ヒアリング当日の質疑応答より					
1. ファクトチェック団体や他のメディアとの連携の在り方・課題について	・メディアリテラシー等の重要性をグローバルで認識。 ・ファクトチェックや Google News Initiative で人員削減があっても、外部の機関やファクトチェッカーなどと連携して取組を実施。 ・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。 ・ファクトチェック組織に様々なツールを提供。 ・2022 年後半、ポインターインスティテュートに 1300 万米ドルの助成金支出。 ・150 万米ドルを拠出して日本ファクトチェックセンター (JFC) の立上げ支援。2023 年には 140 万米ドルの投資。 ・特に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。 ・4 月 2 日の国際ファクトチェックデーでもたくさんの取組を予定。		・日本国内のファクトチェック機関と連携できる可能性は常に継続的に検討している。ぜひ引き続きいろいろな機関と意見交換させていただき、どのような可能性があるのかという検討の機会をいただけたら。 ・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モデレーターから Lead Stories に確認するなどしている。 ・モデレーターのトレーニングに力を入れている。どんどん偽・誤情報として新しいものが出てくる中で、いかにそのキャッチアップをしていくのか、そしてそれをいかにモデレーションに反映していくのかについては努力が必要。 ・メディアとの連携という形は取っていない。報道機関が自らアカウントを開設し、報道の自由・表現の自由の関係で TikTok Japan は意図的にタッチしないという整理。 ・ファクトチェック結果がなかなか広まらないという問題について支援できることはないかということを考えており、例えば、日本ファクトチェックセンターに TikTok アカウントを作っていただいて、より確かな信頼できる情報の発信者として、TikTok の中で情報を発信していただくというのができたらよいのではないかと考えている。	・Tech Accord の中でもコミットしているとおりに、メディア等々との連携は非常に重要で、必須。アメリカ、EU だけでなく、世界各国において、そのような連携が今後必要。 ・我々の方からもメディアにしっかりと連絡をし、アウトリーチをしていこうという計画で、まだ具体的に日本のどこということとは把握はしていないが、広範にやっていく。 ・同時に、選挙や災害時に何かあった際に、テック企業とコミュニティがしっかりと連携することによって、どの画像もしくは動画が本物なのか、どれが AI が生成したものなのかというのを分かるように、それを判断するための情報などもしっかりと啓蒙していきたい。 ・こういった情報が誤解を招くのか、こういった指標があって、それをもって判断できるのかといった検索に必要な知識などもしっかりと啓蒙して、ほかの SNS、OpenAI などとも連携しながらやっていきたい。	
2. 情報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	・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		・Creator Rewards Program について、コミュニティガイドライン違反		・具体的な、例えば、アカウントを停止したといった具体的な定量的な

（設問）	Google	Meta	TikTok Japan	Microsoft	X
運用状況（件数、具体的な事例など）について	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つた場合、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）。 ・例外的な場合だが、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。		のコンテンツはそもそも対象外。コミュニティガイドライン違反した履歴があるとかそういうことも判断されている。したがって偽・誤情報に関するコンテンツには全く収益が還元されていくことがない形で運営。		統計については、ポリシー自体が具体的に導入してからかなり日が浅い、新しいポリシーということもあり、現状で皆様方に本日共有させていただける具体的な定量的な統計はない。 ・ただし、やはり問題のあるこういった投稿によって、これを金銭的に利用していくことはあつてはならないことで、このポリシーを厳格に運用することによって、より高いクオリティのコンテンツが掲載されるということの大きなインセンティブになるというように期待。
3. 選挙関連ポリシーの日本国内での運用状況（直近の国政選挙に際しての認知件数、通報件数、対応件数など）について	・重要なのは信頼できる質の高い情報を上に上げることと、コンテンツモデレーションを行って信頼できない情報をダウングレードしていくこと。 ・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、業界内他社や法執行機関と共有。 ・YouTube にもこうしたリスクを判定するためのチームがあり、ユーザーエクスペリエンスやコミュニティにとって有害なものを分析・研究。最近では生成 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	・ほとんどのデジタルコミュニケーションはインターネット上で行われるので、特定の国のものであるといった属性情報が必ずしもなく、したがってある国に限定した情報を見ることはできない。 ・ある特定の国に対する評価尺度は、その国に対して正確な情報が取れるものにしていきたい。しかし、各国政府から要求されるメトリックスは非常に特定のなものであり、どのように対応していくかに関しては現在議論中。		・日本に特化した情報というのは今手元にはないが、幾つかマイクロソフトとしてコミットメントしておける。 ・虚偽の選挙関連の情報に関して、こういったもの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ですとかアクティビティというのは、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つて、こちらで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しっかりと報告するとともに、LinkedIn では、既に Transparency Report を開示。	・様々な施策を含めて、非常に重要であり、日本でも当然、これが具体的にはプラットフォームのユーザーの皆様、関連の皆様の安全性というところの担保にもつながってまいりため、留まることなく、鋭意強化を進めてまいりたい。

(設問)	Google	Meta	TikTok Japan	Microsoft	X
	にも投資。 ・総務省を含む政府、検討会構成員、市民社会と連携を取っていきたい。			・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・偽情報があつた場合の内容なども、こちらのレポートの中で報告。	
4. AI 生成コンテンツへの対策（選挙時の対策、自動ラベル付け、ポリシーによる対応など）及びその運用状況（ポリシー違反の件数など）について	・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ている。 ・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していく。 ・これらのペナルティはまだ一切料していない。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行う。 ・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、テクノロジーにも力を入れていく。 ・日本では、2023 年 Q3 において、		・悪意ある投稿者が AI 生成ラベルをつけてくれないのではないかという点に関し、自動でラベルをつけられるような機能を開発中。 ・TikTok では 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されるようになっている。 ・AI 生成ラベルをつけないなどのコミュニティガイドライン違反に対しては必ず削除又はおすすめフィードの対象外という形で対処。また、別のガイドラインに違反することもあるので、生成 AI を用いているかどうかだけではなく、他の観点でも審査。	・A I が生成した、もしくは A I 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、A I 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、A I で生成されたものなのか、A I 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。 ・Tech Accord のマイクロソフトのコミットメントとして、LinkedIn も含めて、A I で生成されたものなのかどうかというインディケーターをユーザーに対して提供するという取組も実施。 ・コンテンツの来歴を示すラベルというのが、とても重要な要素。 ・A I に基づくファクトチェックについて、LinkedIn と B i n g に関しては、クレームレビューというものがあつた、このファクトチェックは A I ベースで自動で実施しているわけではない。 ・Start と LinkedIn に関しては、一部自動でのコンテンツのモデレーションなども行っているが、もし自動で何か検出されて、適切でないという判断がなされた場合にも、ユーザーがそれに対して申立てをできるような機会を提供。表現の自由の一環として言っている場合もあつて、そ	・生成 A I については、私ども非常に高い関心を持って状況を今見守っており、いろんな取決めがされているということも認知。 ・いわゆるマニピュレーテッドメディアと呼んでいる形のメディアの問題については非常に重要な問題だと認識しているが、現在、私ども考えておりますのは、業界。 ・いわゆる S N S であるとかという限定的な業界ではなくて、具体的な生成 A I に関係する、A I 自体をモデリングを行っている企業等も含めた大きな枠組みでの各業界がしっかりと連携をしていくということが、いわゆるウォーターマーキングと言われていたような対策についても、いかに広範に連携していけるかというところが非常に重要。 ・その一つの基本的に基盤となつてまいりますのが、ミュンヘンでのセキュリティカンファレンス。安全保障会議、こちらで具体的に、こういう選挙なども含めた部分でのコミットメントが合意されており、こういったものを軸として、いかに広範に業界全体、各業界横断的に対応していくかという視点が非常に重要。

(設問)	Google	Meta	TikTok Japan	Microsoft	X
	コミュニティガイドラインに違反した動画 10 万本を削除。そのうち 97% は機械学習によって自動的にフラギング。また、削除された動画のうち 93% は視聴回数 100 回以下。ある程度日本のユーザーを安全に守っており、悪いコンテンツも速やかに除去できていると考えている。			ういった場合に、何か正当な理由があれば申立てが可能。 ・弊社のオフィス・オブ・A I のほうから、数か月以内にまた別の Transparency Report、透明性報告書というものを発表する予定で、その中でも我々のポリシーなどが新たに発表される予定。	
5. ミュンヘン安全保障会議に際しての「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）」締結を受けて日本国内で実施予定の取組について	・選挙の公正さに関し、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、という 3 つの柱で取組を実施。③に関し、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。 ・選挙という観点から、Gemini に関して非常に慎重なアプローチ（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。メインの取組として、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）。 ・ミュンヘンアコードだけでなく、その他の取組も含め、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。	・この協定によって一夜にして何かが変わるか、全ての選挙に関する課題が解決されるかという点、そうではない。 ・この協定が結ばれた背景・土台として、有害な AI コンテンツからオンラインコミュニティを守るために重要な一歩であるということが謳われると同時に、各企業の継続的な取組が基盤になっているという合意が存在。 ・Meta としては、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）といった取組を進める。	・日本で具体的にどのように対応するかはこれから検討。	・Tech Accord の 8 つのコミットメントについて、グローバルで選挙に関して取組を実施。 ・日本だけではなく、世界的にインパクトが考えられる例として、1 つ目が、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、A I が生成した選挙に関わる虚偽の情報があつた際に、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。 ・候補者の方や担当の方が、例えば A I が生成した情報が画像であれ、ビデオであれ、音声であれ、候補者の方に関して生成されたものが何か	

(設問)	Google	Meta	TikTok Japan	Microsoft	X
				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、B i n gサーチまたはB i n g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・担保しており、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A I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、A I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するという取組も実施。	
6. 電子透かしやフィンガープリント技術を、全てのユーザーやデータが識別・認識されているわけではないインターネットの世界にどのように拡張していくのかについて	・C2PA にサインしたのは 2024 年 2 月であり、取組は始まったばかり。 ・非常に複雑な問題で、1 つの万能薬はなく、複数のアプローチやテクノロジーを組み合わせで対策をしていくことが必要。 ・メタデータに関しても剥がすことが可能であるし、改ざんも可能。ウォーターマークやフィンガープリントに関しても並べ替えたり編集をしたりといった形で損なわれるおそれがある。 ・バランスについていろいろ考えている。1 つクリアなのは、生成 AI は決して悪いものではない。YouTube			・信頼性の向上について、第 1 ステップとしてコンテンツの来歴があるが、これは即効薬で全ての有害なコンテンツを防ぐのに万能薬として効くということではない。非常に重要な要素ではあるが、信頼できる情報源、例えば、メディアなどから画像なり、音声なり、動画なりをしっかりと信頼性を担保したところから入手して、ユーザーに対して提供していく。 ・その中で、情報がどこから来ているのかを、ユーザー側でも考えるようになる。例えば、ここにはコンテンツ来歴というものがないとして	

(設問)	Google	Meta	TikTok Japan	Microsoft	X
	のクリエイターはクリエイティブなプロセスで生成 AI を利用。例えばスクリプトの作成、アイデア出しなど、生産性を上げるために生成 AI を使っているだけであればラベル表示は不要。また、合成コンテンツがどう見ても現実的でない、実物的でないのであればラベル表示は不要。			も、ユーザー側でもリテラシーとして、どこから来ているのか、そして、ほかにどういったところで本物の情報であるということを確認すればいいのかといった、そういったユーザー側でのリテラシーというところもあり、そのような形で、この足りない部分を補完。 ・ 2 点目のポイントが、ハーモナイゼーション、調和という点について、テック企業、Google とか、M e t a といったところが、アドバイザリーボードということで、C 2 P A を立ち上げており、この C 2 P A の標準をもって一貫した形でガバナンスをやっていこうという動きが高まっている。 ・ これらともしっかりと調和をしながらやっていき、C 2 P A は、透明性のある形で、そして、オープンな形で運用されているため、そういったところから出てくる標準などを基に、メディア側、そしてレポーター側、またユーザー側とも調和を図っていくというところが重要。 ・ Tech Accord、C 2 P A 等の様々な取り組み・ツールなども出てきており、特に今後、選挙ですとか、また、A I がさらに進化していく時代においては、こういったものもしっかりと取り入れて、足並みをそろえて調和をさせていくことが非常に重要。	
7. 日本で偽・誤情報対策に関するガイドラインや行動規範を作る必要性・妥当性について	・ EU における行動規範やリスク軽減のための取組といったアプローチが全ての規制当局に対して適切なものに関し、言及する立場にない。 ・ 世界中で様々なコードが存在し、業界主導型のものもあれば、EU のように政府主導型のものもあるとこ			・ 欧州委員会が、災害時や何か事象が起こった際に対応のガイドラインを出したことについて、最善の情報を共有していくというところが非常に重要。ヨーロッパ、そして、それ以外の地域、国がこういった活動を発表しているということは、非常に	

(設問)	Google	Meta	TikTok Japan	Microsoft	X
	ろ、何がその国にとってベストかはその国次第。 ・日本のステークホルダーとの取組として、ここ 2、3 年、意識啓発といったユーザー向けの取組など様々な協働の取組を、例えば SMAJ や SIA とともに重ねてきた。今後、日本でぜひスケールアップしていきたいと考えているのは、例えばミュンヘンアコードのような取組を日本において具体的に展開していくこと。偽・誤情報と戦うためのツール、データセット、情報を、実践的な形で、業界の様々なプレイヤーやファクトチェック組織などと共有していく取組を具体的に進める。			重要。 ・何をレポートするのか、そして、どういった情報を収集していくのかというところに一定の一貫性を持たせるということが非常に重要。それは対応がしやすくなるという観点からではなく、やはり情報の信頼性というところが高まるからということが重要なポイント。 ・複数の国で何度も何度もこういった対応を複数回重ねていくことによって、一貫性を持たせることができる、それによって信頼性を高めることができる。日本とかそれ以外のほかの国も、こういったことを今後やっていく際には、EU やオーストラリアで行われていることを参照しながらやっていくことで、情報とか対応のクオリティも上げていくことにつながる。 ・政府、また様々な組織において、この災害対策のプロトコルというものを持っており、それを基に、テック企業と様々な連携をして、どういった情報が入手可能なのか、どういった情報が有事の際に有益なのかというところを理解し、有事の際に、例えば、そういった情報にユーザーがアクセスしているのかどうか、その情報にアクセスする上で、どういったパスをたどっているのか、例えば、どこのウェブサイトから SNS に行って、またウェブサイトに戻ってきてといった、どのような流れで情報を見ているのかというものをたどれるようにしている。 ・テックインディケーターと呼ばれる情報をたどっていけるような仕組みを、実際有事が発生するずっと前	

(設問)	Google	Meta	TikTok Japan	Microsoft	X
				に立ち上げておく、確立しておくというところがとても重要。 ・組織のインフラに一貫して組み込んで、何かあった際に迅速にレポートをして、危機に対応していく、そういった仕組みをあらかじめ整えておくというところが重要。有事が発生すると、やはり情報も一貫して回らなくなったり、様々な環境への混乱等もあるため、何か起きる前に仕組みづくりをしておくというところが重要。	
8. なりすまし対策（削除等対応の具体的な条件など）について	・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、より強固な対策。すなわち、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。 ・特に公的な有名人のなりすましに関しては長年取り組んでおり、さらにポリシー等を強化していきたい。	・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。そういったコンテンツを見たらぜひ報告をしていただければ幸い。 ・上記ポリシーは、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行っている。 ・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、TikTok 広告ポリシーの禁止コンテンツに該当。したがって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。 ・以上は広告だけでなく、一般コンテンツ・投稿も同じ。	・手元には具体的な数字がないため、LinkedIn に日本での件数がどのぐらいなのかということを確認して、またフォローアップさせていたきたい。 ・スライドの中で、コンテンツが偽情報でないかというモデレーションをやっているという話があったが、2024 年 1 月の状況で、日本ではそういったものはないという形で認識。	・著名人の肖像を使ったような広告に関しては、2 つの取組を御紹介。 ・①弊社では、広告を御利用いただく際に認証プロセスというものを設けており、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というような形。これも完全ではないものの、ペイメント情報が必ずしもつくような形になるので、非常に重要なプロセスの一つ。 ・②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するというようなことになり、幾つかのカテゴリーに関連するところで既に始まっているし、今後とも始まっていく予定。 ・そもそも弊社の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という形になる。 ・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みたい

(設問)	Google	Meta	TikTok Japan	Microsoft	X
					なことがあった場合、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。
9. 政治広告の制限（制限の内容や具体的な条件など）について	・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る。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。 ・広告審査は人と機械、両方でしており、かなり厳密に見ているので、もちろんすり抜けるものもあるかもしれないが、基本的にはリスクは比較的低めなのではないかと考えている。		・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だ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、3つある。 ・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、こういったものが検知された場合については、トロールファームの関連のアカウントについては止めるということを行っており、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また、実際にはなりすましではないが、何らかの存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、いかにもそれの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。 ・ただ、これはいろんな形で迂回するというようなことを先方もやっており、こういったものについては順次対応していくということ、そして、しっかりと我々のプラットフォームを守っていく。 ・③その時々、特に地政学的な、現状で言えば中東情勢であるとか、

(設問)	Google	Meta	TikTok Japan	Microsoft	X
					地政学上のトレンド等に関連したものについても、リアルタイムで対応するように全力を挙げており、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なるが、それ以降、実際のアカウントとして対応したものが80万アカウントある。 ・このような形で複層的に、政治的なメッセージ、あるいはこれに関連するなりすましであったり、偽のID、あるいはtrolleefarmを使った偽のフェイクアカウント、こういったものについては鋭意対応しているというのが現状。 ・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。
10. 利用者からの通報受付体制について	・なりすまし、プライバシー、詐欺・不正などのポリシー違反がある場合、ウェブサイトから簡単に利用者は通報が可能。 ・YouTube に関してコミュニティガイドラインの通報を行うに当たっては、ヘルプセンターも設けている。 ・なりすましや不正、詐欺等に関するポリシーに関して、アプリ上で簡単に通報するフラグを立てることが可能。 ・生成 AI を使った改変や合成に関する通報に関しては、今後、苦情・通報を行うためのウェブフォームをより使いやすいうように変更予定。				
11. IT 業界における共同の取組を成功させ、継続させていくために重要なことは何かについて		・グローバルでボーダーレスなオンラインサービスの一体性が保たれるようにすること（フラグメンテーション）	・現時点で申し上げることはないが、引き続き政府や構成員等と連携させていただければ。	・マイクロソフトとしては、政府の関与の必要性というところに対してオープンにしっかりと受け止めて	

(設問)	Google	Meta	TikTok Japan	Microsoft	X
		ョンの回避)。 ・1つの企業が取り組むのではなく、インダストリーで共通の歩み・協力を進めること。 ・専門的な技術を持った者たちが検討を進めていくこと。 ・テクノロジー企業だけでミスインフォメーションやディスインフォメーションの問題を解決できるわけではないので、社会の各アクター（研究者、市民社会、政府など）が連携して課題に取り組んでいくこと。	・今後、選挙に関する偽・誤情報について、政府とプラットフォームが連携して啓発活動を行うことで、国民全体に効果的に注意喚起することも可能なのではないか。	いきたい。Tech Accord、もしくは、これまでテック業界が構築したものしっかりと連携し協力していきたい。 ・コンテンツの来歴、権威あるコンテンツというところに対して、同じ原則を適用して、信憑性のある情報を提供するというをやっていききたい。	
12. 日本語のコンテンツへの対応状況の集計・開示の可否について		・国の属性に関しては、単に日本語であるかどうかということだけでは特定できない。実際には日本語は世界で使われているので、これを日本だけの情報と特定することはできない。 ・海外でデータを生成して日本を対象とする、あるいは逆に日本でデータを生成して海外だけを対象とする（国内を対象としない）ということもあり得る。 ・日本語のコンテンツへの対応状況に関する数字は現状出せない。データログシステム自体をアップグレード・強化していかないと難しい。			
13. 日本語を母語とする／専門とするコンテンツレビュワーの人数について		・コンテンツレビューの対象は必ずしも自分の理解できる言語だけとは限らない。 ・コンテンツレビュワーのコミュニティは世界 70 言語を対象。その中に日本語も含まれているが、日本語を理解できるわけではないレビュワーが実際には日本語のコンテンツを見ているという場合もあり、日本語を理解できる／日本語でサポートしている人員の数を必ずしも割り出ない。			・まず各国等で具体的にXがどのような形で利用されているのか、そこでの具体的なポスト数であるとか、こういったものを具体的に、あるいは、その利用のされ方を分析した上で、各言語に対応したモデレーターの適正配置を行っているという状況。 ・国ごとといっても、例えば、そこで英語という言語での利用がどのくらいあるのかといったことも含めて、ローカルな言語などの利用状況

(設問)	Google	Meta	TikTok Japan	Microsoft	X
		・人間によるレビューはかなり限定的なリソースしかない。ポリシーのエンフォースメントの 99%以上は機械学習の AI ツールによるもの。当該 AI ツールの中に言語クラシファイアーというものがあり、日本語を含む言語を分類できる。			も含めて配置しており、各国での人数での違いが出てきているというふうに御理解いただきたい。 ・資料のほうには具体的な数値を記載しておらず、一旦セキュリティ及びビジネス上の理由からということで公開していないという形に記載させていただいているが、世界としては2,000名いるという状況。 ・ジャパンスペシフィックなところに関しまして、どの程度御共有させていただくことができるかということに関しては、改めて持ち帰らせていただきたい。
14. 監督委員会への評価について		・監督委員会（オーバーサイトボード）の裁定・勧告について、Meta 側からは全く意見を言うことができない状態にしている。完全に独立した組織。 ・資金提供はしているが、資金の用途に関しては全く口出しできない形になっている。 ・オーバーサイトボードが下す判定・提言、例えばある情報が真実なのかそうでないのかといったことに関する判断は、むしろ民間の企業ができるものではないと考えている。オーバーサイトボードがガバナンスを利かせてそういった判断を下せる独立した機関になっている。 ・またポリシーとして個人の自由な表現、意見の主張の自由も尊重しながら、かつ例えば安全性、個人情報保護といったポリシーもきちんと遵守する形で確立させようとしている。 ・オーバーサイトボードが下した裁定に関し、必ずしも Meta は同じ意見であるとは限らないということにな			

(設問)	Google	Meta	TikTok Japan	Microsoft	X
		るが、オーバーサイトボードは必要な組織であり、特に国民や一般の方々に説明していくという面において非常に重要な役割をしていただいていると思っている。 ・ もう一つ重要な役割として、オーバーサイトボードは Meta が行った対策・施策に対して控訴できる。			
15. 自動モデレーション技術をすり抜けてコンテンツが公開され、問題となった事例について			・ 投稿時の審査だけでなく、一定数の再生回数が進んだ動画についてもモデレーションを実施。 ・ 具体的な事例については回答を控えさせていただきます。		・ 特にSNSだけではなく、ウェブサイト、インターネット全体に関して、今未曾有の非常に厳しい課題に全てのプレーヤーが直面。 ・ これに対処するために膨大なコストが今必要になってきているという状況。SNSだけではなく、例えば、eコマースサイトであったり、あるいは、レビューサイトであったり、ウェブ全体、インターネット全体を取り巻く問題であるというふうに認識。 ・ この問題に対処するためには、機械だけ、あるいは、人手だけということではやはり対応が現在できておらず、基本的には、人に対する人の知識あるいはスキル、それと、機械的な対応というものをやはり組み合わせしていく必要。 ・ 特にマシンの検知について重要なものは、より早期の段階でこういったものを検知することがマシンの場合はできるという可能性が高く、ここをまずマシンを活用し、さらにそれに対して人為的な対応、スキルを組み合わせ、さらにこれに、検知するマシンだけではなく、テクノロジー的なツール、これを人為的なものに組み合わせることで最終的に対処。 ・ 当然、完全なものということは今まだ実現していないし、もちろん誤

(設問)	Google	Meta	TikTok Japan	Microsoft	X
					検知というようなものもあると思うが、これは今、例えばバッドアクターという、いろんなものを仕掛けてくる人たち、あるいはグループに関しては、様々な形でクラウドを利用する、トロールファームをクラウドで構築する、あるいは、それを利用したりということで、広範な形のテクノロジーを利用したエコシステムを介してこういったものを行っている。 ・これに我々が対峙していくためには、コストの問題、こういったものの調査、研究にも、コスト、効率というところも含めて考えていく必要があるのではないかと思うが、これが今グローバルな視点から見た現状の状況。
16. 児童の保護に関する特別な配慮（プロファイリング規制等）の有無について			・偽・誤情報に限ると特別な機能等はないが、そもそも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できない。 ・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、様々な取組を行っている。		
17. コンテンツモデレーションの段階ごとの線引きについて			・偽・誤情報が含まれている可能性があるコンテンツを調査するチームが未確認情報だと判断したコンテンツには「信憑性未確認ラベル」を付与、自動的におすすめフィードの対象外になる。 ・おすすめフィードの対象外になるものは他にも多くある（陰謀論など）。		

（設問）	Google	Meta	TikTok Japan	Microsoft	X
18. 警告ラベルの種類について			・ 主なラベルはコロナ感染症と選挙に関して信頼できる情報にアクセスできるもの。		
19. 信頼できる情報を目立たせる（いわゆるプロミネンス）取組について			・ 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。 ・ 災害情報に関しては偽情報云々以前に、まずは正しく、今何が起き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にしている。 ・ 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。	・ 情報の完全性という点で 2 つ観点があり、信頼性インデックス、信頼性指標というもの、これは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。 ・ ファクトチェックの観点では、クレームレビューというものがあり、外部のファクトチェックを行う組織を使って確認。	・ 我々のポリシーは、例えば、フェイクであったり、なりすまし、ID 自体が偽っている、こういったもの以外については、基本的に、ほかのアカウントを優先的に掲載、何らかの形で優遇しているということは一切行っていない。 ・ ただし、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、こういったものについては、特別なグレーバッジを提供するという形で閲覧していただけるようにしているが、それ以外については、特にファクトファインディングの団体であったり、大手のメディア様であったり、こういったところを特に優先しているということは一切ない。 ・ 例えば、コミュニティノートの中で、実際に情報を検証したりということで、ボランティア、任意でいろんな形でこういう情報提供したり、チェックをしている、ここは実は既存のそういうファクトチェックをする団体よりも早かったりといったことも実際に発生している。 ・ 結論として、そのような形で、特に大手のメディア、あるいはファクトチェッカーの団体様であったり、そういったところに特別な優先度を御提供しているということはない。
20. 国内の研究者への API 等を用いたデータ開示に係る課題・悩み・連携可能性等について			・ 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できる方向で検討中。	・ 持ち帰って、また別途回答させてほしいが、全体的に幅広く連携をしており、一貫してコラボレーションは進めていきたい。	・ 具体的に条件であったり、我々の想定している条件等については、書面という形で用意はできるので、後ほど提供したものを御参照いただきたい。

(設問)	Google	Meta	TikTok Japan	Microsoft	X
					・ここ一、二年、特に新しい項目として、具体的にA Iのトレーニングにこういったものを活用していくということ、A Iのモデルへの活用というところが、研究者の皆様方についても非常に興味がおありの部分であるかと思うが、新しい1つのテーマとして、A P Iの公開に関しましても、慎重に我々も考えていかなければいけない、新しい項目であろうかと認識。 ・長期的に学术界の皆様方とも連携していくということについては、Twitter の時代から、現在も引き続き我々の一つの基本的な考え方。 ・A P Iに関しましては、例えば、今取り組んでいるものとしては、無償でA P Iを公開していく。 ・公共サービス、公共交通サービスなどが一例になるが、こういった部分については、S N Sと公共の交通サービスが連携していけるような形で、無償でこちらを公開するということを、既に実施。 ・我々のS N Sとコミュニティの公のサービスが連携するということによって、さらに様々なものが改善していただくとというのが一つの方向性で、これについては無償で御提供。 ・Twitter については、これは少し財務的な状況、そういった変遷、経緯があったということは否めない分はあるが、今こういったものはまた状況が変わってきており、改めてまた長期的な取組を行っていきたい。 ・A P Iだけではなく、2点、データソースとして御提供しているものも紹介。

(設問)	Google	Meta	TikTok Japan	Microsoft	X
					・①コミュニティノートのデータ。アーカイブという形で、研究をされていらっしゃる研究者の皆様方にも公開。御活用いただくことが可能。 ・②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御利用いただける状態。 ・この手のアルゴリズムで、こういう形でダウンロードしていただいて利用していただける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なというふうに自負しており、API 以外のものについても、鋭意、我々のできる範囲の中で、データ、情報等については公開したい。
21. 日本国内で選挙や自然災害時に実施した対応及びその効果			・能登半島地震のときもそうだったが、災害時にはモデレーションチームの中で情報共有を行っている。このような視点で見えていきたいと思いますというを確認。 ・能登半島地震のときには偽・誤情報を含むコミュニティガイドラインの一斉確認も実施。これは通常のプロセスに上乗せするもの。 ・関連して、危険なキーワードやハッシュタグのモニタリングを強化。 ・地震に関するコンテンツについては、審査する際に特に偽・誤情報が含まれていないのかということに注意を払うようにという指導も入っている。 ・選挙のときも、災害時のように特別な対応をするというのはあり得る。今後のさらなる取組として、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政府に近い NPO が作成しており、ラベルからリンクさせやすい。日本でもそのよ		・災害、あるいは、選挙期間については、通常とは違う事態であったり状況だと考えており、ここに合わせた特別な対策を基本的には切り分けて実施。 ・3段階の、3 Tier という呼び方をしているが、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と、3階層で対応。 ・①人員の部分については、災害あるい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、ここにはマシンラーニングのモデルを活用したり、こういったものも含めて、稼働し、十分な対策を取れるような体制を実施。 ・②ポリシーにつきましては、選挙あるいは災害が発生したときに対応する専用のポリシー、あるいはルールを策定。 ・具体的には、例えば、選挙につい

(設問)	Google	Meta	TikTok Japan	Microsoft	X
			うな選挙情報に関し、特に信頼できる情報が集まった場所を、政府あるいは関連する機関でまとめていただけるようなことがあると、プラットフォームの方でも、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。		ては、選挙関連、あるいは政治的な関連と思われる広告であるとかについては、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プロダクトについても、ここを改変し、強化をするということによって、この機能を強化していく。 ・例えば、マシンラーニングであったり、あるいは、ヒューリスティックなという言い方をしているが、こういった形で、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て、対応できるようなプロダクト上の改定なども鋭意実施。 ・人為的な対応、ポリシーの設定、プロダクトの改定、こういった3層にわたる取組を、緊急事態であったり選挙等については別途行っているという状況。
22. 選挙期間中にコンテンツの削除の判断をするために必要な期間について			・選挙等を含めた偽・誤情報についてもコミュニティガイドラインに含まれている。コミュニティガイドラインや、実際のモデレーターが従っているさらに細かいルールをきちんとローカルの文脈でかみくだいて、モデレーターが理解して対応していくというところを強化。したがってモデレーターも選挙等への対応は常にできる状況。 ・削除等に関する期間は現時点で一定のものが公表できるものではない。		
23. 検索サービスでのモデレーションに関する日本の状況について				・言語等を理解したローカルの方々との連携が重要である認識。 ・LinkedIn は、ファクトチェックと	

(設問)	Google	Meta	TikTok Japan	Microsoft	X
				いう機能があり、各国の言語をしっかりと理解した国内のファクトチェックチームの方々から直接フィードバックを得るということを実施。 ・クレームのレビューのプロセスだけではなく、ローカルのチームのレポート、ファクトチェックを合わせることで、どのような課題があるのかをしっかりと理解し、改善に努めている。 ・日本だけではなく、様々な国において、データソース、さらに多くの情報源をつなげていく、その情報源との連携を改善するように努めている。この情報源というのが信頼できる、かつ権威ある情報源であり、より多くつなげていくことで、さらに信頼ができる、そして、様々な多様性の視点から、特にこのような危機的な情報、災害があるような状態において、エコシステムに対して信頼できる情報を展開する、そのような取組を実施。 ・特に災害時においては、偽情報・誤情報というのは混乱ですとかを招きかねないため、そのような局面においては、特に信頼できる情報を展開するべく、より多くの情報源と連携をしていく。	
24. ニュースフィードのパーソナライゼーションについて				・アルゴリズムの選択については、ユーザー側でアルゴリズムの選択は不可能。 ・こういったアルゴリズムが使われているかについても、幾つか複数の要素が絡み合っており、実際どのように行われているか確認をして、フォローアップとしてお話したい。 ・アルゴリズムの選択はできないが、例えば、こういったトピック、	

(設問)	Google	Meta	TikTok Japan	Microsoft	X
				どういった分野の情報を出したいかというような選択をすることは可能。例えば、旅行といったような興味のある分野に関する情報を出すといった選択をすることは可能。	
25. 検索エンジンにおける、情報の信頼性の確保について				・ B i n g は、最も信頼ができる信頼性の高いサーチエンジンであると自負。他社製品に関しては、我々としては、B i n g が最も信頼に足るものだと認識。 ・ トラストワージネスインディケーター、信頼性のインディケーターというものがあり、それを基にEUへ報告をしているのは、マイクロソフトのみで。 ・ 検索だけではなく、その他の製品に対しても取り入れており。ドメインレベル、メディアレベル、また、情報源のレベルにおいてシグナリングを収集して、信頼性を確認するというのがやはりユーザーにとってはとても重要。単にファクトチェックをするだけではなくて、様々なレベルにおいてシグナリングをもって信頼性を確認し、担保するという取組を実施。 ・ それにより、例えば、信頼性があまり高くないような情報に関しては、透明性がこれだけ担保されていないので、こういった面で信頼性があまりないというような情報を開示することで、それを基にユーザーのほうでナビゲートするといったことも可能。 ・ 情報源が信頼に足るものなのかどうかというところの担保に関しても、とても重要で、B i n g だけではなく、ほかの製品群に関しても、そういった情報をユーザーに提供す	

(設問)	Google	Meta	TikTok Japan	Microsoft	X
				ることで、その理解を基に情報を得て、それを表現の自由に生かしていただくように取り組む。	
26. 外部からの情報提供の重要性、日本における取組への影響について				・ いろいろな様々な要素、要因があって、それを基にサーチの結果を示し、それを基に権威あるものであるかというような判断をしているという話をしたが、サードパーティのデータは、もちろん重要ではあるが、それが基で決定的なものになるということではなく、たくさんあるファクター、シグナルの一つ。 ・ 例えば、選挙に関連する権威ある情報ということに関しては、選挙、実際に選挙を運営している当局と、直接そこに情報を取りに行つて、正しい直接の情報、最新の情報を確認して、それをトラストシグナルという形で反映しながら表示。 ・ 常に追加のデータソース、データ源というものをグローバルで模索しており、日本においても、幅広いところで、この追加の情報源を得ながら、信頼できる情報提供をしていきたい。 ・ 一つの情報源だけでなく、複数から情報源を得ていくというところが非常に重要。例えば、最近、Reporters Without Borders という組織とパートナーシップを組み、ここがその信頼性の評価なども行っている。こういったコラボレーションや、情報源との提携・連携を今後もさらに拡大し、進めていきたい。	
27. コンテンツモデレーションの機能の有効性を社内で評価する取組について				・ LinkedIn、Microsoft Start に関しては、我々のほうでそのようなレポートをして、我々自身のトランスペアレンシー、また、我々自身の評価がどういったレベルで行っている	

(設問)	Google	Meta	TikTok Japan	Microsoft	X
				のかといったメトリックスも収集しながら、評価、そして、アセスメントを実施。 ・例えば、フェイクアカウントやなりすまし、偽情報といったものに関しても検知を行っており、99.6%が自動で検知され、対応。ユーザー側からレポートされる前に検知ができたものが99.7%ということで、そのような形で、継続的に計測、そして、評価、アセスメントを我々としては行うことで対応を実施。 ・コンテンツモデレーションのポリシーがどれだけ有効かというものに関しても、我々のほうで評価し、それを Transparency Report として展開。	
28. コミュニティノートを実際に導入した前後で、偽・誤情報の流通・拡散状況に何らかの影響					・効果自体については、具体的に検証できつつあると考えており、フォローアップ自体が35%、あるいは、具体的なシェアであったり、こういったものが大幅に低下するというようなところも確認。 ・この部分については、さらに推移を見ていきたいと思うが、効果は基本的には定量的にも表れているという認識。
29. コミュニティノート等、ボランティアベースの取組の限界について					・この数年という時間軸で、非常にスローなペースで徐々に拡大。この拡大のペースが今スローだというのは、まさしく、ボランティアの皆さんに依存しているということもあり、ただ、考えられるようなバイアスであったり、こういったものを当然検知し防いでいく。あるいは、返していくということも当然必要になるので、そういったものも含めて、これまたスローペースで拡大をして

(設問)	Google	Meta	TikTok Japan	Microsoft	X
					きている。 ・一つの重要なポイントというのは、アルゴリズムだと思っている。このアルゴリズムによって、具体的な、我々側のほうから今度はスクリーニングをして、バイアスであったり、こういったものを防いでいくということ、ここが一つ重要。 ・もう一つは、具体的なデータ、特にコードを公開する、パブリッシュするということが非常に重要。公開すると同時に、こういったコードであったり、アルゴリズムについても、透明度を持って、トランスペアレントな形で改善をしていくということ、ここが非常に重要。 ・これによって、具体的なユーザーの皆様、それから、学術関係の各先生方も含めた皆様方に公開し、公開した形で改善をしていくということによって、公正な形でバイアスを避け、さらに、こういった協力者、ボランティアのベースも拡大していくことができるのではないか。 ・また、公開することによって、常にこれをリアルタイムでモニターをして、問題点については対処がすぐできるような形としていきたい。 ・実際には、民主化、デモクラタイゼーションというところを旨としているが、その一方で、我々自身が何らかの問題については、これを守ってガードしていく義務も当然そこにあると考えており、その中で、このアルゴリズムを公開し、常に皆様方にモニタリングしていただく中で、公開し、改善していけるような、そういう体制を持つということが、非常に重要なアプローチ。

(設問)	Google	Meta	TikTok Japan	Microsoft	X
30. スパムと分類された投稿の中に、誤情報あるいは虚偽情報に該当する件数について					・ 具体的数字の内訳の中で、どの程度が偽・誤情報に関わるものだったのかという部分は、ぱっと出せるものもなく、また、偽・誤情報に係るものが、必ずしもこのポリシーによって対応されないケースもある。 ・ 2022年から2023年に対して、Platform manipulation and spam policy のバイオレーションの数字が一気に上がったのは、弊社側でそういったものを、いわゆるバッドアクターをどのように取り締まるかというところの対応をかなり強化した。 ・ イーロン・マスク自身がbotを非常に嫌っているところがあり、彼のこれにける情熱というのはすさまじい。さんざんユーザーの皆様から文句を言われたようなこともあったが、例えば、ベリフィケーションのプロセス等、始めるに当たって様々なコメントはあったが、そういったことが今確実にここに効き始めているのかなと思うし、ポリシーだったりディテクションの観点でも、AI等を使ったりしながら、確実に数字は強まっている。 ・ 2023年という年が、LLMが非常に躍進した年だったのかなと思っており、そういった点で、バッドアクターが非常に増えたという外的要因もあった。
31. コミュニティノートの運用の透明性について					・ コミュニティノートの参加についての基準というのは、明確な基準を持っている。 ・ 決して何らかの形でペイドユーザーだけということではなく、そちらのページにアクセスをしていただき、詳しい基準、条件等が記載して

(設問)	Google	Meta	TikTok Japan	Microsoft	X
					いるので、ぜひ御参照いただきたい。 ・コミュニティノートの担当VPが、どういった課題感が考えられるのかみたいところで、各マーケットによってひょっとしたら違う状況があるのかもしれないというようなことも言っており、何か気になる箇所があれば、その辺りを例えば調査をしていく等、何らか御教示いただけると非常にありがたく、弊社側でも御協力できることがないかなというふうにも模索している。
32. コードを公開することによる一般利用者へのリーチについて					・一般のユーザーにとってこれが分からない、専門的なものだという御指摘についてはおっしゃるとおり。 ・タブについても、フォーユーというものと、アカウントをフォローのタブがあるが、その間の部分というのは必要なかもしれないし、確かに、今のアルゴリズムのコーディングについては、そこの専門家の方であったり、あるいは、そこに関連するエンジニアの方向けということになるので、現状ではそのようになっているという認識。