# 生成AIで加速するデジタル空間の偽・誤情報に対する政策の現状と方向性

吉田　弘毅（YOSHIDA, Koki）
総務省　情報流通行政局
情報流通振興課　企画官

# 1. デジタル空間における情報流通を巡るリスク・問題

# インターネット上での偽・誤情報等の流通の顕在化



- 多くの利用者が情報を収集・閲覧するプラットフォームサービス等の**インターネット上では、偽・誤情報等も流通するなどの問題が顕在化**。
- 一因として、**偽・誤情報等は、SNS上において正しい情報よりも速く、より広く拡散する特性**があること等が指摘。

## ◆ メディアごとの偽・誤情報を見かける頻度

- インターネット上のメディアやまとめサイトにおいては、40%強が週に１回以上接触。
- インターネット以外のメディアにおいては、30%強が週に１回以上接触。

問）直近の１ヶ月の間で、あなた自身が偽情報・誤情報※だと思う情報を次に示すメディアの中でどの程度見かけますか。
※ここでは、虚偽、または、誤解を招くと考えられる情報/ニュースを指します。

【出典】令和４年度国内外における偽情報に関する意識調査（総務省）

## ◆ 事実と誤情報の拡散ダイナミズム

- 誤情報は事実よりも遠く、速く、深く、広範囲に拡散。
- Twitter上で事実が1000人以上にツイートされるのは稀だが、誤情報はもっと多くの人にツイートされた。
- 最初の投稿がリツイートされるまでの時間は誤情報の方が事実よりも20倍短く、1500人に届くまでにかかる時間も６倍短い。

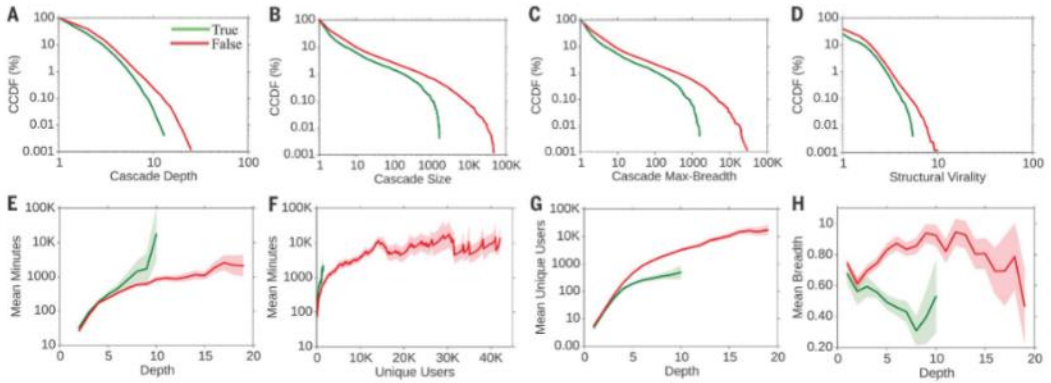



**Fig. 2. Complementary cumulative distribution functions (CCDFs) of true and false rumor cascades.** (**A**) Depth. (**B**) Size. (**C**) Maximum breadth. (**D**) Structural virality. (**E** and **F**) The number of minutes it takes for true and false rumor cascades to reach any (E) depth and (F) number of unique Twitter users. (**G**) The number of unique Twitter users reached at every depth and (**H**) the mean breadth of true and false rumor cascades at every depth. In (H), plot is lognormal. Standard errors were clustered at the rumor level (i.e., cascades belonging to the same rumor were clustered together; see supplementary materials for additional details).

【出典】Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. Science, 359(6380), 1146-1151.

# インターネットにおける偽・誤情報の事例



| 年 | エリア | 概　要 |
|---|---|---|
| 2011 | 日本 | 東日本大震災直後、「今回の地震は人工地震である」、「被災地で外国人による犯罪が横行している」等の偽・誤情報（流言飛語）が拡散。 |
| 2016 | 日本 | Twitterに、ライオンが路上に立つ画像を添付したデマが投稿。瞬く間に拡散しリツイートは１時間で２万件超。熊本市動植物園の業務を妨害したとして偽計業務妨害の疑いで逮捕。災害時にデマを流し業務妨害での逮捕は全国初。不起訴。 |
| 2016 | 英国 | EUからの離脱を問う国民投票に際し、「英国はEUに毎週３億5000万ポンドを支払っている」という偽・誤情報が拡散。 |
| 2016 | 米国 | 大統領選挙に際し、各候補者について「ローマ法王が支持を表明した」、「あるピザ店を拠点とする児童売春組織に関与している」等の偽・誤情報が拡散。後者の拡散は、噂されたピザ店に実際に銃を持って押し入り、発砲する人物まで現れる事態につながった（いわゆる「ピザゲート事件」）。 |
| 2016 | ドイツ | 移民によるテロ事件の発生後、事件と無関係の難民とメルケル首相（当時）との写真が利用され、同首相がテロリストと関係があったかのような偽・誤情報が拡散。 |
| 2017 | フランス | 大統領選挙に際し、「マクロン候補（当時）が租税回避地にペーパーカンパニーや銀行口座を保有している」という偽・誤情報が拡散。 |
| 2019 | 日本 | 常磐自動車道であおり運転や傷害行為に及んだ人物の同乗者であるとし、無関係の人物が特定され、誹謗中傷を受けた。 |
| 2020 | 世界 | 新型コロナウイルス感染症の流行に際し、「第５世代移動通信システム（5G）が流行に寄与している」、「ワクチンにマイクロチップが含まれている」等の偽・誤情報や、医学的な根拠のない感染予防効果を標榜する健康商品等のインターネット広告が拡散。 |
| 2022 | 世界 | ロシアによるウクライナへの軍事侵攻を巡り、特定の国・個人・企業がいずれかの陣営を支持又は非難しているかのように見せかけた偽画像・偽情報等が拡散。 |
| 2022 | フィリピン | 大統領選挙に際し、候補者に不利な事実の存在を否定するなどの内容の偽・誤情報が拡散。 |
| 2023 | 日本 | X（旧Twitter）に、財務官の氏名や肩書、写真等を使用して本人のアカウントと偽ったなりすましアカウントが出現、550人ほどがフォロー（後にアカウント凍結）。 |
| 2023 | 日本 | 東京電力福島第一原子力発電所からのALPS処理水の海洋放出に際し、無関係の画像等を使用して「魚が大量死している」、「奇形の魚が発見された」等とする偽・誤情報が拡散。 |
| 2023 | 世界 | イスラエルとパレスチナ武装勢力ハマスとの衝突を巡り、無関係の画像等を使用して現地の状況を実際よりも軽く見せたり、逆に深刻に見せたりする偽・誤情報が拡散。 |

【出典】各種ウェブサイトを基に作成

# 生成AIやディープフェイク技術を利用した巧妙な偽・誤情報の事例



| 年 | エリア | 概　要 |
|---|---|---|
| 2019 | マレーシア | 閣僚が同性と性行為（マレーシアでは禁止）をしている動画が偽造され、拡散。 |
| 2021 | 日本 | 東北地方で発生した地震直後に、官房長官が記者会見で笑みを浮かべる偽画像が拡散。 |
| 2021 | 欧州 | ロシアの議員のディープフェイク動画と気付かずに欧州の議員がビデオ電話会議を実施。 |
| 2022 | 欧州 | ロシアによるウクライナへの軍事侵攻に際し、ゼレンスキー大統領が市民らに投降を呼びかける内容の偽動画が拡散。 |
| 2022 | 日本 | 台風の発生時に、静岡県で住宅が水没したとする偽画像が作成され、SNSに投稿・拡散。 |
| 2023 | 米国 | 政治活動家が、バイデン大統領が第三次世界大戦の開始を告げる動画を作成。<br>作成者はAIで作成した旨を説明したが、多くの人が説明をつけないまま動画を拡散。 |
| 2023 | 米国 | 国防総省の近くで爆発が起きたかのように見せかけた偽画像が拡散。株価が一時的に下落する事態に発展。 |
| 2023 | 日本 | 生成AIを使って民放の実在のアナウンサーの声や動きを再現し、投資情報サイトへの登録を促す内容の偽動画がSNS上で拡散。 |
| 2023 | 日本 | 総理大臣の声と画像を使って卑猥な言葉で語りかけているように見せかけ、実在する民放の報道番組で放送されているかのように編集された偽動画がSNS上で拡散。 |

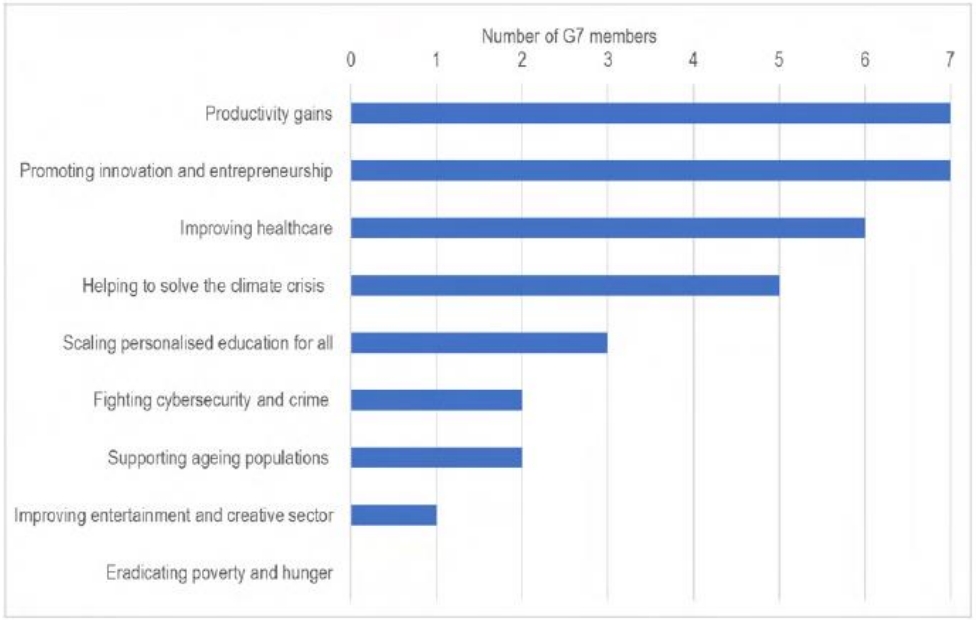
【出典】令和５年版情報通信白書（総務省）ほか各種ウェブサイトを基に作成

# 「生成AI」技術の急速な進展とリスク



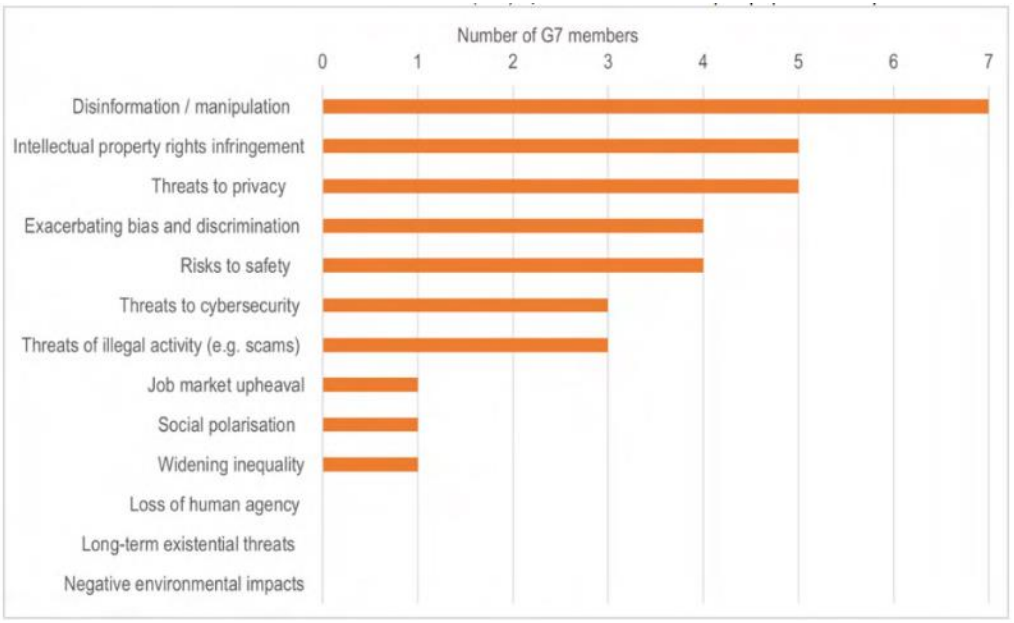
- **大規模言語モデル（LLM:Large Language Models）**の開発が進み、従来人間が得意としてきた、情報を生成・創造する目的で用いられる**生成AI（Generative AI）**技術が急速に進展し、生産性向上等が期待。
- 他方、偽情報・情報操作、知的財産権侵害、プライバシー侵害、偏見・差別の助長、安全上のリスク等のリスクをもたらすとの指摘。特に、**偽情報・情報操作については、G7構成国全てがリスクとして認識**。

## ◆ G7構成国が選択した生成AIの活用の機会
(あらかじめ与えられた選択肢の中から5つを選択)




Note: The figure aggregates responses from seven respondents to the question: “From your country or region’s perspective, what are the top five opportunities generative AI presents to help achieve national and regional goals? (Please select five options)”.

## ◆ G7構成国が選択した生成AIに関するリスク
(あらかじめ与えられた選択肢の中から5つを選択)



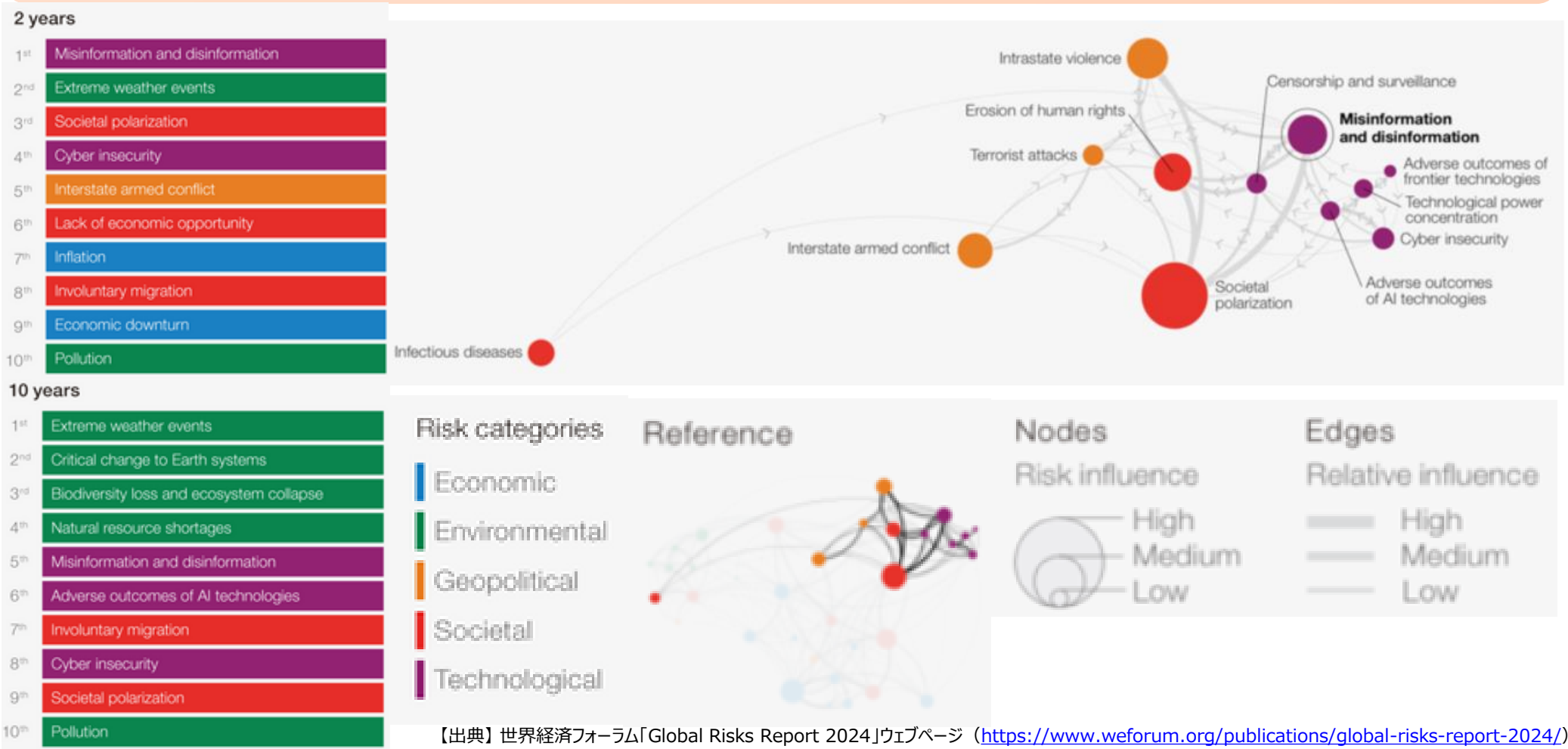

Note: The figure aggregates responses from seven respondents to the question: “From your country or region’s perspective, what are the top five risks generative AI presents to achieving national and regional goals? (Please select five options)”.

【出典】G7 Hiroshima Process on Generative Artificial Intelligence (AI), Report prepared for the 2023 Japanese G7 Presidency and the G7 Digital and Tech Working Group, OECD (2023)

# グローバルリスク報告書2024（世界経済フォーラム）



- 世界経済フォーラム（WEF：World Economic Forum）「グローバルリスク報告書2024」（2024年1月）において、**<u>今後2年間で予想される最も深刻なリスクとして、AIによる偽・誤情報を警告</u>**。
- 今後2年間にいくつかの主要経済国（例えば、バングラデシュ、インド、インドネシア、メキシコ、パキスタン、イギリス、アメリカ）で選挙が実施される中、**<u>改ざんされた情報と社会不安の結びつきがリスクの中心</u>**。
- 偽・誤情報のリスクは、選挙のみならず、**<u>現実の認識もより多極化し、公衆衛生から社会正義までに至る公共問題に関する議論にも浸透</u>**。また、今後10年間では、5番目に位置付け。




【出典】世界経済フォーラム「Global Risks Report 2024」ウェブページ（https://www.weforum.org/publications/global-risks-report-2024/）

# 2. 総務省における政策の検討

# 「デジタル空間における情報流通の健全性確保の在り方に関する検討会」



## 1．検討事項

① **デジタル空間を活用したサービスの普及・情報通信技術の進展等の状況**：（例）生成AI、メタバースなど

② **新たな課題と各ステークホルダーによる対応状況**：
（例）生成AI等による巧妙な偽・誤情報の生成や拡散に伴う社会的な影響の深刻化、メタバースにおけるデータの取得・利用に係る対応の重要性、国境を越えた情報・データ流通の広域化や迅速化に伴う国際的な協調の必要性など

③ **今後の対応にあたっての基本的な考え方**：
（例）基本理念：信頼性のある自由な情報流通、表現の自由、知る権利、青少年を含む利用者保護、デジタルシティズンシップなど
各ステークホルダーの役割：デジタルプラットフォーム事業者、生成AI事業者、仮想空間関係事業者、通信・放送事業者、利用者など

④ **デジタル空間における情報流通の健全性確保に向けた具体的な方策**：
（例）多様なステークホルダーによる協力関係の構築、ファクトチェックの推進、幅広い世代に対するリテラシーの向上、情報発信者側を含む自主的取組の推進、研究開発の推進、国際的な対話の深化、生成AI・メタバース関連事項など

## 2．構成員

※法律（憲法・メディア・情報）、技術、データ分析、認知科学・心理学、ジャーナリズム、プライバシー、サイバーセキュリティ、消費者保護等の専門家。以下、座長・座長代理以外は、50音順。

- 宍戸 常寿（東京大学大学院法学政治学研究科教授）【座長】
- 山本 龍彦（慶應義塾大学大学院法務研究科教授）【座長代理】
- 生貝 直人（一橋大学大学院法学研究科教授）
- 石井 夏生利（中央大学国際情報学部教授）
- 越前 功（国立情報学研究所情報社会相関研究系教授）
- 江間 有沙（東京大学国際高等研究所東京カレッジ准教授）
- 奥村 信幸（武蔵大学社会学部教授）
- 落合 孝文（渥美坂井法律事務所・外国法共同事業 プロトタイプ政策研究所所長・シニアパートナー弁護士）
- クロサカ タツヤ（慶應義塾大学大学院政策・メディア研究科特任准教授）
- 後藤 厚宏（情報セキュリティ大学院大学学長）
- 澁谷 遊野（東京大学大学院情報学環准教授）
- 曽我部 真裕（京都大学大学院法学研究科教授）
- 田中 優子（名古屋工業大学大学院工学研究科基礎類教授）
- 増田 悦子（公益社団法人全国消費生活相談員協会理事長）
- 水谷 瑛嗣郎（関西大学社会学部メディア専攻准教授）
- 森 亮二（英知法律事務所弁護士）
- 安野 智子（中央大学文学部教授）
- 山口 真一（国際大学グローバル・コミュニケーション・センター准教授）
- 山本 健人（北九州市立大学法学部准教授）
- 脇浜 紀子（京都産業大学現代社会学部教授）

※ その他、関係団体や関係省庁がオブザーバとして参加

## 3．スケジュール

- 令和５年11月７日に第１回会合を開催し、ヒアリング・論点整理等を経て、本年９月にとりまとめを公表

# 「デジタル空間における情報流通の健全性確保の在り方に関する検討会」



- 令和５年11月７日に第１回会合を開催し、ヒアリング・論点整理等を経て、本年９月にとりまとめを公表。
- とりまとめでは、デジタル空間における情報流通の健全性確保のための具体的な方策として総合的な対策を公表。

## 基本的な考え方

- ◆ サイバーセキュリティやプライバシー等の関連分野を踏まえた社会全体で対応する枠組み
- ◆ 信頼性のある情報の流通促進と違法・有害情報の流通抑制の両輪による対応
- ◆ 個人レベルとシステムレベルの両面及び相互作用による対応
- ◆ プレバンキングとデバンキング※の両輪による対応
- ◆ 流通・拡散する情報とそれに付随するデジタル広告への信頼性に対する相互依存関係を踏まえた対応

※ プレバンキング：偽・誤情報等が流通・拡散する前の備え（リテラシー向上等）
デバンキング：偽・誤情報等が既に流通・拡散した状況での事後対応（ファクトチェック等）

## 総合的な対策

### 1 普及啓発・リテラシー向上

①プレバンキングの効果検証等有効な方法及び取組の推進
②普及啓発・リテラシー向上に関する施策の多様化
③マルチステークホルダーによる連携・協力の拡大・強化

### 2 人材の確保・育成

①検証報道等の信頼性のある情報を適時に発信する人材
②コンテンツモデレーション人材
③リテラシー向上のための教える人材

### 3 社会全体へのファクトチェックの普及

①ファクトチェックの普及推進
②ファクトチェック人材の確保・育成
③関連するステークホルダーによる取組の推進

### 4 技術の研究開発・実証

①偽・誤情報等対策技術
②生成AIコンテンツ判別技術
③デジタル広告関連技術

### 5 国際連携・協力

①普及啓発・リテラシー向上・人材育成の国際連携・協力
②偽・誤情報等対策技術の国際標準化・国際展開の推進
③欧米等とのバイやG7・OECD等とのマルチ連携・協力の推進

### 6 制度的な対応

①情報伝送PF事業者による偽・誤情報への対応
②情報伝送PFサービスが与える情報流通の健全性への影響の軽減
③マルチステークホルダーによる連携・協力の枠組みの整備
④広告の質の確保を通じた情報流通の健全性確保
⑤質の高いメディアへの広告配信に資する取組を通じた健全性確保

# 総合的な対策（制度的な対応）



## 1 情報伝送PF事業者による偽・誤情報への対応

**大規模な情報伝送PF事業者を対象とし、次の方策を中心に、具体化を進めることが適当。**

**①違法な偽・誤情報に対する対応の迅速化**

- 行政法規に抵触する違法な偽・誤情報に対し、行政機関からの申請を契機とした削除等の対応を迅速化（窓口整備、一定期間内の判断・通知　等）　ただし、前提として、行政機関による申請状況の透明性確保等が不可欠

**②違法な偽・誤情報の発信を繰り返す発信者への対応**

- 特に悪質な発信者に対する情報の削除やアカウントの停止・削除を確実に実施する方策について、その段階的な実施を含め具体化

**③違法ではないが有害な偽・誤情報に対する対応**

- 違法ではないが有害な偽・誤情報への対応は、影響評価・軽減措置の実施を求める枠組みの活用を含め、事業者による取組を促す観点が重要
- 情報の可視性に直接の影響がないコンテンツモデレーション（収益化停止等）を中心とした対応について、迅速化や確実な実施を含め、利用者の表現の自由の保護とのバランスを踏まえながら具体化

**④情報流通の態様に着目したコンテンツモデレーションの実施**

- 送信された情報の内容そのものの真偽に着目せず、情報流通の態様に着目してコンテンツモデレーションを実施する方策について具体化

**⑤コンテンツモデレーションに関する透明性の確保**

- 基準や手続の策定・公表、人員等の体制に関する情報の公表　等

## 2 情報伝送PFサービスが与える情報流通の健全性への影響の軽減

**情報伝送PF事業者による社会的影響等の予測・軽減措置の実施**

- 情報伝送PF事業者のビジネスモデルがもたらす将来にわたる社会的影響等を事前に予測し、軽減措置を検討・実施

## 3 マルチステークホルダーによる連携・協力の枠組みの整備

**①連携・協力の目的**（行動規範の策定・推進、軽減措置の検証・評価　等）

**②協議会の設置・役割・権限　等**

## 4 広告の質の確保を通じた情報流通の健全性確保

**①広告事前審査の確実な実施と実効性向上**

- 審査基準の策定・公表、審査体制の整備・透明化、本人確認の実施等

**②事後的な広告掲載停止措置の透明性の確保**

- 基準や手続の策定・公表、人員等の体制に関する情報の公表　等

**③事後的な広告掲載停止措置の迅速化**

- 外部からの申請窓口の整備・公表、一定期間内の判断・通知　等

**④事後的な広告掲載停止措置の確実な実施**

## 5 質の高いメディアへの広告配信に資する取組を通じた健全性確保

**①広告主・代理店による取組促進**（経営陣向けガイドライン等の策定）

**②広告仲介PF事業者による取組促進**

# ご清聴ありがとうございました。