

情報の誠実性のための 国連グローバル原則

マルチステークホルダー型の行動に向けた提言

(国連広報センターによる日本語仮訳 2024年10月作成)

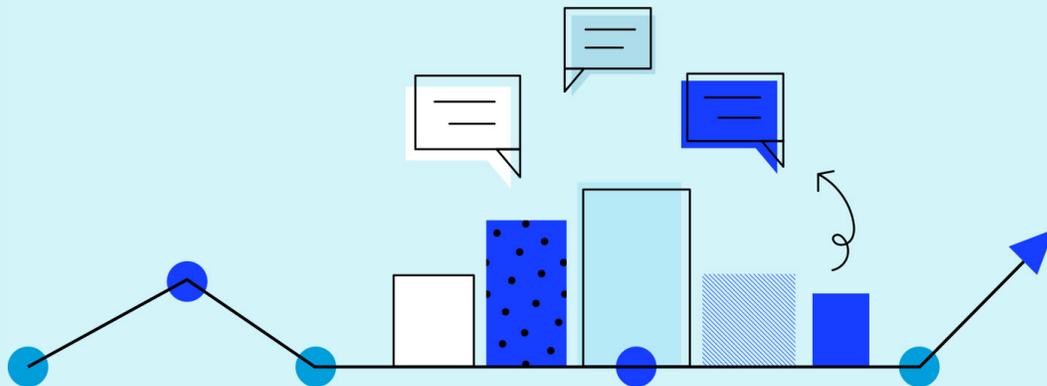


**United
Nations**

目次

デジタル時代の情報エコシステム	3
情報の誠実性と持続可能な開発目標（SDGs）	4
「情報の誠実性のための国連グローバル原則」策定に向けて	5
情報の誠実性のための国連グローバル原則	7
社会的信頼とレジリエンス（強靱性）	8
健全なインセンティブ	10
人々のエンパワーメント	12
独立した自由で多元的なメディア	14
透明性と研究	16
行動要請	18
テクノロジー企業	19
人工知能（AI）関連主体	25
広告主	27
その他の民間主体	29
報道機関	30
研究者と市民社会	32
各国政府	34
国連	38
今後の道のり	40
参考資料	41

デジタル時代の 情報エコシステム



テクノロジーの進歩によって、コミュニケーションはほんの数十年で革命的に進化し、個人やコミュニティをそれまで考えられなかった規模でつなげつつ、知識の普及や文化的豊かさの向上、持続可能な開発にとって未曾有の機会を作り出しています。そして多くの意味で、情報エコシステムの誠実性に対する期待も高めました。それは、表現の自由が十分に確保され、包摂的で安全で安心な開かれた情報環境の中で、差別や憎悪のない正確で信頼できる情報が誰でも手に入れられるエコシステムです。

これらの前進によって、情報の大量拡散が可能になる一方で、あらゆる種類の主体が歴史的に例を見ない量と速さ、拡散性をもって誤情報や偽情報、ヘイトスピーチを拡散することも容易になったために、情報システムの誠実性がリスクにさらされています。このようなリスクには、人工知能(AI)テクノロジーが急速に躍進する中で生じている現在の、そして新たな、さらには未来の脅威が含まれています。

このようにして情報空間の誠実性が侵害されれば、人々が人権を行使できる能力が損なわれ、地球上で平和や繁栄、住み続けられる未来を実現するための取り組みが妨げられることにもなりかねません。情報の誠実性強化が、現代の喫緊の課題の一つとされているのは、そのためです。

情報の誠実性には、人権や平和な社会、持続可能な未来を擁護する多元的な情報空間が欠かせません。そこには、すべての人に信頼の醸成や知識の向上、個人の選択を促すデジタル時代を実現することも含まれています。

情報の誠実性の促進には、人々があらゆる種類の情報と考え方を求め、受け取り、伝えるとともに、干渉されことなく意見を持てるよう、人々をエンパワーメントすることも含まれます。そのためには、ますます複雑化する情報環境の中で、個人がプライバシーと自由を守られながら、情報空間の中を安全に歩めるようにしなければなりません。

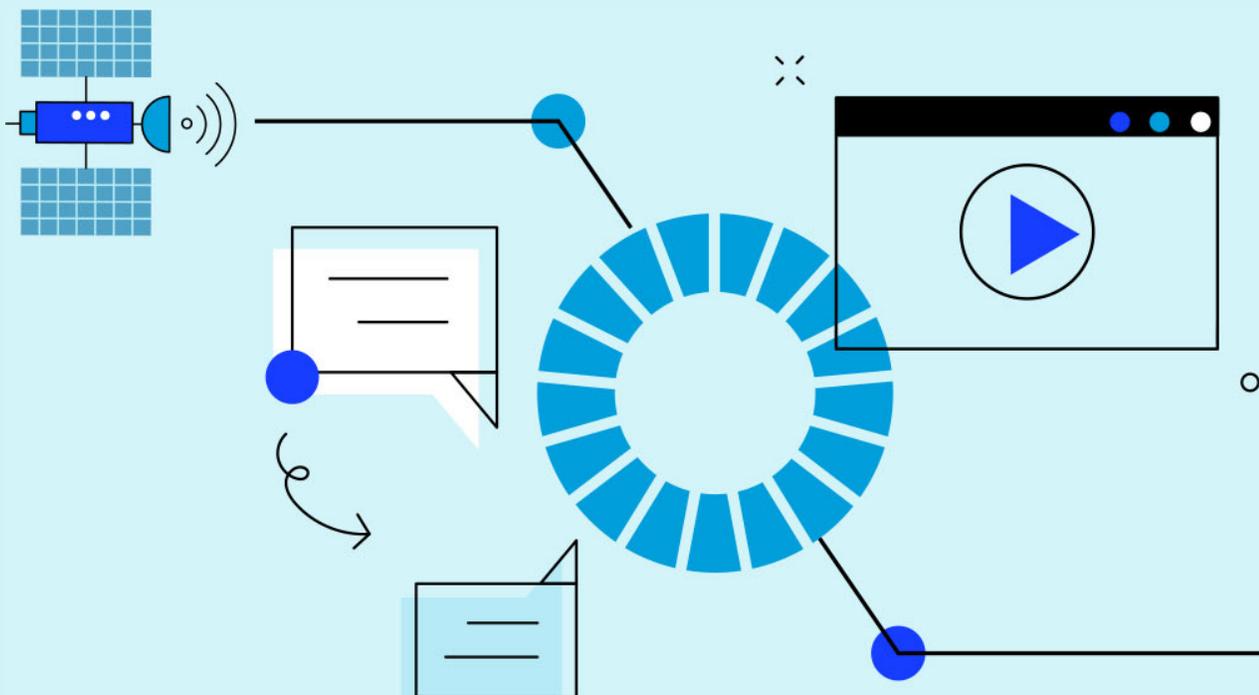
情報の誠実性と 持続可能な開発目標 (SDGs)

情報の誠実性強化のための取り組みは、持続可能な開発目標 (SDGs) の達成に向けて、これまでの成果を守り、さらに前進させるためにも欠かせません。情報の誠実性が損なわれれば、特にグローバル・サウスの国々で、SDGs を達成するうえでの既存の脆弱性がさらに悪化しかねないからです。

脆弱な立場にあり、社会から疎外された状態にある集団は、不当に大きな影響を受けます。例えば、SDGs を達成するためには、より多くの女性が全世界で労働力に加わる必要があります。しかしながら、多くの国では差別的な法律や政策が存在することに加えて、女性を黙らせ、公的領域から締め出すことで、女性を組織的に従属させる手段として、ジェンダーに基づくヘイトスピーチや偽情報、暴力が用いられています。これは女性の参加に破壊的で長期にわたる影響を及ぼし、女性の声を抑圧したり、自主検閲を強いたり、職業上、風評上の損害を与えたりすることで、苦勞して勝ち取ったジェンダー平等の前進を脅かすことにもなりかねません。

気候変動対策を損なうために情報空間を悪用しようとする企みが、この課題の緊急性をさらに高めています。商業的利益が絡むことも多い組織的な偽情報キャンペーンは、気候変動対策の目標に向けた行動を遅らせるか頓挫させるために、人間の活動による気候変動とその原因や影響について科学的に合意された礎を否定するかこれに疑問を投げかけることを目的としています。活動家や科学者、メディア関係者など、気候危機に関する情報を提供したり、これに対処したりしようとする著名人は、ヘイトスピーチや脅迫、嫌がらせの対象となっています。

健康の促進や飢餓の撲滅から、平和や公正の促進、教育の発展、不平等の縮小に至るまで、SDGs の全分野を通じ、情報の誠実性を強化しようとする対応は、持続可能な未来を達成し、誰一人取り残さないための取り組みを後押しすることになるでしょう。



「情報の誠実性のための 国連グローバル原則」策定に向けて

国連はすべての地域で、情報の誠実性に関し、加盟国やユース団体、メディア、学术界、民間セクターの代表を含む市民社会と広範かつ多様な協議を行いました。ステークホルダーは国レベルでの話し合いやバーチャル会議、二者間会合のほか、全世界に発信された公開オンライン・フォームを通して意見を表明しました。

これらの協議では、あらゆる地域と状況で適用でき、かつ、あらゆる個人の要求に取り組み、特に脆弱な立場に置かれ、社会から疎外された状況にある集団のニーズに寄り添う提言を集めることが必要だということが明らかになりました。

これを受け、「情報の誠実性のための国連グローバル原則」は、より健全な情報エコシステムを目指す様々なステークホルダーによる行動の指針となる総合的な枠組みを提供しています。この枠組みは、情報の誠実性強化のための5原則からなり、その各々に主要なステークホルダー・グループへの提言が含まれています。

その原則とは、社会的信頼とレジリエンス（強靱性）、独立した自由で多様なメディア、透明性と研究、人々のエンパワーメント、そして健全なインセンティブの5つです。その中心にあるのはいずれも、人権に対する揺るぎないコミットメントです。

グローバル原則は、各国や市民社会、民間セクター、その他のステークホルダーがこれまでに成し遂げた幅広い取り組みや進歩を認識するとともに、これらを踏まえてつくられました。そして、かつてない規模、速さ、そして強さ

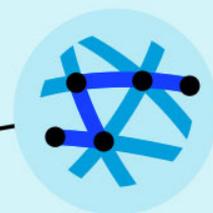
でグローバルな連帯と幅広い対策が必要とされていることを認識しながら、あらゆる職業とあらゆる言語、状況で情報の誠実性を保護、促進するための統一的な出発点を提供しています。

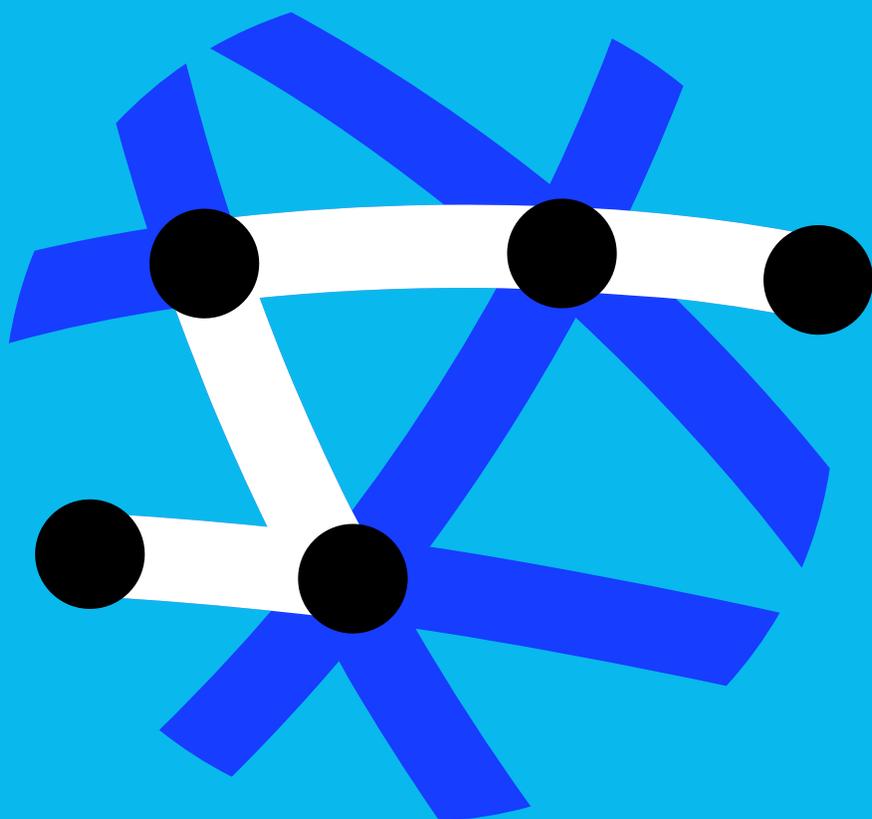
グローバル原則は個人、国連、各国政府、メディア、市民社会組織、さらにはテクノロジー、広告、広報（PR）業界にまたがる営利企業を含む官民の組織に、国際法に謳われた権利や自由に沿って情報の誠実性を実現するための幅広い提携を結成するきっかけを提供しています。

グローバル原則は「私たちの共通の課題」と国連事務総長の政策概要8「デジタル・プラットフォームに関する情報の誠実性」で提案された考え方に基づいています。

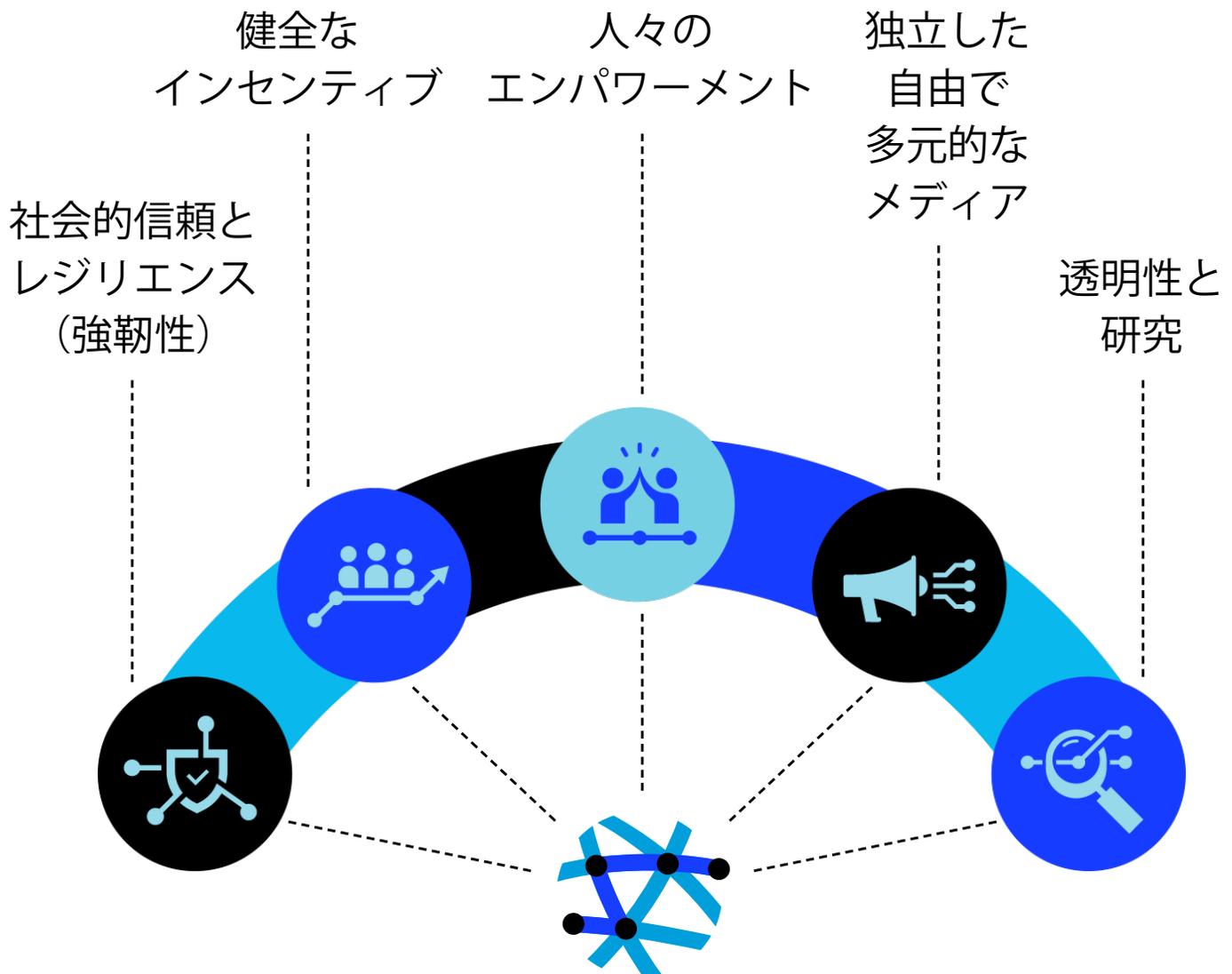
グローバル原則は、国際人権法を含む国際法に根差しているだけでなく、これと関連する「ビジネスと人権に関する国連指導原則」、「デジタル・プラットフォームのガバナンスに関する UNESCO ガイドライン」、「ジャーナリストの安全と不処罰の問題に関する国連行動計画」、「人工知能の倫理に関する UNESCO 提言」、「ヘイトスピーチに関する国連戦略・行動計画」を補完します。グローバル原則は、国連加盟国が「未来のための協定」と「グローバル・デジタル・コンパクト」の策定に向けて検討を進める際のリソースにもなります。

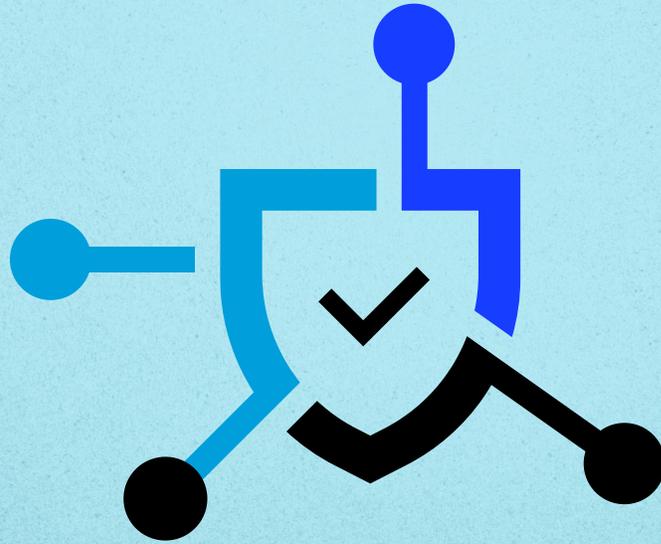
このようにして、グローバル原則は情報の誠実性に対する国連の揺るぎないコミットメントをさらに反映すると同時に、将来に向けた国連の活動の指針としても策定されています。





情報の誠実性のための 国連グローバル原則





社会的信頼とレジリエンス（強靱性）

社会全体の信頼とレジリエンス（強靱性）は、情報の誠実性の鍵となる要素です。本分野における信頼とは、公的な情報と情報源を含め、人々がアクセスする情報と情報源に対する信頼性、および、情報がエコシステム全体に流れることを可能にするメカニズムに対する信頼性を指します。レジリエンスとは、社会が情報エコシステム内の混乱と操作行為に対処できる能力を指します。

信頼とレジリエンスは、情報エコシステムを戦略的、政治的または金銭的な目的で悪用しようとする国家や非国家主体が駆り立てる行為に対して脆弱です。場合によっては極めて組織的に展開されるこうした行為は、幅広い危害を及ぼし、人々が科学や事実を批判的に評価する能力を危険にさらすおそれがあります。

テクノロジー分野の大企業は情報エコシステムで強大な力を有し、他社や広告主、報道機関、個人のユーザーを含むステークホルダーが情報を扱い、情報にアクセスする方

法に法外な影響力を及ぼします。生成 AI などの人工知能（AI）テクノロジーの進歩は、わずかなコストで情報空間に大規模なリスクを及ぼす手段を生み出しました。AI が生成または仲介した、本物あるいはオリジナルだと主張されるコンテンツは非常に信じられやすく、感情に響き、検知が難しく、しかもアルゴリズムを用いるプラットフォームやメディアを通じて一気に広まりかねません。これによって、信頼の低下が急激に生じ、加速化し、深まるおそれもあります。

情報の誠実性へのリスクに対処するには、デジタルの信頼と安全につながる大胆で将来志向の革新的な実践を、言語や状況の違いを超えて一貫して実行する必要があります。こうした実践は、脆弱な立場に置かれ、社会から疎外された状況にあり、潜在的な危害に不当に多くさらされている集団の知見を反映するべきです。

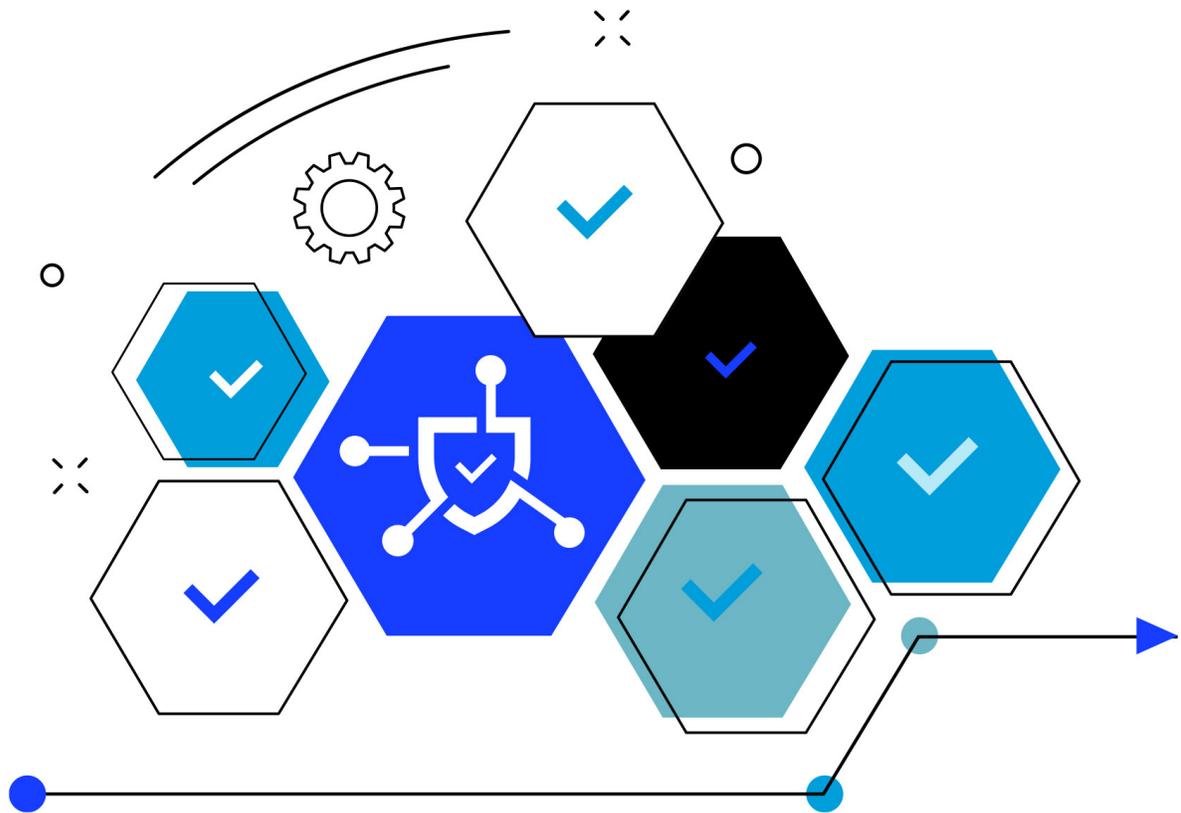
女性、高齢者、子ども、若者、障害者、先住民族、難民や無国籍者、LGBTIQ+ の人々、民族的・宗教的少数者集団を、特に考慮する必要があります。

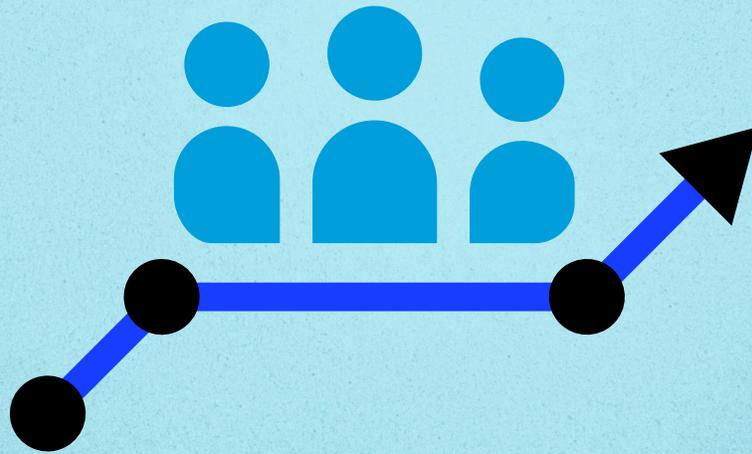
若者や子どもの多くは、その生活の大部分をオンラインで費やし、デジタル・チャンネルから膨大な量の情報を得ています。すでに情報空間へのリスクの矢面に立たされているだけでなく、今後も新興の技術やメディアのトレンドから最も直接的な影響を受けることになるでしょう。

人々は、多種多様な情報源にアクセスし、自分自身が包摂され、平等に扱われ、社会経済的安心や政治的エンパワーメントがあることを感じられる場合、よりレジリエント（強靱）で、このようなリスクを察知し、対応するための備えがよりできている傾向にあります。そうでない場合、これらのリスクが拡散する温床が大きくなりかねません。よって、対応策では、長期的なレジリエンスの強化に関する根底的な社会的ニーズを認識するべきです。

公益に資する行動にコミットしているステークホルダーはいずれも、情報空間を共通の利益のために活用することによって、絶えず進化するコミュニケーション環境の現実に対応できます。これは選挙や自然災害、人為的な危機など、情報空間へのリスクが高く、そのリスクが社会の二極化を深めたり、人々が公的領域に参加する能力を損なったり、極端な場合には、暴力の扇動に利用されたりしかねない重大な社会的転機において、特に不可欠となります。

活動家やジャーナリスト、人道支援従事者、さらには平和維持要員、選挙支援要員、科学者、医療専門家などを含む国連の人員は標的となりかねず、悲惨な結果が生じかねません。オンラインでの嫌がらせやその他の陰湿な戦術によって、声を上げられなくなったり市民社会スペースが縮小したりするおそれもあります。こうした個人を守る協調的な取り組みは最優先課題です。





健全なインセンティブ

健全なインセンティブを整備するには、収益を上げる主な手段としてターゲティング広告やその他のコンテンツの収益化に依存する今日のビジネスモデルが情報エコシステムの誠実性に及ぼす、深刻な影響に取り組まなければなりません。

こうしたビジネスモデルは大小あらゆる規模の企業、とりわけ、デジタル・プラットフォームを運用するテクノロジー企業に未曾有の成長機会をもたらし、数限りない人々が作り上げ、かつ利益を得ているクリエイター経済を生み出しました。このようなビジネスモデルは、偽情報や憎悪を広める者にも金銭的なインセンティブと機会を与えています。そうした者たちは、テクノロジー企業が広告主とクリエイターの潜在的収益を最大化するために人々の反応を引き出すことを優先するアルゴリズムを用い、そのためにユーザーの行動を追跡してデータを収集するというアテンション・エコノミーを悪用しているのです。人々の

反応を最大に引き出すことが多いのは、意見を二極化させ、強い感情を引き起こすことを目論むメッセージであり、その結果、アルゴリズムは有害なコンテンツに報酬を与え、これを増幅させることになりました。

このようなビジネスモデルを悪用する主体には、情報操作を目論む者に加えて、国家や政治家、民間セクターの組織が時には国境を越えて組織的な情報操作キャンペーンを行うために起用した主流のPR企業も含まれます。

テクノロジー部門は、デジタル広告のプロセスを、人間の監督が最低限にしか行き届かない、複雑かつ不透明なものに設計しました。これは広告技術（アドテック）のサプライチェーンを構成する多くの主体にとって好都合であるものの、最も大きな利益を得るのは大手テクノロジー企業です。

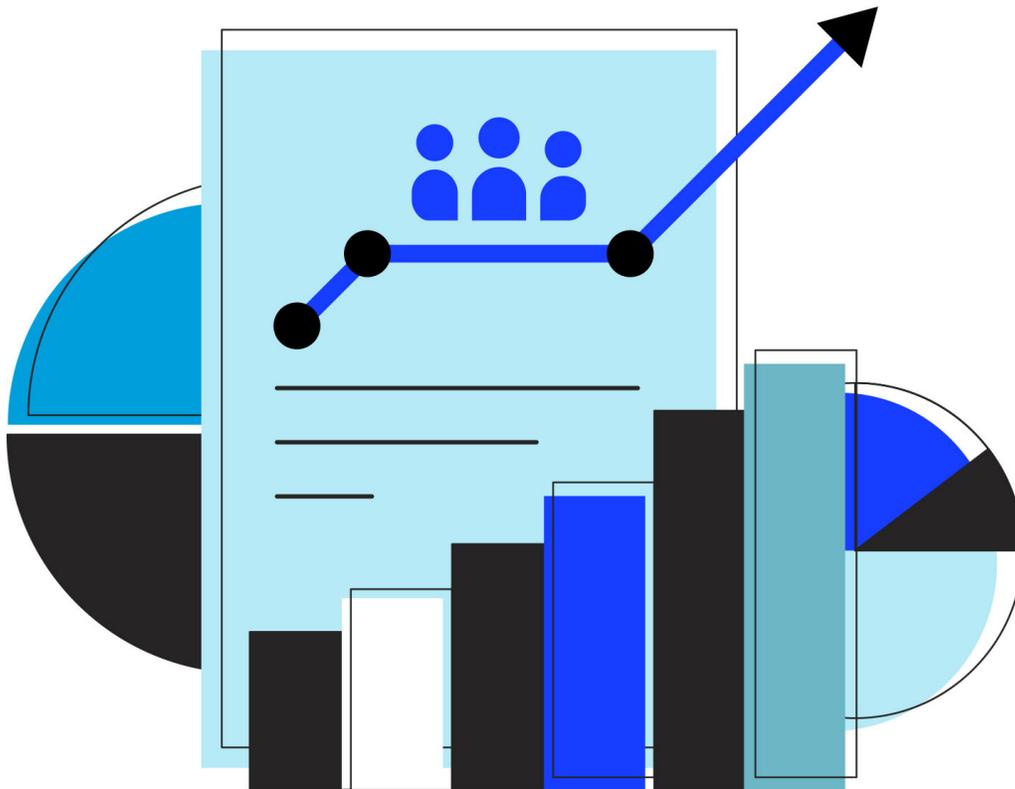
このような不透明な構造により、広告予算が図らずも、広告主が支援を意図していなかったかもしれない個人や組織、考え方のために使われかねず、ブランドにとって重大なリスクとなりかねません。このような形で広告を出稿すれば、広告キャンペーンの効果とブランドの安全性に悪影響が及ぶおそれもあります。

アドテックで優位に立つ一握りの企業には、同時に、自社が所有するプラットフォームに広告基準を設ける責任もありますが、この基準の実施は不完全で一貫性を欠くこともあります。

このような形で情報の誠実性が侵害されていることから、インセンティブ構造を根本的に転換させる必要があるこ

とは明らかです。その実現は、人権を指針とし、行動追跡と個人データに基づいてアルゴリズムによってターゲットを絞るプログラマティック（運用型）広告に依存しないビジネスモデルの採用によって可能となります。

広告主は、情報の誠実性を強化しつつビジネスの論理にもかなう形で、情報エコシステムから利益を得ることができます。テクノロジー企業が積極的に現状のビジネスモデルを手放すとは見られませんが、広告プロセスを広告主に対してさらに透明にし、広告配信者が人権面で責任ある広告方針を守ることによって、インセンティブを健全化することはできます。広告主は、透明性のあるサプライチェーンへの統制力を強めることにより、投資収益を改善することもできます。





人々のエンパワーメント

情報エコシステムを利用する個人のエンパワーメントを図るためには、人々がそれぞれのオンライン体験を自ら管理し、消費するメディアについて十分な情報に基づいて決定し、自由な自己表現ができるようにしなければなりません。人々のエンパワーメントを図るためには、多様で信頼できる情報源に常にアクセスできることが必要です。

デジタル空間は多くの意味で、地理的境界を越えて進歩への想いを共有する人々をつなげることで、公的領域への包摂的な参加を促す触媒となってきました。このような空間を良いほうに活用すれば、個人のエンパワーメントを図り、しばしば社会的に排除や疎外されている人々に力を提供する一助となる可能性もあります。

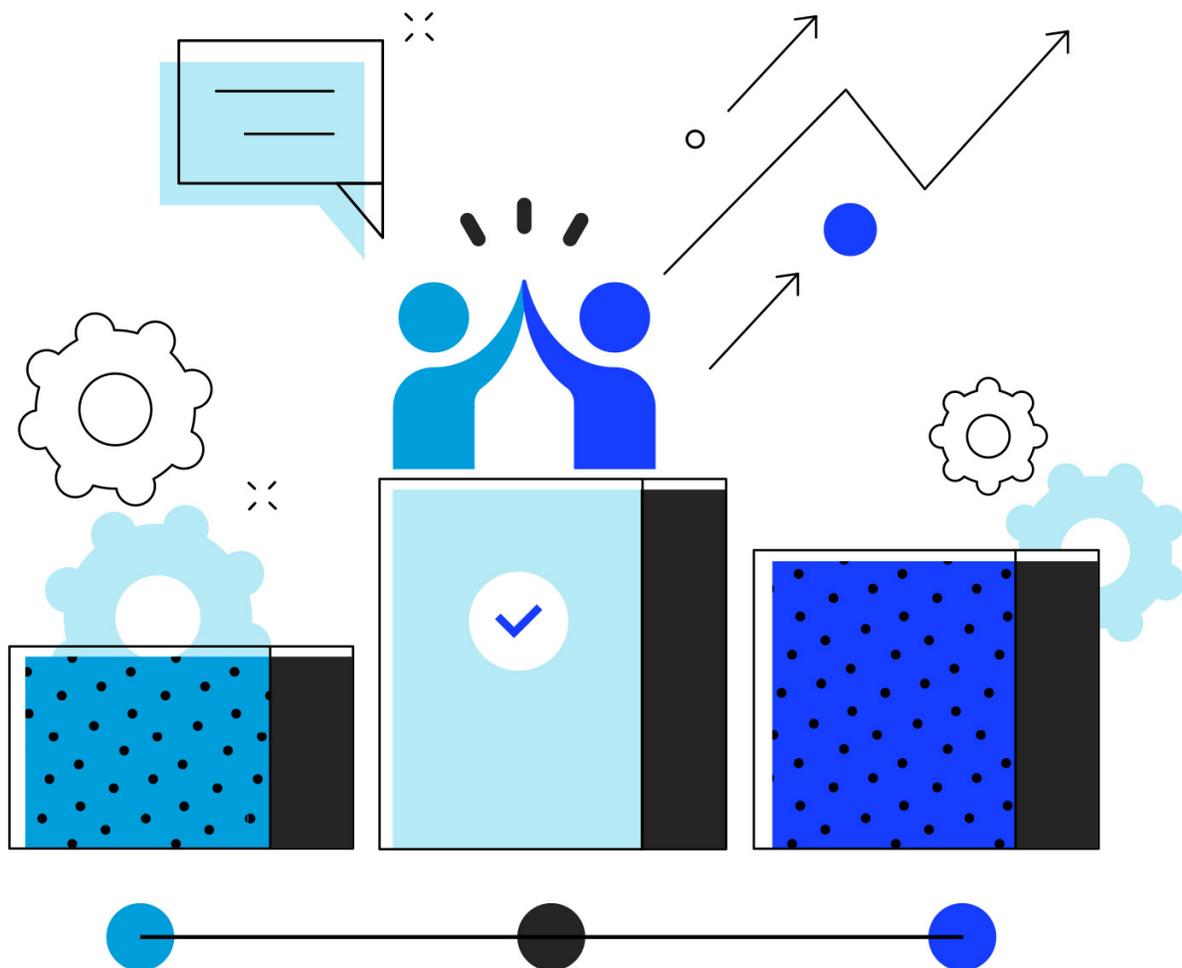
その一方で、デジタル技術は本質的なエンパワーメントを阻害するおそれもあります。ユーザーは個人データの用途についても、大手テクノロジー企業によって個人別に選定されたアルゴリズム・コンテンツについても、ほとんど自身で管理することはできません。また、情報プロバイダーが特定のコンテンツを優先づけ、推奨するために用いている判断基準やメカニズムに対して、ユーザーが理解したりアクセスしたりすることも妨げられています。

テクノロジー企業は、ユーザーのプライバシー権を尊重し、信頼と安全、プライバシー方針、データ利用のあらゆる側面について、ユーザーが意見やフィードバックを提供できるよう、ユーザーのエンパワーメントを図るべきで

す。また、異なるプロバイダーが提供する幅広いサービスとの互換性を含め、ユーザーの決定権と選択権を強めるべきです。

メディア・リテラシー、情報リテラシー、デジタル・リテラシーのトレーニングに関する取り組みでは、あらゆる人のエンパワーメントに関心を集中させるべきです。特に、女性や高齢者、子ども、若者、障害者、脆弱な立場に置かれ、社会から疎外された状況にある集団が直面する特有の課題に焦点を置く必要があります。

インターネットへの接続率は上昇しているものの、世界の3分の1は今でもオフライン状態にあります。オンライン状態の人々でさえ、アクセスが不十分であればインターネットの資源を十分に活用する能力が阻害され、情報空間でのリスクに対する脆弱性が高まってしまいます。インターネット接続への障壁が急速に低くなる中で、新規ユーザーのエンパワーメントを図るとともに、アクセスできない人々に対しては、安全で生産的なオンライン体験に必要なデジタル・リテラシー能力を身に着けるための取り組みを導入する必要があります。





独立した自由で多元的な メディア

独立した自由で多元的なメディアがなければ、情報の誠実性は実現しません。

自由な報道は法の支配の基盤であり、民主的社会の礎石として十分な情報に基づく市民の話し合いを可能にし、権力者の責任を追及し、人権を守る役割を果たします。女性や脆弱な立場に置かれ、社会から疎外された状況にある人々を含め、ジャーナリストやメディア関係者が常に安全かつ包み隠さず報道し活動する自由を認められ、あらゆる個人が多元的で信頼できるニュース・ソースに一貫してアクセスできるのであれば、報道は自由であるとみなされます。

メディアには、信頼できる正確な情報を提供し、情報空間に対するリスクを軽減するうえで、特別な役割と責任があります。しかし、表現の自由に対する権利が認められているにもかかわらず、自由で検閲と妨害を受けない報道機関やその他のメディアを含めた報道の自由は、大きく継続的な脅威に直面しています。メディア関係者はオンラインでもオフラインでも、嫌がらせや脅迫、暴力の対象となっており、それが自主検閲や職業上のリスクの増大といった結果を招くこともあります。

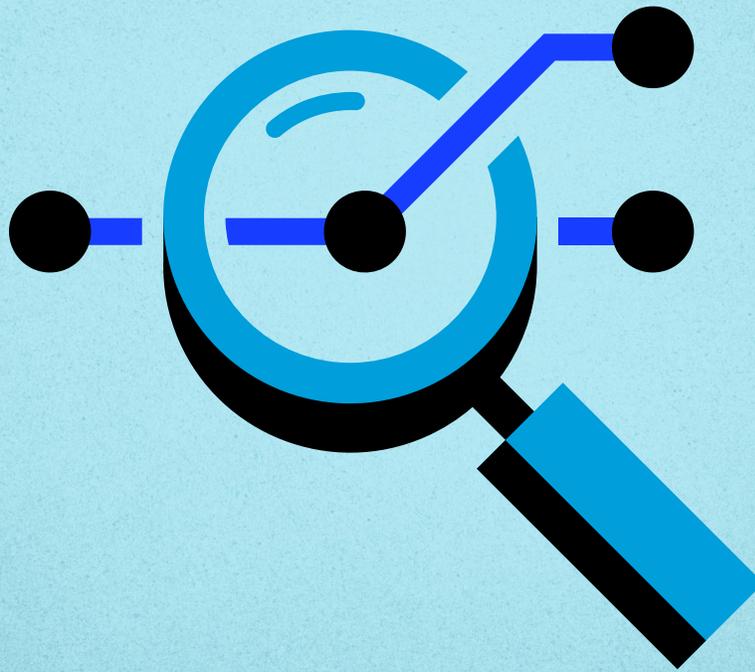
同時に、報道業界は広告収入を大手テクノロジー企業が支配するデジタル空間に奪われるという苦難も抱えています。このような要因により、企業の利害がメディア各社に対する締め付けをさらに強め、メディアの多様性を脅かすとともに地域に根付いたジャーナリズムや公益のためのジャーナリズムを損なうことを許しています。編集基準が強固に守られなければ、メディアは情報の誠実性に対するリスクを高め、増幅することになり、これがオンラインとオフラインの空間で相互作用を起こすおそれもあります。

メディア・インフラが限定的なために、市民ジャーナリストが地域の住民にとって極めて重要なサービスを提供す

るという状況を鑑み、公益に資する報道機関やジャーナリスト、メディア関係者を支援するために強力な対応が緊急に必要とされています。このような対応には、地元の実践者を活用したしっかりとした、かつ持続的なメディア開発支援を含めることができます。

各国政府とテクノロジー企業は、情報の流れと政策を決定づけるうえで、大きな影響力を及ぼすため、報道の自由とジャーナリストの変わらぬ安全を確保するための取り組みを強化すべきです。





透明性と研究

テクノロジー企業やその他の情報プロバイダーが透明性を向上すれば、情報がどのように拡散されているか、個人データがどのように用いられているか、情報の誠実性に対するリスクにどのような取り組みがなされているかに対する理解を深めることも可能になります。

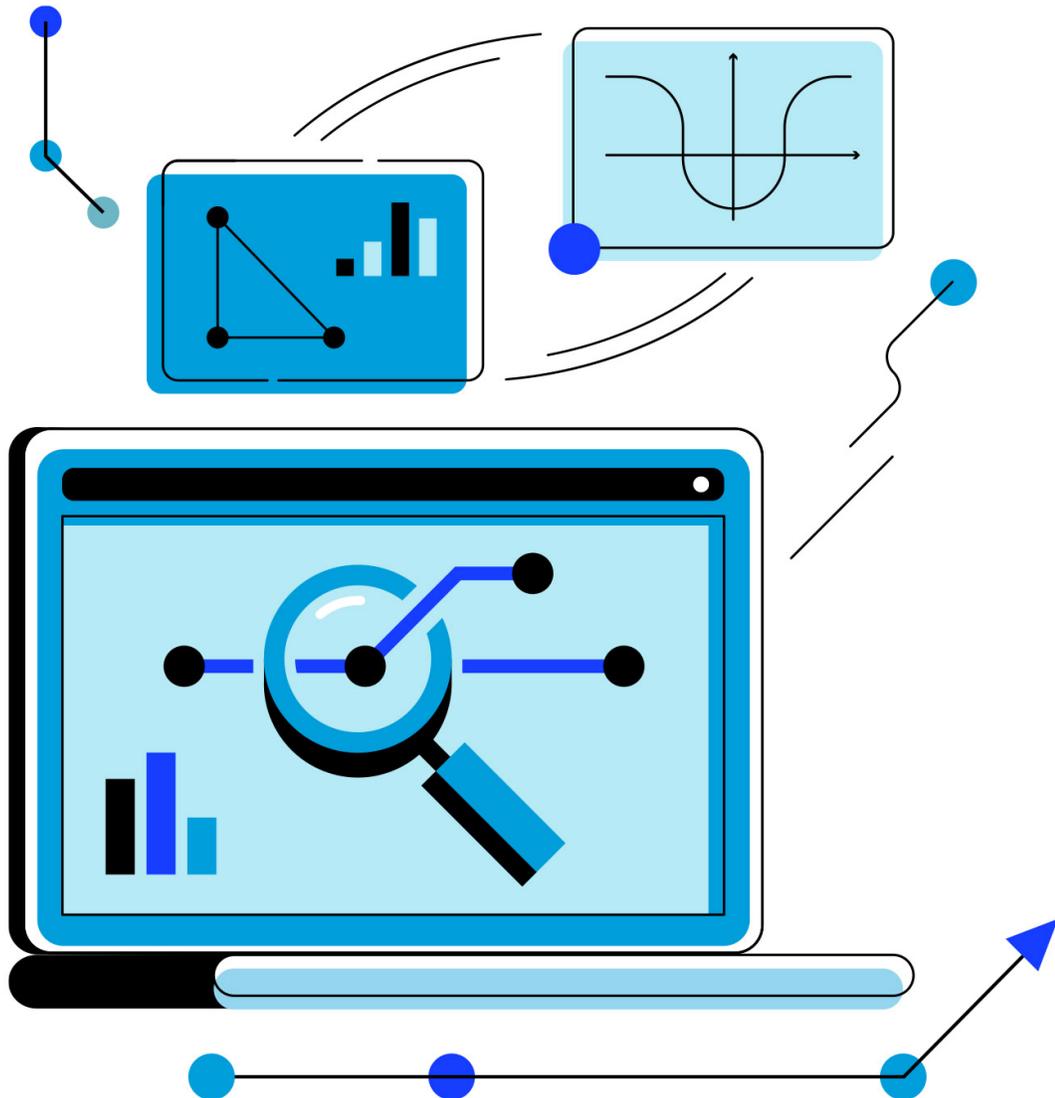
しかし、力の不均衡が透明性に対する障壁を生んでいます。一握りのテクノロジー企業が、かつてないほどの大量のデータにアクセスするとともに、一部のメディア・オーナーとともに、また、時には国家や政治的経済的主体とも密接に連携しつつ、情報エコシステムに対する強大な支配力を及ぼしているからです。

さらには、テクノロジー企業が本社を置く少数の国による透明性に関する規制の面での選択が、世界全体に大きく影響しています。こうした不均衡はしばしば公益にかなう研究を制約したり、平等を確保し十分にサービス対象・研究対象になっていない状況やコミュニティのニーズを充足するうえでの取り組みを妨げたりするおそれがあります。

AI テクノロジーは、その影響の全貌がまだ見えておらず、情報エコシステムの研究と理解に追加的な課題をもたらしています。

情報環境に対するさらにニュアンスを汲んだグローバルな理解を確立し、情報の誠実性を促進するためにより的確なエビデンスに基づく行動を強化するためには、データや知見の入手可能性と品質、利用可能性を向上させる必要があります。幅広い多様な研究者にプライバシーを守っ

たデータへのアクセスを確保すれば、研究の不足部分と不平等を埋めるための集团的取り組みを強化できるでしょう。学術研究者やジャーナリスト、市民社会が恐怖や嫌がらせを受けずに、その非常に重要な職務を果たせるよう保護、支援しなければなりません。



行動の要請

以下の提言の目的は、先の5原則を、情報エコシステム全体のステークホルダーが実行できるステップへと落とし込むことにあります。下記の提言は包括的な青写真となることを意図されているため、各国政府の法的義務からテクノロジー部門の責任、さらにはメディアと市民社会のベストプラクティスに至るまで、幅広い領域を取り扱っています。



ステークホルダーに向けた 提言

- テクノロジー企業
- 人工知能（AI）関連主体
- 広告主およびその他の民間主体
- 報道機関
- 研究者および市民社会組織
- 各国政府および政治的主体
- 国連

テクノロジー企業

技術に対する規制が限定的な場所を本拠地とすることが多い大手テクノロジー企業には、巨大な力があります。こうした企業は、ユーザーの行動について収集した大量のデータから利益を上げ、国境を越えた情報の流れを形成し、世界的な規模でデジタル体験を支配することが可能です。

この力の不均衡を是正するためには、透明性と独立した監督をともに優先する枠組みが必要です。ユーザーは自身のデータとオンライン体験を管理するとともに、明確な苦情申し立てや救済の手段を持って然るべきです。危機下や紛争下の状況を含め、その製品やサービスのデザインと使用が人権と社会の結束に及ぼす結果について、テクノロジー企業の責任を求めるためには、説明責任を求めるメカニズムが必要です。

それは、情報の誠実性を侵害し、人権を損なう機能を特定するために、プラットフォーム構造を批判的かつ透明な形で評価する必要性を意味します。表現の自由と情報へのアクセスを守りつつ、このような侵害を予防、軽減する戦略を実施すべきです。

偽情報と憎悪が最大限の露出と大きな利益を生み出すべきではありません。ターゲットを絞ったプログラマティック広告に依存しない商業的に実現可能なビジネスモデルを新たに作り出せば、イノベーションを促し、ユーザーのエンパワーメントを図り、公共の利益にも資することができます。この多面的なアプローチは、ユーザーの権利を尊重し、信頼できるオンライン環境を育てることができる、よりバランスの取れた情報エコシステムの整備にもつながります。

提言



a. 設計から実装に至るまで、安全とプライバシーを統合する。 設計、開発、実装、廃止の各段階を含め、製品とサービスのライフサイクル全体に強固な安全とプライバシーの方針を組み込む。その中で、人間とAIが生成するメディアの双方に一貫した形で、安全とプライバシーの方針を適用する。社会的な転機の前とその周辺時期を含め、能動的に社会的リスクを最小化して潜在的な危害を軽減するために、独立の第三者機関と協力し、あらゆる商品とサービスに関して継続的に人権に対するリスクを評価し、公開する。脆弱な立場に置かれ、社会から疎外された状況にある集団や、市民社会のメンバー、しばしばオンラインで標的にされているその他の人々の保護とエンパワーメントのために措置を取る。また、テクノロジーの利用を通じて生じるかそれによって増幅される、ジェンダーに基づく暴力やその他の暴力形態に対処する措置を取る。AIテクノロジーに起因する情報エコシステムの誠実性に対するリスクが蔓延する可能性を含め、新たに生じつつある課題に取り組むためにイノベーションを起こす。製品開発の全段階での人員配置において、また、信頼・安全チームの多様性と包摂性を確保する。リスクと方針の評価が、リーダーシップを含め社内のあるあらゆるレベルと部署で共有され、組織全体で理解されるよう、内部の情報共有に関する手順を確立する。信頼・安全性に関する方針が一貫して確実に実行されるようにする。



b. ビジネスモデルを再評価する。プラットフォームの構造が情報エコシステムの侵害を助長し、人権を損なうことにつながっているかどうか、また、つながっているとすればそれはどのように起きているかを評価するとともに、表現の自由を尊重しつつ相応の軽減・救済措置を取る。ターゲットを絞ったプログラマティック広告に依存せず、公益に資する革新的で商業的にも実現可能なビジネスモデルを模索する。



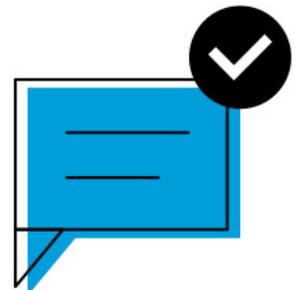
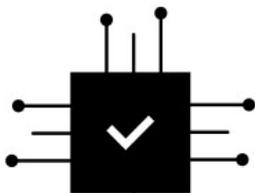
c. 子どもたちを守る。年齢認証やペアレンタル・コントロールなど、子どもの権利を保護、擁護するための措置を確立、実行する。テクノロジーの利用によって発生または増幅する子どもの性的搾取や虐待を防止し、これに対処する方針と実践を実行する。子どものための特別な通報や申し立てのメカニズムを設け、公開する。



d. 資源を配分する。リスクの水準に応じ、社内で信頼・安全性に取り組む資源と専門能力を十分かつ継続的に割り当てる。社会文化的、言語学的な背景と業務の使用言語、および、脆弱な立場に置かれ、社会から疎外された状況にある集団、特に紛争下や不安定な状況下にある人々の異なるニーズに対応するために、十分な資源を割く。



e. コンテンツモデレーション（インターネット上の不適切なコンテンツの監視と必要に応じた削除や拡散抑制などの規制）の一貫性を確保する。独立の第三者機関との協力で、国際的人権基準に沿ったコンテンツモデレーションの手順を開発し、その方針があらゆる業務領域で恣意的でない形で一貫して確実に実行されるようにする。手動と自動のコンテンツの監視と収集に十分な資源を配分し、これをあらゆる言語と業務内容に一貫して適用する。プラットフォーム・コミュニティの基準に違反し、人権を損なうコンテンツに対し、アルゴリズムによる増幅の制限、ラベリング、収益化禁止といった措置を取る。コンテンツモデレーション方針の実施と、あらゆる言語と業務内容でコンテンツモデレーションに充てられる資源に関し、細分化したデータを公表する。





f. 労働基準を守る。 国際労働基準と国際人権法と整合する勤務環境を提供するとともに、コンテンツモデレーションの担当者を含め、信頼・安全への取り組みに関わるあらゆる労働者に福祉、安全および質の高いトレーニングを確保するための取り組みを優先する。



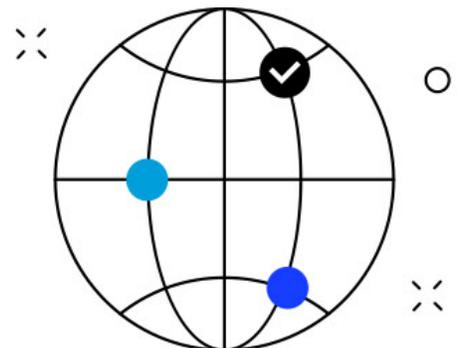
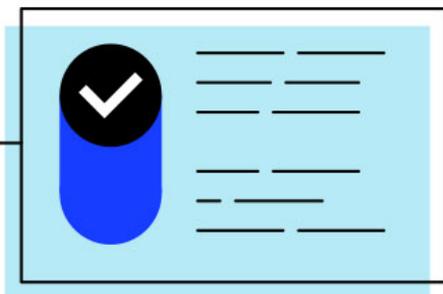
g. 独立した監督機能を設ける。 外部の第三者による定期的な人権監査を実施する。監査の対象に次のものを含む：利用規約とコミュニティー規定；信頼・安全および広告に係る方針；リスクマネジメント；あらゆる言語と業務内容に関わる広告とレコメンダーシステム（ユーザーにコンテンツの「おすすめ」を選定するシステム）の影響；コンテンツモデレーション；苦情・異議申し立て手順；透明性のメカニズム；そして、研究者によるデータへのアクセス。脆弱な立場に置かれ、社会から疎外された状況にある集団に対する、また、ジェンダーの平等および子どもの権利に対する製品とサービスの影響を評価する。これらの監査結果を公表し、あらゆるユーザーがアクセス、理解できるようにする。



h. 業界基準を策定する。 市民社会やその他のステークホルダーと連携し、役割と責任を明確に定めた業界の説明責任に関する枠組みを共同で策定し、監査済みの公開報告書の作成と第三者による監督、また、プライバシー、リスクマネジメント、信頼・安全に係る強固な基準の遵守を約束する。脆弱な立場に置かれ、社会から疎外された状況にある集団と不安定な状況に置かれた集団のニーズについて具体的な規定を設け、人権に対するリスクを測定し、これに対処する効果的な方法を確認する。リスクは多様な情報空間を横断して広がりかねず、各情報空間には悪用されかねない独自の設計上の欠陥や方針上の隙間があるという認識のもと、プラットフォームとサービスの間の協力を確保する。



i. 危機対応を向上させる。 リスクが高い領域で活動するステークホルダーと連携しつつ、危機下と紛争下で迅速かつタイムリーに対応できる、早期警報と対応拡大の手順を確認する。公益に資する信頼できる正確な情報に、明確でタイムリーにアクセスするためのメカニズムを確認する。





j. 政治的プロセスを支援する。 選挙やその他の政治的プロセスの前とその最中に、あらゆる製品とサービスの人権に対するリスク評価を行い、その結果を公表する。情報の誠実性を守るための関連方針をすべて実行し、政治家候補を含む公的な領域でしばしば標的とされている女性や他の集団に対する偽情報や嫌がらせ、暴力に対処するための措置を取る。



k. ステークホルダーと協働する。 各国政府や学术界、市民社会、子ども、若者主導の組織、技術コミュニティなど、幅広いステークホルダーと積極的に関わり合い、情報エコシステムの誠実性に対するリスクの理解を深めるとともに、それに応じて信頼・安全のメカニズムを拡張、調整する。



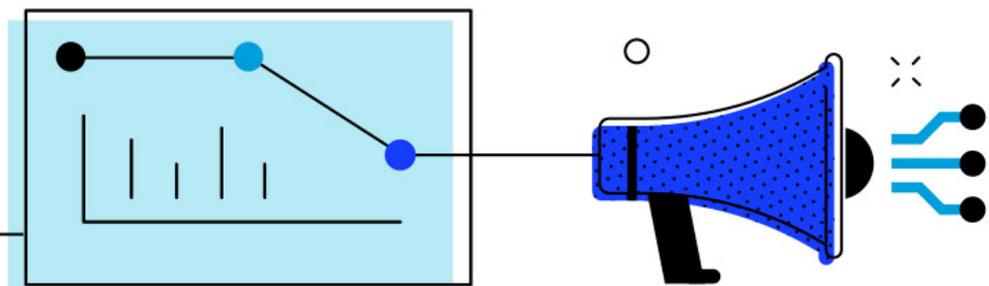
l. 強固な苦情対応のメカニズムを確立する。 脆弱な立場に置かれ、社会から疎外された状況にある集団に係る特別なプロセスを含め、ユーザーと非ユーザーのための苦情、通報、異議申し立ておよび救済のメカニズムを、確実に透明性があり、安全、安心かつアクセス可能なものとする。組織的な偽装行為など、通報・苦情対応メカニズムの悪用を防ぐ手順を確立し、履行する。



m. 明確な方針を発信する。 利用規約や方針、コミュニティ規定、実行手順を、子どもも含め、簡単にアクセス可能で一貫した理解可能なものにする。ニュースや政治的コンテンツに関する方針、ガイドラインおよび規則をすべて明確にする。



n. 広告方針を実行する。 広告とコンテンツの収益化に関し、明確で強固な方針を確立、公表、実行する。既存の出版・アドテック関連のパートナーシップを継続的に見直し、アドテック・サプライチェーン内のパートナーがこの方針を守っているかどうか評価する。毎年、方針の履行の実効性と、その他取られたいかなる措置についても報告、公表する。





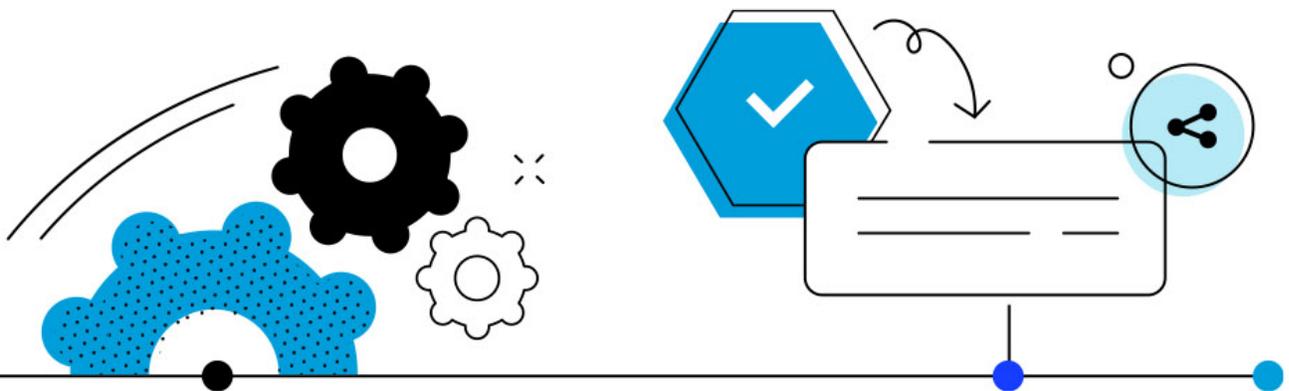
O. 広告の透明性を実証する。 あらゆる広告を広告だと分かるように明示して、広告主や、ターゲティングに用いられたパラメーター、また、AI で生成または編集されたコンテンツに関する情報をユーザーに対して透明にする。広告のソースまたは広告の支払者、広告費の額、ターゲット層に関する情報を掲載した完全でアクセス可能、最新かつ検索可能な広告ライブラリ（広告に関するデータベース）を維持する。特定の時間枠で広告がどこに掲載されたのか、また、広告出稿とブランドに隣接して表示される他のコンテンツに関する管理とサービスの正確さや実効性に関し、詳細なデータを広告主と研究者に提供する。収入源と、広告主やコンテンツ・クリエイターとの分配取り決めに関し、透明性のある報告を行う。AI が生成または編集したコンテンツの明示を含め、政治的広告はすべてラベリングするとともに、ユーザーがなぜターゲットとされたのか、広告には誰がどれだけだけの金額を支払っているのかに関し、簡単にアクセスできる情報を提供する。



P. メディアの安全性と多様性を支援する。 消費者が幅広いメディア・ソースにアクセスできるようにするために、多元的なニュース・コンテンツが発信される環境を整備する。編集の独立性を尊重しつつ、独立系の自由で多元的なメディア、特に多様な言語と状況で活動する地域と市民のジャーナリズムを支援する。ジャーナリストとメディア関係者の権利をオンラインで守るために必要なあらゆる措置を取る。特に選挙や自然災害、人為的な危機などの社会的な転機となる時期にジャーナリストが直面するリスクに鑑み、ジャーナリストとメディア関係者を嫌がらせや虐待、暴力の脅しから守ることに資する明示的で透明な規定を設ける。特に、女性ジャーナリストを標的とする行為を軽減し、対処することを目的として、信頼・安全性に係る方針と実践をアップデートする。



Q. データへのアクセスを提供する。 各専門分野の学術研究者やジャーナリスト、市民社会、国際機関に対し、ユーザーのプライバシーと知的財産権を尊重しつつ、情報の誠実性をより良く理解し、政策やベストプラクティスの参考とし、説明責任を改善するために必要なデータへのアクセスを提供する。社会的なリスク、様々な異なるコミュニティや人々に対する影響、AI テクノロジーの活用が持つ意味合い、SDGs の達成への潜在的な影響、リスク軽減措置の実効性を含め、情



報エコシステムの誠実性に関する効果的な研究を可能にするために、このような情報は細分化して提供する必要がある。この情報には、次のものを含む：コンテンツのランク付け、推奨、配信、そして注意喚起することに関してアルゴリズムがどのようなトレーニングを受けているかについての説明を含むアルゴリズム搭載型レコメンダーシステム；削除、禁止または降格されたアカウント；そして、あらゆる言語と状況での信頼・安全に向けた資源の配分。また、アクセス可能で機械可読なフォーマットで、最低限の費用で研究者へのデータ提供を促す。



R. 情報開示を確保する。 政府からのコンテンツ削除または掲載の要請を公表する。資金の提供やその他の支援を含む、ファクトチェック機関とのあらゆる協働と、政治団体や政治家候補に提供した資金を開示する。



S. 管理と選択肢を提供する。 インフォームドコンセントを確保するとともに、他のサービスとの互換性などを含み、ユーザーが自身のオンライン体験を簡単に管理できるよう、使いやすいツールや機能を提供する。そうすることで、より幅広い選択を可能にし、ユーザーが閲覧するコンテンツや、そのデータがどこで、どのように使われているかに関して、インフォームドコンセントを提供する。



t. AI コンテンツにラベルをつける。 組織レベルでソリューション開発に投資したうえで、AI が生成または編集したコンテンツに明確なラベルをつけ、ユーザーがこのようなコンテンツを簡単に見分けられるようにすることで、情報エコシステムの誠実性全般に対するユーザーの信頼を損なうのではなく強化する。その中には、AI により生成または編集されたコンテンツを特定するメタデータの情報を含める。



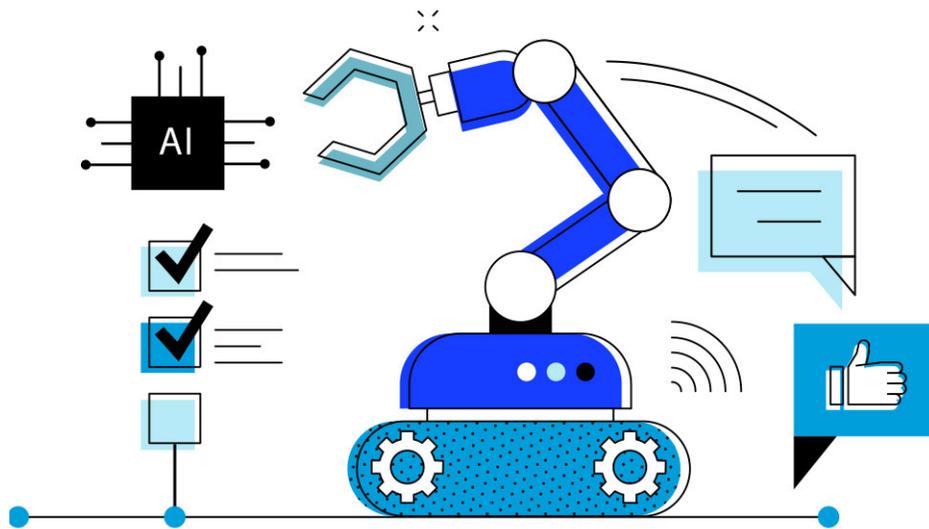
U. プライバシーを確保する。 データの収集、利用、共有、売却および保存で、ユーザーのプライバシーが尊重されるようにするとともに、アルゴリズムの決定に関するものを含め、自身の個人データがどのように使われたか、また、他の主体がどのように個人データを入手し共有したかに関する情報をユーザーが簡単にアクセスできるようにする。



V. デジタル・リテラシーを育む。 アルゴリズムの機能、効果および影響に対する市民の理解促進を含め、デジタルスキルを高めるために、メディア・リテラシーと情報リテラシーの推進を支援する。リテラシーと能力構築の資源を、あらゆる言語と業務領域、特に不安定な状況に関するものに充当する。子どもと若者に安全関連のトレーニング教材を提供する。リテラシーへの取り組みの実効性に関し、外部の第三者による評価を可能にし、その結果を公表する。

人工知能（AI）関連主体

人工知能システムのライフサイクルの
少なくとも一つの段階に関わる
政府と官民の主体



人工知能（AI）テクノロジーが急速に進化する中で、AIテクノロジーの能力は私たちの世界を変えようとしています。日常的な作業の自動化から科学的発見の補助に至るまで、莫大な潜在的利益があります。しかし、このプロセスと並行して安全、安心で信頼できるAIの設計、開発、展開、廃止を確保する必要性が大きくなります。

トレーニング・データにバイアスや多様性の欠如があれば、AIシステムは誤解を生む情報を生成し、不公正を永続化させかねません。現実味のあるコンテンツを生成する能力が大規模に悪用されれば、情報エコシステムの誠実性にリスクが生じるおそれもあります。

新たに生じつつあるリスクは、AIテクノロジーのライフサイクルで透明性と公正性を優先することによって、軽減できます。AIがそのライフサイクル全体を通じ、安全かつ責任ある形で設計、開発、展開、廃止されるようにするためには、政府やテクノロジー企業、学術研究機関による協調的な取り組みが必要です。これらのステークホルダーが協働すれば、AIが社会と人間のウェルビーイングに資するようにできるのです。

提言



a. 安全、安心で信頼できる AI を確保する。 AI テクノロジーの安心、安全で信頼できる設計、開発、展開、使用および廃止を確保するための措置を取る。AI テクノロジーの悪用、人間の監督を経ない AI テクノロジーへの過度な依存、これに関連して地域や社会的状況にまたがって信頼がさらに崩れる可能性など、AI 分野内のイノベーションや進歩で情報エコシステムの誠実性にとってリスクとなり得るいかなる物事にも対処し、これを公けに発信する。市民のウェルビーイングにとって重大な問題に関する信頼できる包摂的な情報源について AI のトレーニングを行うとともに、ジェンダーや人種のバイアスを含め、トレーニング・データに由来するバイアスを軽減するための措置を取る。女性や子ども、若者、その他の脆弱な立場に置かれ、社会から疎外された状況にある集団に対するものを含め、社会的リスクを最小化して潜在的危害を積極的に軽減するために、多種多様なステークホルダーと連携して人権に対するリスクを評価する。



b. 独立監査を委託する。 機関に所属する研究者と個人の研究者が、企業の脆弱性に関する開示方針の遵守などの適切なセーフガードを伴う形で AI モデルの独立監査を行えるよう、アクセスと法的、技術的な安全領域を提供することを確約する。独立監査の結果や、AI 関連のリスクについてのデータ（有害な差別や「ハルシネーション」、すなわち事実のように見えるが全くのどっち上げのコンテンツが生まれる可能性など）、および、潜在的な危害を防止、緩和し、対処するために取られた措置について、市民が確実にアクセスできるようにする。



c. 知的財産を尊重する。 知的財産権を尊重し、オリジナルの記事を含め、AI ツールのトレーニングに用いた知的財産の使用について、公正な報酬を確保する。



d. データの出所を示す。 真正性証明や電子透かし、ラベリングなど、可視および不可視の形態の、出所に関するソリューションや方針を開発、実施する。使い勝手の良いラベリングの標準化に向けてマルチステークホルダーで取り組む。



e. リテラシーを支援する。 組織レベルで、情報の誠実性に対するリスクに焦点を置きながら、AI の仕組みや情報消費者全体への影響に対する、市民の理解を高めるためのリテラシー向上を目的としたイニシアチブの開発と展開に投資する。



f. ユーザーのフィードバックを可能にする。 ユーザーのプライバシーを保護しつつ、不正確または誤解を招く出所情報をユーザーが警告または通報できるようにする。



広告主

広告主は、偽情報やヘイトスピーチから利益を得ようとする者に対する金銭的インセンティブを断つために支援することにより、情報エコシステムの誠実性に独自の影響力を及ぼすことができます。そうすることで、広告主は自社

のブランドをより良く守り、重大なリスクに対処し、自社の企業理念に沿った事業を営みつつ業績を高めることもできます。

提言



a. 人権面で責任ある広告を確立する。 広告が情報空間に対するリスクを高めないようにするとともに、子どもの権利を含む人権を擁護するためのセーフガードを確立する。機敏なデータやユーザー属性に基づく差別的なターゲティング手法は避ける。包摂・除外リストやアドベリフィケーション・ツール（広告の掲載に関する質を検証するためのツール）、人による審査などの方法を通じ、公益ジャーナリズムなど情報の誠実性を高める報道機関やプラットフォームに出稿する。アドテック企業に対し、ウェブサイトやチャンネルが収益化を始める前に満たさねばならない基準を公開するよう求める。



b. 業界基準を活用する。 業界基準を活用することで、情報の誠実性に対するリスクを最小限に抑え、ブランドの安全性確保に資する明確な方針を策定する。



c. 提携する。業界全体や市民社会と協働し、広告の影響評価や、広告とコンテンツの収益化に由来するリスクと潜在的危害の体系的な軽減を含め、情報の誠実性について得たベストプラクティスや教訓を適時共有する。



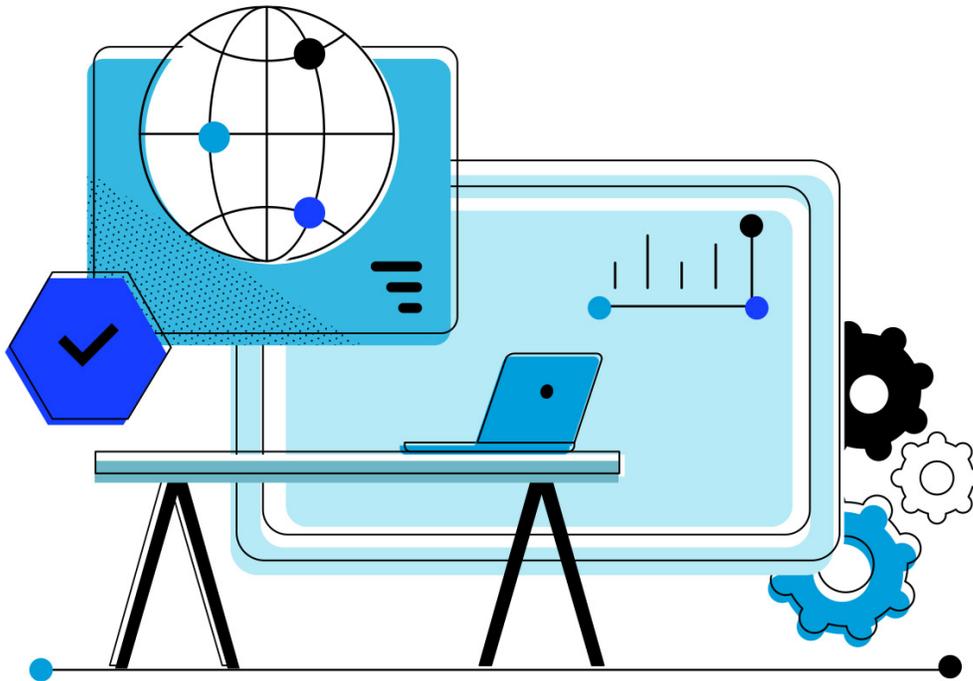
d. データを請求する。どこに広告が表示されたかを示す詳細なデータを請求し、出稿前に適合性審査を行うことで、広告出稿に隣接して表示される他のコンテンツの全面的かつ詳細な全体像を継続的に把握する。広告キャンペーンの徹底的な監査を行う。



e. 透明性を義務化する。アドテック企業に対し、アドテック・サプライチェーンの全面的検証を可能にする透明性基準を採用するよう要求し、ログレベルでの出稿とブロッキングのデータ（広告表示をブロックしたデータ）を含め、クライアントや研究者とすべての広告キャンペーンのデータを共有する。



f. 監査を行う。アドテック企業に、アドエクスチェンジ（複数の広告ネットワークやメディア等をまたいで広告枠を取引できる仕組み）供給パートナーに対する独立の第三者による監査と審査を行うよう要求する。



その他の民間主体

テクノロジー部門に直接関わらない其他幅広い民間主体の行動も、情報空間に影響を及ぼします。これは情報の誠実性を損なうこともあれば、これに寄与することもあります。企業には、表現の自由に対する権利や知る権利など

の人権を尊重する責任がある一方で、その他のステークホルダーと協調的パートナーシップを形成し、情報エコシステムの健全化に寄与することもできます。

提言



a. 誠実性を守る。 表現と言論の自由の権利を含め人権を擁護するとともに、財務的目標やその他の戦略的目標達成のために、情報エコシステムの誠実性に対するリスクを故意に拡散したり助長したりすることを控える。



b. リテラシーに投資する。 関連の市民社会の主体と連携し、その専門知識を借りることで、組織レベルで従業員のメディア・リテラシーと情報リテラシーに投資する。

報道機関

独立した自由で多角的なメディアは、公益に係る問題に関する情報を市民に届け、市民参加を促し、権力者の説明責任を促進するうえで欠かせない役割を果たします。

しかし、メディアの独立性と多様性に対する直接的および間接的な脅威や地域・公益ジャーナリズムの衰退は、こうしたメディアの不可欠な機能を損ねるおそれがあります。

職業上の基準を強固に維持しなければ、報道機関が情報の誠実性を侵害することにもなりかねません。倫理的な報道と編集の実践や、透明性へのコミットメントを質の高いトレーニングや勤務条件で補強すれば、ジャーナリストは不可欠のサービスを提供するだけでなく情報エコシステムの誠実性に対するリスクによって崩れたバランスを回復することにも貢献できます。

提言



a. 情報の誠実性について報道する。 情報エコシステムに対するリスクについて積極的に報道し、市民に情報を提供するためにデータに基づく調査報道の能力構築に投資する。メディアを消費する人々の信頼を維持、確保できるよう、情報の出所を含め、編集に厳密なプロセスと基準を用いる。市民の参考となるファクトチェック・メカニズムを確立する。



b. 危機対応を提供する。 情報エコシステムの誠実性に対するリスクが高まりかねない緊急事態や危機的状況で、市民への迅速な情報を無償で提供することを確約する。



c. 職業上の基準と倫理基準を維持する。 不偏性と編集の独立性を重視しつつ、公益に資するべく策定されたプロフェッショナルで倫理的なジャーナリズムに関してグローバルに認められた規範と実践を支持し、遵守する。また、自らのコンテンツを検証する自主的な説明責任のメカニズムを積極的に採用する。倫理的で正確、かつ不偏的な報道を前進させるとともに、「ソリューション」や「建設的」ジャーナリズムの手法を採用することを含め、イノベーションを促し、コミュニケーション環境の変化への適応力を高めるためのスキルをアップデートするために、質の高いトレーニングを定期的実施する。ユーザーが自ら選択、消費するニュースについてより良い情報を得られるよう、資金源や所有構造、金銭的インセンティブを開示する。



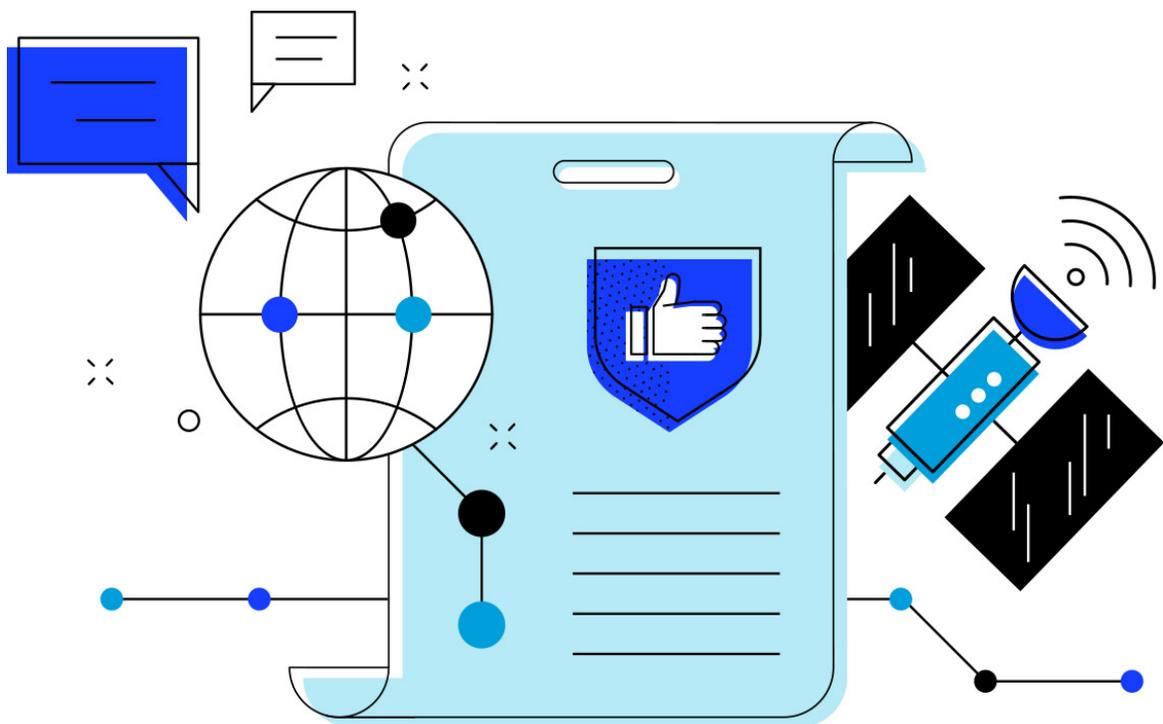
d. AI を倫理的に使う。 掲載や放送の際、AI が生成または編集した素材に明確なラベルをつけるなど、AI テクノロジーの倫理的利用について強固な方針を確立する。その中に、コンテンツが AI によって生成または編集されたものであることを特定できるメタデータの形で情報を含める。



e. 透明かつ人権面で責任ある広告を確立する。 広告が情報空間へのリスクを高めないようにするためのステップを踏む。ニュースと意見、スポンサーが付いているコンテンツを明確に区分し、意見記事の財源と潜在的な利益相反に関する透明性を確保する。広告と AI が生成または編集した広告や記事体広告に明確にラベルをつける。広告収入源、および、明確でアクセス可能な広告方針と実践について透明な報告を行う。



f. 労働基準を尊重する。 国際労働基準、国際人権法と整合する勤務条件を確立するとともに、デジタル空間を含め、ジャーナリストの福祉と安全の確保に寄与する取り組みを優先し、女性のジャーナリストやメディア関係者に対する差別、虐待、嫌がらせ、暴力の脅威に特に注意を払う。



研究者と市民社会

研究者と市民社会組織は、情報エコシステムの誠実性に対するリスクの多面的な影響を理解し、これに対処するための取り組みに欠かせない役割を果たします。研究者や市民社会による取り組みは、特に社会的に疎外され、脆弱な立場に置かれた集団にとっての、情報空間に対するリスクを

明らかにし、アドボカシーのためのエビデンス基盤を整備し、レジリエンスを高めることに役立つ可能性があります。研究者の知見と効果的な解決策との間の隔たりを埋めるためには、協調的なパートナーシップと知識の交換が不可欠です。

提言



a. 協働する。 地域や文脈の異なるステークホルダーと連携し、情報エコシステムの誠実性を強化するための効果的で倫理的なアプローチを共有する。



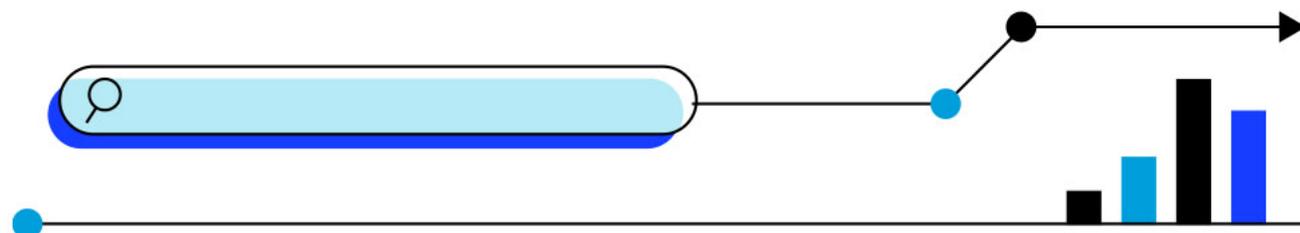
b. 誠実性と倫理基準を守る。 人権を擁護し、情報の誠実性に対するリスクを故意に拡散または助長するのを控える。あらゆる研究は倫理的かつ透明に、そしてプライバシーに配慮して行う。



c. オープンアクセスを促す。 研究結果を自由に入手できるようにし、分野間の協働を促進するなどして、オープンアクセスを確保するよう取り組む。



d. 包摂的な研究を強化する。 研究が進んでいない脆弱な立場に置かれ、社会から疎外された状況やコミュニティに特に配慮しながら、情報の誠実性に対するリスクが SDGs に及ぼしかねない影響を含め、地域や言語、分野を横断した情報エコシステムに関する学際的研究を模索する。このようなリスクと関連の危害を測定するための厳密な手法を開発する。

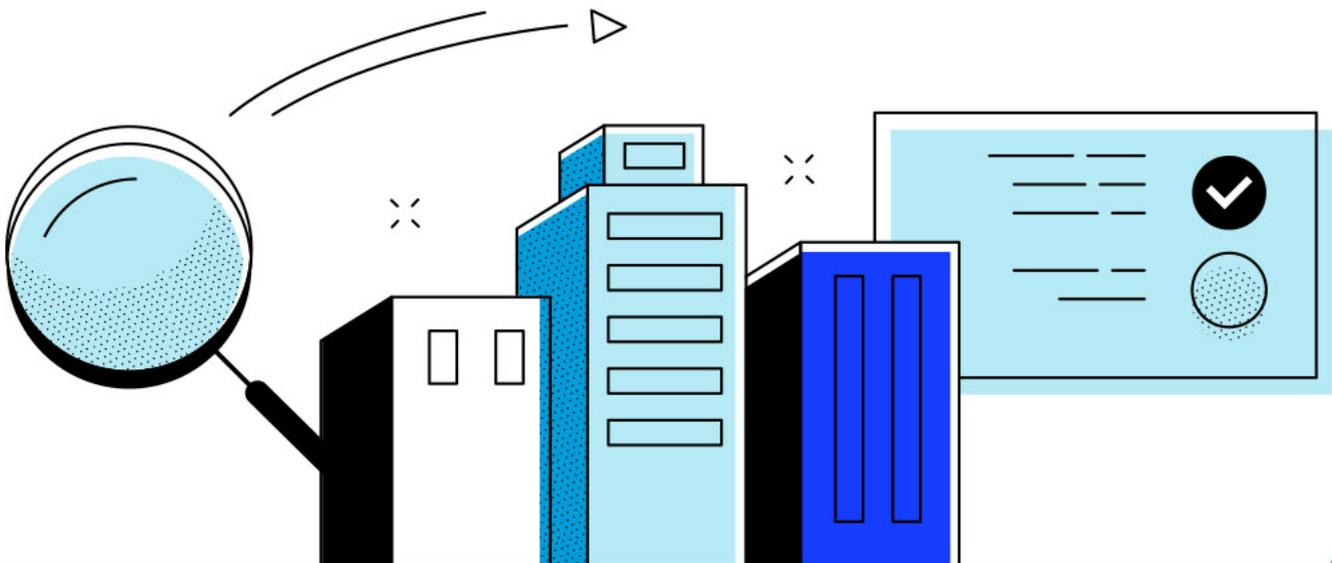




ファクトチェック機関と ネットワーク向けの提言

a. 職業上の基準を維持する。 専門性と倫理への確約を徹底するとともに、組織の構成やガバナンス、財源、所有者、および業務慣例に明確に表れる形で、独立性、無党派性、透明性の基準を遵守する。

b. 資金調達に関する情報を開示する。 財源のほか、テクノロジー企業、メディア、市民社会組織などのステークホルダーとの協働に関して情報開示する。



各国政府

各国政府には、「情報の誠実性のための国連グローバル原則」を強化するうえで欠かせない責任があります。これはまず、人権、特に情報を求め、受け取り、伝える権利を含む表現の自由に対する権利を尊重し、擁護し、推進するという国家の義務から始まります。

各国政府は、その法規制上の権限、公共資源への支配力、国内的にも国際的にも提携を構築する能力をはじめとする要因により、情報空間を形づくるうえで中心的な役割を担います。各国はその人権義務の一環として、効果的な政策、立法、規制および判決を通じ、人権侵害に関する調査、処罰、救済の適切なステップを踏むことで、国内や管轄内で企業による侵害から人権を守らなければなりません。

情報エコシステムに取り組める技術的、財政的能力は国によって異なります。インフラの整備の度合いや技術と資金

に対するアクセスにおける隔たりは、デジタル格差を助長しています。その一方で、多くのテクノロジー企業はグローバル市場のほぼ全体に浸透して支配しつつあるものの、本拠地をグローバル・ノースの少数の国々に置いています。

すべての国が情報エコシステムに寄与し、その利益を得られるようにするには、各国が人権を尊重しながら、デジタル接続を拡大できる能力を高め、「AI 格差」が生まれる可能性を予め排除し、情報空間におけるリスクに適切に対処できる能力を強化できるよう、緊急かつ持続的に取り組むことが必要です。こうした取り組みは最終的に情報の誠実性を強化し、人権を推進するとともに、SDGs の達成に寄与するでしょう。

提言



a. 人権を尊重、擁護、推進する。 国際的な人権の基準と規範に基づき、表現と言論の自由の権利をはじめとする人権を尊重、擁護、推進する。市民社会の全面的な参加を得ながら、情報の誠実性の様々な要素に取り組むための規制やその他の措置を、国際人権法を含め、該当する国際法と整合させる。そうすることで、それらの規制や措置を人権の強化と信頼の構築に向けたさらに幅広い取り組みの一環として位置付ける。表現の自由に対する権利の制限を例外的なものとし、制限を設ける場合には、国際人権法に基づく要件に適合させる。つまり、他者の権利または評判、または、国家の安全保障や公共の秩序、公衆衛生や道徳の保護に必要な範囲にするとともに、比例原則を遵守する。権利の制限が事実上、表現の自由を抑圧することがないようにする。国際人権法を含め、国際法と整合する形で個人データのプライバシー保護を採用し、これを実効的に執行する。



b. 誠実性を守る。 国内的にも国際的にも、故意に偽情報を拡散したりヘイトスピーチを伝達したりする情報活動を実施、または助長することを控える。いかなる形でも、インターネットを遮断したり速度を低下させることを控える。国連の平和オペレーションをその権限と活動に影響を与えない情報エコシステムの誠実性に対するリスクから保護することに関する決議を含め、関係する国連安全保障理事会決議を守り、実施する。



c. 人々を守る。 子どもの特殊なニーズと権利に取り組みつつ、女性や LGBTIQ+ の人々、民族的または宗教的少数集団など、オンラインでもオフラインでも情報空間で標的とされることが多い脆弱な立場に置かれ、社会から疎外された状況にある集団の、法律面および実践面での保護とエンパワメントを確保する取り組みを再確認し、さらに進める。国際人権法上の義務を遵守して、戦争のためのプロパガンダや差別、敵意または暴力の扇動に相当する民族的、人種的または宗教的憎悪の提唱を法律により禁止する。



d. 情報へのアクセスを提供する。 十分なサービスを受けていないコミュニティのためのアクセスを促進しながら、報道機関を含め、すべての人々が理解、利用できるあらゆる言語とフォーマットで、差別なく、公的に有している情報への迅速なアクセスを提供する。危機的状況下において、信頼できる正確な情報へのアクセスを確保する。コミュニティを積極的に巻き込み、公的機関への信頼を得るために倫理的で信頼できる情報発信の方法を採用する。



e. 報道の自由を確保する。 ジャーナリストやメディア関係者、ファクトチェッカーをあらゆる形態の差別や虐待、嫌がらせ、暴力の脅威から守るために力強い措置を取ることで、自由で存続可能な独立の多元的メディア環境を確保、保護、推進する。その際に、女性や脆弱な立場に置かれ、社会から疎外された状況にある集団の人々に特に注意を払う。法律と政策により、デジタルコンテンツ・クリエイターと市民ジャーナリストの権利を尊重、保護する。



f. 研究者と市民社会を守る。 学問の自由を尊重し、威嚇や嫌がらせ、報復行為から研究者と市民社会を守る。



g. 透明性を提供する。 テクノロジー企業と報道機関に対して課された要件とデータに関する要請について、完全な透明性を提供する。倫理的でない採用慣行や金銭的インセンティブなど、情報の誠実性を損なう不透明で欺瞞的なロビー活動やテクノロジー企業と政策立案者間の利益相反に対応する措置を取る。



h. グローバルな連帯と能力構築、開発支援を強化する。特に開発途上国において情報の誠実性を強化し、情報空間へのリスクに対するレジリエンス（強靱性）を高めるための能力構築を支援する国家間の協働とパートナーシップに参加する。あらゆる言語で、AIテクノロジーに関するものを含むデジタル・リテラシー、情報リテラシー、メディア・リテラシーのトレーニングと能力構築、啓発プログラムに資金を配分する。その際に、全面的な透明性を確保する。開発途上国が自国で主導して、情報エコシステムの誠実性へのリスクに対する社会的レジリエンスを構築し、メディア・リテラシーと情報リテラシーに関する十分なトレーニングを実施し、公益メディアを補強するために実施する取り組みに、こうした目的に特化した適切な開発援助の提供を含め、支援する。図書館を含む公的機関によるリテラシー関連のトレーニングと資源へのアクセス改善に向けた活動を支援する。



i. 政治参加を促進する。選挙に関わるあらゆるステークホルダーが選挙プロセス全体を通じて正確かつ迅速な情報にアクセスできるよう、そのアクセスを保護する。包摂的な政治参加とリーダーシップを促進し、あらゆる形態の差別、虐待、嫌がらせ、暴力の脅威からの保護を含め、公的領域における女性の権利を擁護するための措置を取る。



j. 包摂的な公益に資する研究を優先する。AIテクノロジーの新たな、また、未知の能力と影響などに照らし、情報の誠実性に係る専門領域全体で倫理的基準と審査に従う独立の研究を優先し、投資し、支援する。情報エコシステムの誠実性に対するリスクがSDGsに及ぼしかねない影響を含め、地域や言語、分野を横断して実施される研究を支援する。その際に、特にサービスや研究の対象となっておらず、リスクを抱えた状況やコミュニティに重点を置く。国内と国家間の公正な情報共有を可能にするために研究結果へのオープンアクセスを促進し、周知する。



k. リテラシーを育成する。デジタルスキルを低学年から、フォーマル、インフォーマル双方の教育カリキュラムに統合しながら、焦点を絞ったメディア・リテラシーと情報リテラシーの活動の推進を通じて、批判的に情報に基づいて公の場で発言することを推進する。背景と年齢が異なるあらゆる人々の具体的な社会・文化・言語面のニーズを考慮しつつ、オンラインの権利、デジタル情報環境の仕組み、個人データの用途に関して、子どもを含む市民の理解と認識を積極的に高める。女性、子ども、若者、高齢者、障害者、そして新たにオンラインに加わろうとしてる数十億の人々を含め、脆弱な立場に置かれ、社会から疎外された状況にある人々と集団のリテラシーのニーズを優先する。AIテクノロジー関連の具体的問題についてリテラシー向上に取り組むとともに、新たな、そして生じつつあるテクノロジーや課題を反映し、リテラシー向上への取り組みを継続的にアップデートする。



l. 子どもと親、保護者、教育者のエンパワーメントを図る。安全で責任あるデジタル空間における行動、オンライン・メディアの活用、子どもが持つ表現の自由と知る権利に対する理解に関し、子どもと親、保護者、教育者に継続的な資源を提供する。より安全なオンライン体験を目指すメディア・リテラシーと情報リテラシー関連のガイドラインとイニシアチブを策定する際には、若者のデジタルを使いこなす能力を活用しつつ、あらゆる当事者が関与するようにする。



あらゆる政治的主体向けの提言

政治的プロセスに関与し、
これに影響を与える個人、集団および組織

a. 選挙の誠実性を維持する。有権者資格、投票、開票および結果に関するものを含め、情報の誠実性を損なおうとする試みを控えるとともに、そうした試みを公然と非難する。

b. 包摂性を守る。女性や脆弱な立場に置かれ、社会から疎外された状況にある集団の人々をはじめとする候補者や公人を標的とした誹謗や嫌がらせを公然と非難し、これに対処するための措置を取る。

c. 透明性を提供する。広告の財源や、データに基づくターゲティング手法の使用を含め、情報発信の透明性を維持する。

国連

「情報の信頼性のための国連グローバル原則」は、国連とその国際公務員にも適用されます。国連はグローバル原則を遵守することにより、国際社会における情報の誠実性を責任ある形で管理する模範を示し、これに説得力を持たせます。

情報エコシステムの誠実性を強化する活動を拡大すれば、平和を確保し、持続可能な開発を促し、すべての人の人権を擁護、保護するという国連の使命の遂行にも寄与することでしょう。

国連の約束



a. 取り組みを拡大する。 多様な状況と言語を横断して、文脈に即した研究、モニタリング、リスク評価、コミュニティとの対話、提携の構築をはじめとする方策を通じて、情報の誠実性強化に向けた取り組みを強化する。予防と軽減、対応を強化し、生じつつある機会と課題を明らかにするために、国連のプログラムや活動に情報の誠実性を統合する。



b. 能力構築に向けたイニシアチブを支援する。 開発途上国のニーズに特に注意を払いながら、若者向けのトレーニングを含め、情報の誠実性強化に役立つスキル開発のイニシアチブを提供することにより、各国での能力構築を支援する。



c. アドボカシー活動を展開する。 十分なサービスが受けられない状況や、脆弱な立場に置かれ、社会から疎外された状況にある集団に特に注意を払いながら、グローバルなレベルと各国、各コミュニティで、グローバル原則を推進し広める。SDGsの実現に向けた取り組みを支援しながら、社会の結束に積極的に貢献し、情報の誠実性へのリスクに対するコミュニティのレジリエンスを強化する。



d. 専門能力を増強する。 国連のマנדート遂行と実質的な優先事項に影響する情報エコシステムの誠実性へのリスクに対して、革新的で文脈に即したアプローチを開発する中核的な部署を国連事務局に設ける。その部署は、必要に応じて他の部署と調整し、国連システム全体に貢献する。



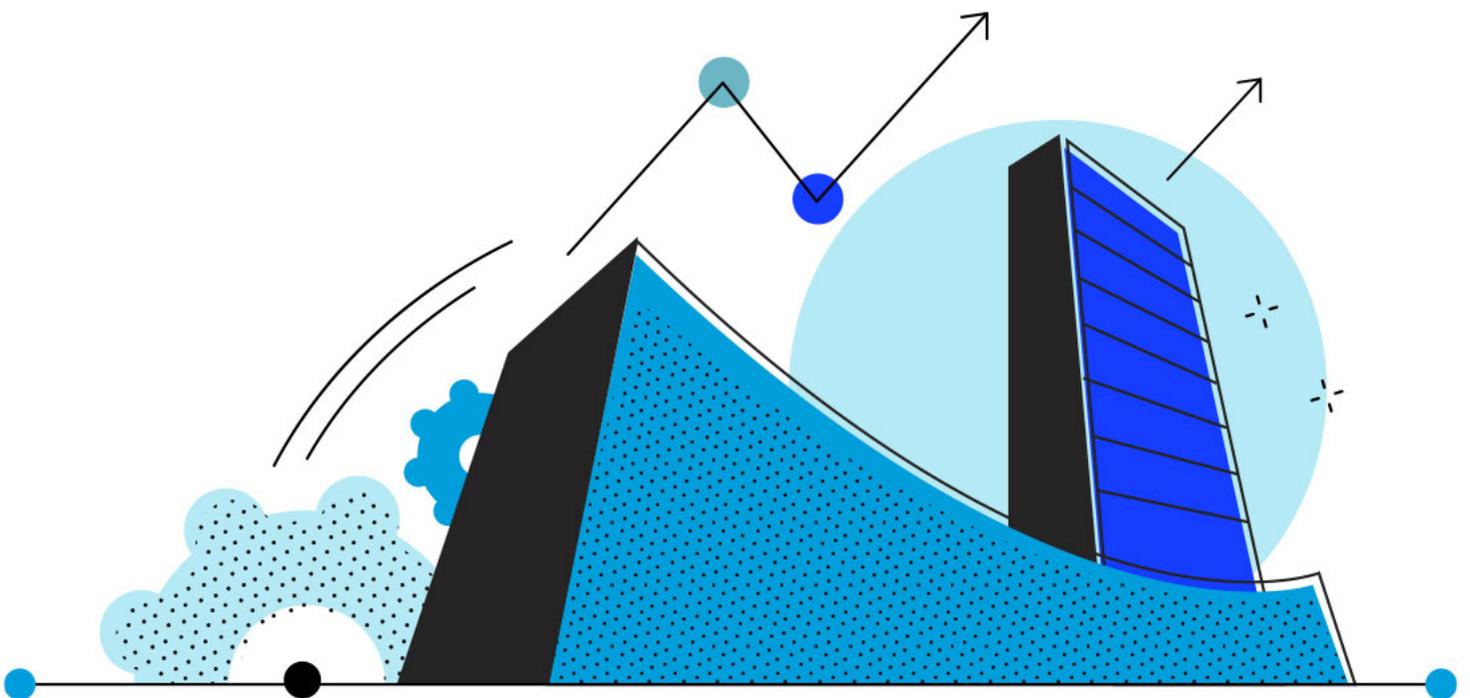
e. 機動的な広報戦略を策定する。 共通の利益のために、また、国連が奉仕するすべての人々のニーズをより良く充足するために、デジタル空間とオフライン空間を活用しながら、エビデンスに基づき、革新的で機動的な広報戦略を個々の目的に合わせて策定する。



f. 多言語の資料を提供する。 グローバル、地域、国内レベルでの取り組みを支援するために、多様な状況に適用できる研究やガイダンス、ベストプラクティスを提供する、情報の誠実性に関する多言語のオンライン資料のハブを設ける。



g. マルチステークホルダー型の行動計画を支援する。 既存のメカニズムを活用し、国際的な能力構築と調整に関する国連の専門性と経験を動員しながら、各地域と国内でマルチステークホルダー型の行動計画と提携を支援する。



今後の道のり

情報エコシステムの誠実性に対するリスクが高まり、簡単に利用できる最新の AI テクノロジーが進歩していくことを考えれば、情報の誠実性を強化する緊急性はいくら強調しても足りません。本グローバル原則は、世界が複雑なデ

ジタル時代を突き進み、「未来サミット」で多国間の解決策を見出そうとする中で、情報の誠実性を保護、促進するための行動に向けた包括的で統一的な枠組みを提供しています。

この目的を達成するために、ステークホルダーには以下の行動が強く求められています。

- ✓ すぐに行えることの枠組みとして「情報の誠実性のための国連グローバル原則」の遵守を公約するとともに、これを採用し、積極的に公表する。
- ✓ グローバル原則を活用し、能力構築を目的とするものを含め、市民社会や学界、メディア、政府、国際的な民間セクターから多様な専門性とアプローチを結集し、**情報の誠実性に関する幅広い部門横断的な提携**を結成し、これに積極的に参加する。その際は、専門のユース諮問グループなどを通じ、若者の全面的かつ意味ある参加を確保する。
- ✓ 草の根の取り組みを支援して学びを得るためにコミュニティーを巻き込むとともに、若者の全面的かつ意味ある参加を確保しながら、地域、国内、地方のレベルで協働による**マルチステークホルダー型の行動計画**を策定する。

あらゆる部門のステークホルダーは「情報の誠実性に関する国連グローバル原則」を受け入れることにより、連帯を実証し、信頼や知識、あらゆる個人の選択権を育む、情報エコシステムの再活性化への道のりを協働して築くことができます。

付録

参考資料

1. United Nations Secretary-General’s “Our Common Agenda policy brief 8: information integrity on digital platforms” (2023)
<https://www.un.org/sites/un2.un.org/files/our-common-agenda-policy-brief-information-integrity-en.pdf>
2. UNESCO Guidelines for the Governance of Digital Platforms (2023)
<https://www.unesco.org/en/internet-trust/guidelines>
3. Report of the Secretary-General, “Countering disinformation for the promotion and protection of human rights and fundamental freedoms”, 2022 (A/77/287)
<https://www.ohchr.org/sites/default/files/2022-03/NV-disinformation.pdf>
4. UNESCO Recommendation on the Ethics of Artificial Intelligence (2021)
<https://unesdoc.unesco.org/ark:/48223/pf0000381137>
5. United Nations Strategy and Plan of Action on Hate Speech (2019)
www.un.org/en/genocideprevention/documents/advising-and-mobilizing/Action_plan_on_hate_speech_EN.pdf
6. Rabat Plan of Action on the prohibition of advocacy of national, racial or religious hatred that constitutes incitement to discrimination, hostility or violence (2012)
<https://www.ohchr.org/en/documents/outcome-documents/rabat-plan-action>
7. Guiding Principles on Business and Human Rights (2011)
https://www.ohchr.org/sites/default/files/documents/publications/guidingprinciplesbusinesssh_en.pdf