

総務省「デジタル空間における情報流通の健全性確保の在り方に関する検討会」

Withフェイク2.0時代における 偽・誤情報問題の未来と求められる対策

2023.11

山口真一 博士（経済学）

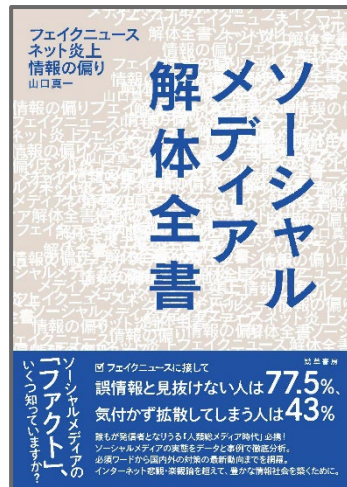
国際大学GLOCOM准教授

syamaguchi@glocom.ac.jp

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

- 東京大学 客員連携研究員
- シエンプレ株式会社 顧問
- 日本リスクコミュニケーション協会 理事
- 株式会社エコノミクスデザイン シニアエコノミスト
- 日本経済新聞Think!エキスパート
- グリー株式会社 アドバイザリーボード
- 株式会社メルカリ アドバイザリーボード
- Polimill株式会社 アドバイザー
- クリエイターエコノミー協会 アドバイザー
- 科学技術・学術政策研究所 (NISTEP) 専門調査員
- 早稲田大学ビジネススクール 兼任講師
- 中央大学国際情報学部 兼任講師
- 東洋英和女学院大学国際社会学部 兼任講師
- 内閣府「AI戦略会議」 構成員
- 総務省「デジタル空間における情報流通の健全性確保の在り方に関する検討会」 構成員
- 文部科学省「情報モラル教育推進事業の検討委員会」 構成員
- 厚生労働省「年金広報検討会」 構成員
- 厚生労働省「被用者保険の適用拡大に向けた委員会」 構成員
- 公正取引委員会「独占禁止政策協力委員」 委員

1986年生まれ。博士（経済学・慶應義塾大学）。2020年より現職。専門は計量経済学、ネットメディア論、情報経済論等。NHK等のメディアに多数出演・掲載。KDDI Foundation Award、組織学会高宮賞、情報通信学会論文賞（2回）、電気通信普及財団賞を受賞。主な著作に『ソーシャルメディア解体全書』（勁草書房）、『正義を振りかざす「極端な人」の正体』（光文社）等がある。



偽・誤情報問題の現在地

1



偽・誤情報問題の拡大

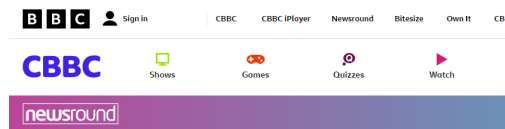
- 2016年は「偽・誤情報元年」といわれる。2016年の米国大統領選挙では多くの偽・誤情報が拡散され、選挙前3か月間で、トランプ氏に有利な偽・誤情報は3,000万回、クリントン氏に有利な偽・誤情報は760万回シェアされた。
- 2020年以降は新型コロナウイルスやコロナワクチン関連の偽・誤情報が広く拡散され、WHOが”Infodemic”と警鐘を鳴らした。
- 教育コンテンツなどの体を取った偽情報が青少年に及ぼす影響が懸念されている。
- 日本でも、災害時のデマ投稿、新型コロナウイルス関連、政治的なものなど、多くの偽・誤情報が拡散している。



「5G電波がコロナウイルスをまき散らしている」という偽・誤情報を信じて、基地局を破壊する活動まで行われた

<https://www.nikkei.com/article/DGXMZO58443970U0A420C2XR1000/>

<https://www.who.int/emergencies/diseases/novel-coronavirus-2019/advice-for-public/myth-busters>



AI used to target kids with disinformation

© 16 September - Comments

Investigative BBC journalists working in a team that analyses disinformation, information that is deliberately misleading and false, found more than 50 channels in more than 20 languages spreading disinformation disguised as STEM [Science Technology Engineering Maths] content.

虚偽の科学情報を含む動画を作成する YouTubeチャンネルが「教育コンテンツ」として子供に推奨されていることを指摘

<https://www.bbc.co.uk/newsround/66796495>

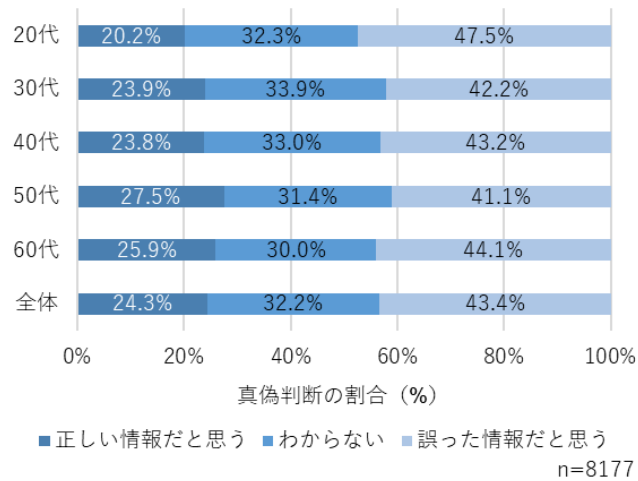


安倍元首相が被災地を訪れた際の写真が、スタジオで撮影されたかのように捏造された画像が拡散された。台湾でもFacebookやLINEで拡散され、台湾でもファクトチェックがされている

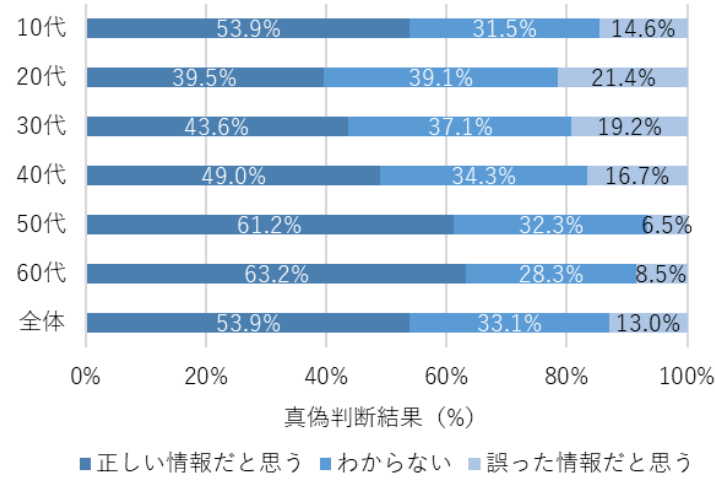
<https://infact.press/2020/03/post-5303/>

多くの人々が偽・誤情報を誤っていると気づけない

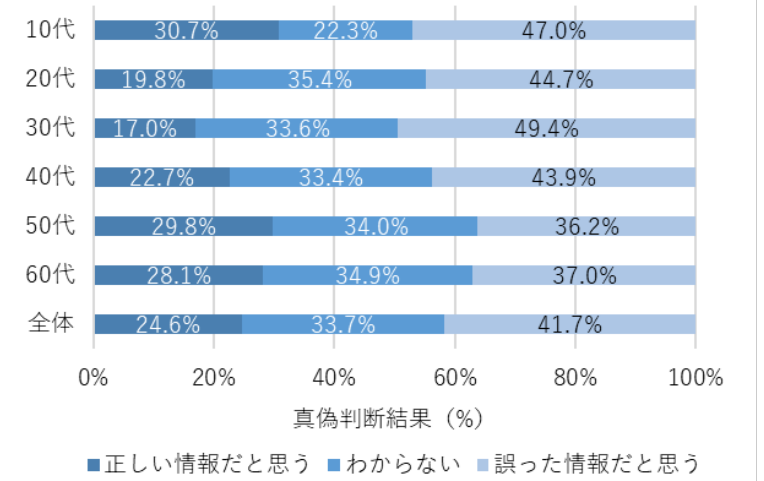
- コロナワクチン関連の偽・誤情報、政治関連の偽・誤情報、陰謀論を各6件、合計18件を使って調査した結果、コロナワクチン関連は37.1%、政治関連は26.4%、陰謀論は19.1%の人が、少なくとも1つ以上を見聞きしていた。
- 見聞きしたうえで誤っていると気づいている人は、政治関連では平均して13%しかいない。コロナワクチンと陰謀論では多いものの、それでも43.4%と41.7%にとどまっている。
- 年代別に判断結果を見ると、50代や60代といった中高年の方が、若い世代よりも誤っていると気づきにくい傾向が見られた（とりわけ政治関連の偽・誤情報と陰謀論において）。偽・誤情報は若者だけの問題ではないといえる。



偽・誤情報の真偽判断結果（コロナワクチン関連）



偽・誤情報の真偽判断結果（政治関連）

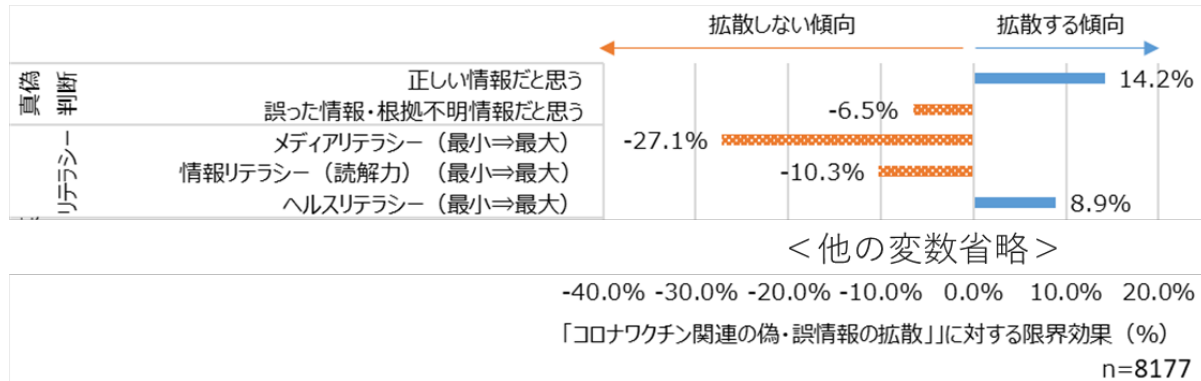


陰謀論の真偽判断結果

山口真一・谷原吏（2022）「Innovation Nippon 2021 わが国における偽・誤情報の実態の把握と社会的対処の検討 一政治・コロナワクチン等の偽・誤情報の実証分析一」、<https://www.glocom.ac.jp/activities/project/7759>
 山口真一・谷原吏・大島英隆（2023）「Innovation Nippon 2022 偽・誤情報、陰謀論の実態と求められる対策」、<https://www.glocom.ac.jp/activities/project/8839>

拡散するのは偽・誤情報を信じている人、リテラシーの低い人

- 偽・誤情報の拡散行動を分析したところ、偽・誤情報を信じている人は、誤っていると気づいている人に比べて非常に拡散しやすい傾向にあることが分かった。例えばコロナワクチン関連の偽・誤情報であれば、20.7ポイントも拡散確率が高い。
- また、メディアリテラシーや情報リテラシーが低い人ほど拡散することも明らかになった。例えば、メディアリテラシーが最も高い人と最も低い人で比較すると、コロナワクチン関連の偽・誤情報を拡散する確率が27.1ポイントも異なる。
- 偽・誤情報は、事実のニュースよりも約6倍も速く拡散することが明らかになっている。
- 人々が接している情報空間というのは、偽・誤情報を信じている人や、メディアリテラシー・情報リテラシーが低い人が拡散しやすい空間であるといえる。



$$\begin{aligned} \text{logit}[P(\text{Share}_{ijt} = 1)] &= \log\left(\frac{P[\text{Share}_{ijt}]}{1 - P[\text{Share}_{ijt}]}\right) \\ &= \alpha + \beta_1 \text{Determine}_{ijt} + \beta_2 \text{Literacy}_{ij} + \beta_3 \text{Media}_{ij} \\ &\quad + \beta_4 \text{People}_{ij} + \beta_5 \text{Characteristics}_{ij} + \gamma_1 \text{FN}_{jt} \end{aligned}$$

拡散確率に与える影響 (回帰分析結果から抜粋)

※政治関連の偽・誤情報の分析でも同じ傾向が見られた。

Yamaguchi, S., & Tanihara, T. (2023). Relationship between Misinformation Spreading Behaviour and True/false Judgments and Literacy: An Empirical Analysis of COVID-19 Vaccine and Political Misinformation in Japan. *Global Knowledge, Memory and Communication*, forthcoming.

Vosoughi, S., Roy, D.K., & Aral, S. (2018). The spread of true and false news online. *Science*, 359, 1146-1151. <https://doi.org/10.1126/science.aap9559>

偽・誤情報は選挙結果を左右する可能性がある

- 2つの実際の政治関連の偽・誤情報を使って実証実験をした結果、偽・誤情報を見て支持を下げた人は少なくなかった。
- 特に弱い支持をしていた人ほど偽・誤情報によって支持を下げやすい傾向が見られた。弱い支持の人というのは人数でいうと多い人たちであり、偽・誤情報は選挙結果に影響を与えうる。

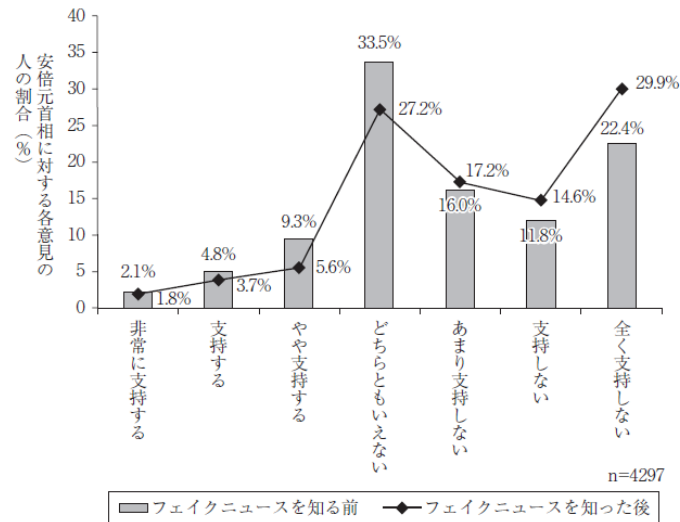


図 3.12 フェイクニュースを知る前後の安倍元首相への支持の分布

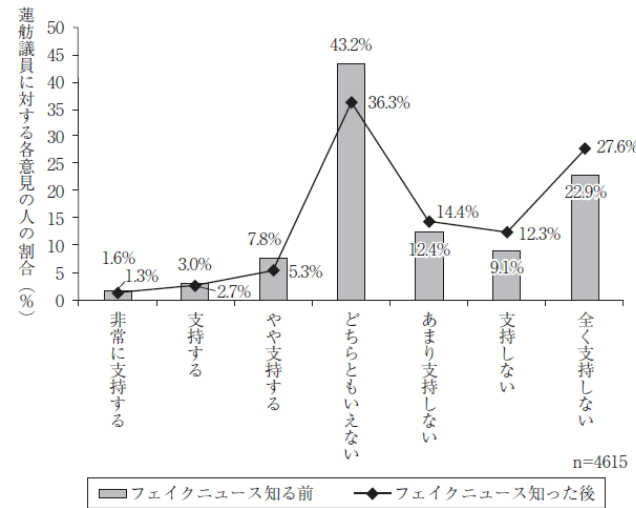


図 3.13 フェイクニュースを知る前後の連舫議員への支持の分布

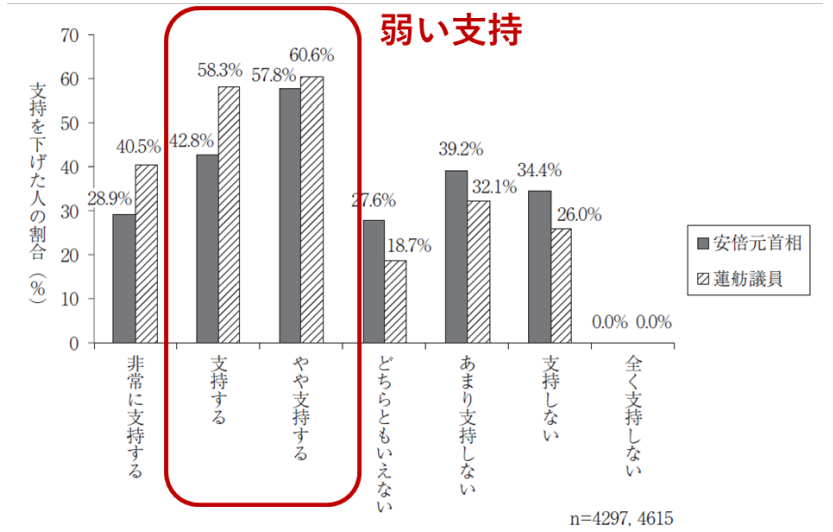


図 3.14 フェイクニュースによって支持を下げた人の割合 (最初の支持別)

生成AIが情報環境に もたらすインパクト

2



ディープフェイクの大衆化が起こった

- AI技術の発展により、誰もがディープフェイクを使えるディープフェイクの大衆化が起こった。これにより、偽・誤情報が爆発的に増加することが予想され、withフェイク2.0時代に。
- 既に災害・戦争などの有事に社会を混乱させる投稿、明確に政治的意図をもって流された情報、詐欺行為など、様々なものに利用されている。今後、裁判などの証拠画像・映像の捏造も増えるだろう。

ドローンで撮影された静岡県の水害。
マジで悲惨すぎる...



午前4:39 · 2022年9月26日 · Twitter for Android

静岡県の水害時に拡散したデマ投稿（2022年）
<https://www.itmedia.co.jp/news/articles/2209/26/news180.html>



ファッションモデルがイスラエル支持を表明する偽動画が拡散
3000万回以上表示
<https://www.nikkei.com/article/DGXZQOGN177EW0X10C22A300000/>



偽音声が発信に使われることも頻発。企業を対象としたものも
<https://www.gizmodo.jp/2023/05/beware-of-fraud-using-ai.html>

世論工作の大衆化

- 誰もが安価に世論工作をしたり、社会を混乱させたりする力を手に入れた。
- 世論工作はビジネスにもなっている。例えば、ある組織が4万件におよぶアバターを作成し、AIでSNSへの投稿を自動作成して世論工作をするビジネスをしていた。既にいくつかの国の選挙で行われており、日本語によるものまでであるという。
- SNSや動画共有サービスと生成AIを組み合わせることで、国際世論の工作・誘導が容易になった。

4万件におよぶ「世論工作アバター」とAI自動生成のフェイクニュースを駆使して、世界の選挙に介入する「影の組織」――。

国際的な調査報道メディアのコンソーシアムが、「世界33の主要選挙に介入した」というイスラエルの世論工作グループ「チーム・ホルヘ」の実態について、一斉にキャンペーン報道を展開している。

「チーム・ホルヘ」は独自のシステムを使い、代表的なソーシャルメディアアカウントや携帯電話番号までを持った4万件に及ぶ架空のアバター（キャラクター）を生成。このシステムには、AIによって世論工作用のツイートなどを自動生成、拡散する機能もあるという。

これらのアバター軍団とフェイクニュースを自在に操作し、クライアントの注文に応じて、世界的に世論工作を展開しているという。

AIを駆使して世論工作をビジネスにしていると指摘
<https://news.yahoo.co.jp/byline/kazuhirotaira/20230220-00337981>

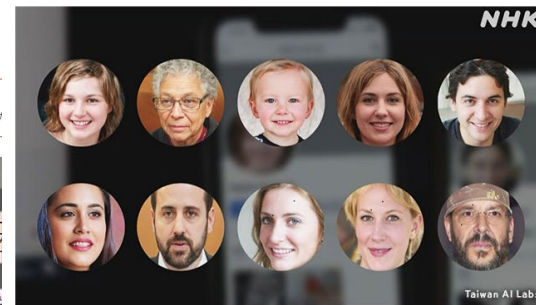
これらの投稿をしたアカウントの顔画像は、欧米風の男児の顔だった。

こうした不審な顔画像が用いられていたアカウントは、ほかにも複数確認され、いずれも、蔡英文政権や欧米に批判的な内容などを拡散させていた。

(投稿抜粋)
「邪悪な資本主義同盟」「東朝が勝ち西負け



SNSで投稿していた実態のアカウント



すべてAIで生成された顔とみられる

これらの人物は実在するのか。

AIで生み出された顔写真を使ったアカウントで偽情報を拡散
<https://www3.nhk.or.jp/news/html/20230824/k10014170241000.html>



イスラエルとハマスの衝突 100万回以上見られた偽動画など33に
イスラエル・ハマスの衝突でも多くの偽画像・偽映像が拡散
<https://www.nhk.or.jp/minplus/0016/topic035.html>

増える誤報と情報環境の悪化

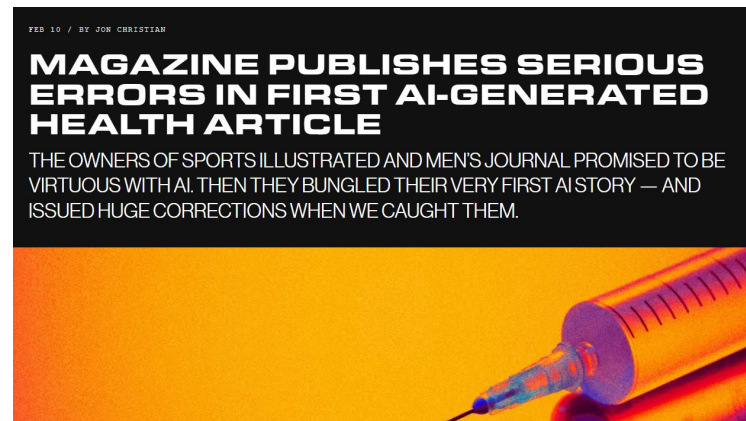
- インターネット上に生成AIによる偽・誤情報があふれ、情報環境が悪化すると予想される。
- 偽・誤情報が蔓延する中で、ある程度の情報の質が担保された「信頼できる情報源」の存在は重要となる。しかし……
 1. メディアのふりをした偽情報の拡散が起こることが考えられる。
 2. メディアが偽画像・映像を取り上げ、誤った報道をする件数が今後急増する可能性がある。
 3. 生成AIを報道に活用することで、誤報が増えることも指摘される。



アメリカ国防総省の近くで爆発が起きたとする偽の画像がネット上で拡散し、株価が一時、下落する騒動に発展

- ブルームバーグを装った「ブルームバーグ・フィード」というアカウントも投稿したことで、株価にも影響。
- インドの主要テレビ局も誤って放送。

<https://www3.nhk.or.jp/news/html/20230523/k10014075821000.html>



生成AIを導入したメディアで、健康コンテンツ1本の中に18もの不正確な記述が見つかった

<https://futurism.com/neoscope/magazine-mens-journal-errors-ai-health-article>

Rise of the Newsbots: AI-Generated News Websites Proliferating Online

NewsGuard has identified 49 news and information sites that appear to be almost entirely written by artificial intelligence software. A new generation of content farms is on the way.

NewsGuardはほとんどAIツールによって書かれたと思われる49のニュース・情報サイトを特定した

<https://www.newsguardtech.com/special-reports/newsbots-ai-generated-news-websites-proliferating/>

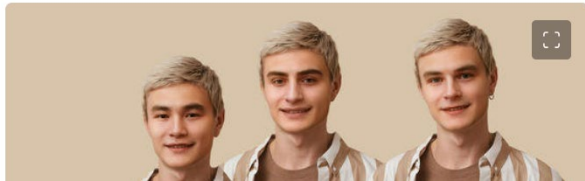
生成AIの最も大きな懸念はディープフェイク

- Microsoftのブラッド・スミス社長は、AIで最も懸念しているのはディープフェイクだと述べた。
- スペイン企業による米国調査では、63.5%の人が、ChatGPT最新モデルで作成されたテキストを、人間が作成・編集したものと誤って回答した。また、米国の美術コンテストでは、AIによって制作された作品が1位となって話題となった。
- この先1、2年で、いよいよテキスト・画像・映像あらゆる領域で、人の目では検証できないレベルになるだろう。

GIGAZINE + フォローする

Microsoftの社長がAIで最も懸念しているのは「ディープフェイク」だと発言

5月26日

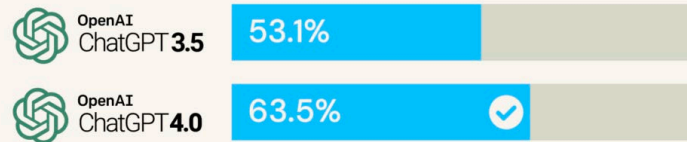


Microsoft社長が、AIで最も懸念しているのはディープフェイクだと発言

<https://futurism.com/neoscope/magazine-mens-journal-errors-ai-health-article>

CHATGPT STATISTICS

How many people thought ChatGPT-generated copy was made by a human?



Read the full report at tooltester.com/en/blog/chatgpt-statistics

tooltester

大半の人が、AIによって生成されたテキストかどうか判断できない

<https://stablediffusionweb.com/#demo>

AI作品が絵画コンテストで優勝、アーティストから不満噴出

© 2022.09.08 Thu posted at 06:59 JST



ジェイソン・アレンさんがAIを使って制作した絵/Courtesy Jason M. Allen

既にAIが生成した絵画が絵画コンテストで優勝している

<https://www.cnn.co.jp/tech/35192929.html>

事実がフェイクになる

- AIによる偽画像・偽映像が蔓延すれば、事実の写真・映像についても「AIが作成したものかもしれない」と疑う必要がある。
- これまでの偽情報と異なるのは、情報を少し検証するだけでは真偽が分からないケースが増えることである。そのため、技術による対抗が必要になるが、AIかどうか判断する技術も完ぺきではないため、誤判定の問題も出てくる。信頼のよりどころがなくなっていく。

「赤ちゃんの遺体画像を"AI生成"と判定」イスラエル・ハマス衝突、AIフェイクの本当のリスクとは

平和博  エキスパート | 桜美林大学教授 ジャーナリスト
11/6(月) 7:10



極めてリアルな生成AIフェイク画像が与えるインパクトは、人々が「フェイクに騙される」ことだけではない。

本物に対しても「これは本物か？」との疑念を広げ、事実を「フェイク化」する危険をはらむ。そして、AI判定の難しさが、さらに事態を複雑にする。

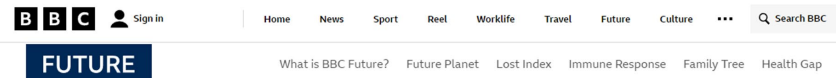
真偽の境目が、加速度的に濁っていく。

AI生成と誤判定したとされる事例

<https://news.yahoo.co.jp/expert/articles/737b2ba5ae41bcad3c6dce7cb1a6184e02e3a50c>
投稿はXより

偽情報の生まれる背景は「経済」と「政治」

- 偽情報が生み出される背景には、①経済的理由、②政治的理由、の主に2つがある。
- 経済的理由については、アテンション・エコノミー*が広まる中で、広告収入目当てに偽情報を流す事例が後を絶たない。世論工作を請け負うビジネスも。
- 政治的理由については、2016年の米国大統領選挙や沖縄県知事選挙、ロシアのウクライナ侵攻など、様々な場面において偽・誤情報が作られている。
- これらの動機は生成AI時代になっても基本的には変わらず、より規模が拡大するだろう。誰でも簡単に参入できるようになったためである。



'I was a Macedonian fake news writer'

In North Macedonia, there's a small industry of websites publishing misleading and inflammatory political articles targeted at US readers. Simon Oxenham meets a woman who worked there.

If you ignored the content, the typical day of a "fake news" writer would seem like any office job. Every morning, Tamara would open her laptop to a fresh email with a link to a spreadsheet. This document contained eight stories based on the other side of the world from her, in the US. The spreadsheet would also contain eight deadlines, each set just a few hours later. Her job was to rewrite each story before her deadline.

* 「関心経済」のことで、情報が指数関数的に増加してとても人々が読み切ることができない時代において、情報の質よりも人々の関心をいかに集めるかが重視され、その関心や注目の獲得が経済的価値を持って交換財になるということを指す。システム1（速い思考）を刺激することが収入につながる。



Russia 'meddled in all big social media' around US election

マケドニア共和国の学生が大量にフェイクニュース記事を生産していた

<https://www.bbc.com/future/article/20190528-i-was-a-macedonian-fake-news-writer>

ロシアによる米国大統領選挙への介入

<https://www.bbc.com/news/technology-46590890>

The research details a **vast campaign spearheaded by the Internet Research Agency (IRA)** - a Russian company that has been described by the United States Intelligence Community as a troll farm with ties to the Russian government.

The report says Russia had a particular focus on targeting conservatives with posts on immigration, race and gun rights.

There were also efforts to undermine the voting power of left-leaning African-American citizens, by spreading misinformation about the electoral process.

Another report, also released today by the Senate, by the research firm **New Knowledge**, similarly highlights Russia's efforts to target African-Americans.

It explains how Russia's IRA were focused on "developing Black audiences and recruiting Black Americans as assets," which included encouraging activists to stage rallies.

社会として偽・誤情報問題 にどう対処するか

3



政府として何ができるか①：法律の功罪

- AIに関する強い法規制（生成AIの利用禁止など）ありきで考えるべきではない。最小限の規制で、ネガティブポイントを抑えて社会的厚生を最大にすることが重要。強い法規制は適切な活用や技術革新を阻害するだけでなく、結局利用を止めることはできず、問題がより見づらい場所で起こることにつながる。
- 偽・誤情報に対する安易な法規制も、表現の自由にネガティブな影響を与える。
- 最初は限定的な運用でポジティブな効果を生み出している、やがて拡大解釈されてネガティブな効果を生み出すという、slippery slopeの問題がある。誹謗中傷や偽・誤情報といった線引きの難しいものでは特に起こりやすい。実際、マレーシアやロシアなど、偽・誤情報対策を強化する法律で、政権に反対する報道や政治家を取り締まるのに使われている例が少なくない。

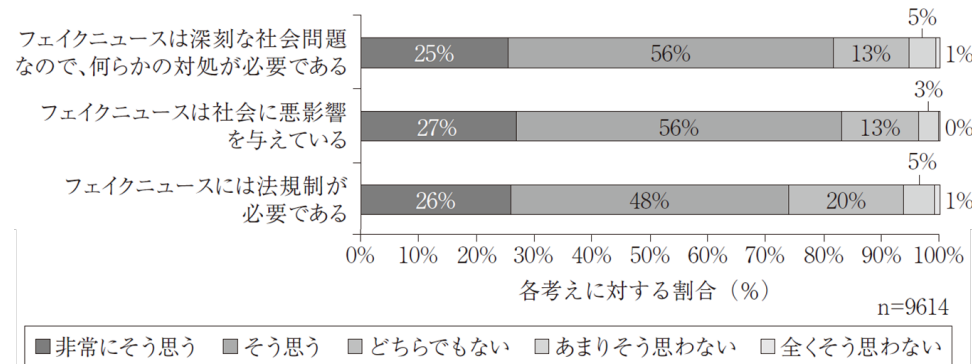


図 3.4 フェイクニュースについてどう考えているか

フェイクニュースに法規制が必要と考える人は74%存在しているが、表現の自由の観点から法規制には慎重な必要がある

山口真一（2022）『ソーシャルメディア解体全書』、勁草書房

ロシア、フェイクニュースと見なせば禁錮刑に 欧米メディア取材停止

ワウクリナ情報
 編集者 = ロンドン、松尾一郎 2022年3月5日 20時26分



ロシア下院=2月22日、ロイター

ロシアのプーチン大統領は4日、ロシア軍の活動に関する報道や情報発信のうち、ロシア当局が「フェイクニュース」（偽情報）と見なした場合に、記者らに対して最大15年の禁錮刑を科せる法案に署名した。刑法に新たな規定を加えて改正する形がとられている。強力な情報統制が敷かれるなか、欧米主要メディアも相次いでロシア国内での取材活動の一時停止などを表明した。

海外では、既にフェイクニュースの法規制が政府に悪用されている事例もある

<https://www.nikkei.com/article/DGXMZO50821000Z01C19A0000000/>

<https://digital.asahi.com/articles/ASQ356J15Q35I1PE00N.html>

マレーシア、偽ニュース対策法廃止へ 下院が再び可決

東南アジア + フォローする

2019年10月9日 23:47



【シンガポール=中野貴司】マレーシアの連邦議会下院は9日、フェイク（偽）ニュース対策法を廃止するための法案を賛成多数で可決した。偽ニュース対策法はナジブ前政権が18年5月の総選挙前に駆け込みで成立させた法律で、恣意的な運用が可能になるとの懸念が強かった。複数の地元メディアが報じた。

政府として何ができるか②：透明性の確保など

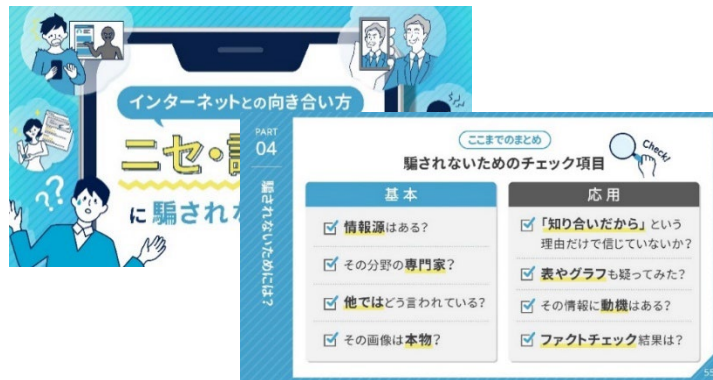
- 業界の自主規制・自主的取り組みを後押しするような役割が期待される。事業者との連携・コミュニケーションが必要。
- 重要なのは目指すべき社会の姿を考え、具体的な透明性・アカウンタビリティの確保を促すことである。得られた結果からエビデンスベースで有効な対策を検討していくことが求められる（効果的な施策の横展開など）。
- グローバルプラットフォーム事業者が多い中で、ローカルな透明性と、ユーザに日本語で対応できる体制を整えることを求めていく必要がある。
- 技術の発展に伴って問題は拡大していく。特にディープフェイクは人間が判断するのは困難になるため、事業者と連携して継続的な把握を行うことが重要である。
- 現行法で犯罪に当たる行為に対して厳正に法的対処をしていくことで、愉快犯なども抑止する。



DALL-Eで作成

求められるメディア情報リテラシー教育の拡充

- 情報の発信だけでなく、受信（メディアや情報の環境・特性など）も含めたメディア情報リテラシー教育を、老若男女に実施していくことが求められる。情報社会においてメディア情報リテラシーを高めることは、教育を受けた人が生きるうえで欠かせないだけでなく、社会全体にとってもこのうえなく必要なことといえる。
- 欧米ではメディアリテラシー教育が進んでおり、義務教育に入っているケースも少なくない。
- どのようなリテラシーが重要なのか、研究によって特定し、そのエビデンスを踏まえた教育啓発を行うことが重要。
- 講座の実施、インフルエンサーを活用した幅広い啓発、教育課程への導入など、多角的な啓発手段で広げていくことが求められる。



総務省と開発した教育啓発教材

分かりやすさを心掛けている。また、「読み上げれば講義ができる」講師向けガイドラインを作成している。

https://www.soumu.go.jp/use_the_internet_wisely/special/ni-segojouhou/

社会技術研究開発事業における 2023年度新規採択プロジェクトの決定について 「SDGsの達成に向けた共創的研究開発プログラム (情報社会における社会的側面からの信頼形成)」

新規採択プロジェクト全（課題解決型）
ニュース発信者と受信者間における「トラスト」形成
研究代表者：藤代 裕之（法政大学 社会学部 メディア社会学科 教授）

概要	研究開発への参画・協力機関
フェイクニュースやヘイトスピーチなどが民主主義社会を揺るがしている。その背景には発信者と受信者が相互作用し、ニュースが生成・拡散されるニュース生態系の「汚染」がある。生態系には、報道機関だけでなく、偽・誤情報を拡散するSNSメディア、世論工作を行う組織など、玉石混交の発信者が存在しているが、発信者の運営・取材・編集方針は多くの場合不透明であり、受信者が信頼性の高い発信者を判断する手がかりが不足している。本プロジェクトは「こたつ監査」の発展、生成AIによる偽情報や動画が増大しており、信頼性判断を一層困難にしている。 本研究開発プロジェクトでは、ニュース生態系における「トラスト」形成の阻害要因を分析する。発信者側としては偽・誤情報対策など信頼性向上の取り組みを国内外で調査する。受信者側としてはニュース採録時における信頼性判断の実態を調査する。また、偽・誤情報対策に有効なリテラシーを特定する。これらの調査から、信頼性判断の手がかりについて不足している要素などを可視化し、分析する。 分析を踏まえ「指標」「制度」から課題解決を目指す。発信者では報道機関などの取り組み事例を分類して信頼性に必要な要因や構造を明らかにし、発信者が公表する項目や表記手法を検討する。その際に「指標」が持つ権力性を緩和するためのガバナンスや「指標」が有効性を持つための規制のあり方を検討する。受信者では、それぞれの関心に基づき信頼性を判断できる「指標」を開発し、判断をサポートするシステム開発も行い、連携する団体や施設の協力を得て実証的に評価検証を行う。 発信者と受信者の相互作用により、信頼性の高いニュースが生態系を構築することで両者間に「トラスト」が形成される社会を目指す。	・法政大学 ・関西大学 ・武蔵大学 ・鳥取県立大学 ・成蹊大学 ・東京工業大学 ・福岡大学 ・株式会社リ・パブリック ・株式会社ゆめみ ・ユニークビジョン株式会社 ・井藤士ドットコム株式会社 など

今後の実証研究で重要なリテラシーを特定予定
<https://www.jst.go.jp/pr/info/info1646/index.html>



総務省、国際大学GLOCOMが監修に入った啓発キャンペーン
インフルエンサーと連携
https://www.youtube.com/playlist?list=PLQntWbrycbJcpM6aVvc5g_nP_HMxPF5weB

技術による対抗

- 人のリテラシーで対応できる範囲を超えるため、AI作成か判断する技術やオリジネーター・プロファイルなど、技術的対抗が求められる。
- 民間での研究開発の活性化を促すと同時に、研究支援を通して研究機関・大学での一体的な開発を促進する。重要なのは技術が大衆化され、誰もが容易に活用できるようにすることである。
- 情報環境が汚染されることを防ぐため、ファクトチェック組織やメディア企業は特に先端的対抗技術が利用できる環境にあることが望ましい。

偽動画 9割見破るAI

東大が開発、世界最高水準 米メタも封じ込め急ぐ

2022年6月7日 2:00 [有料会員限定]

保存

🔍 📄 📧 📱 🐦 📘 📌

「ディープフェイク」と呼ばれる偽動画への対策が進化してきた。東京大学は人工知能（AI）を訓練し、9割前後と世界最高水準の精度で偽動画を見破る手法を開発した。米メタ（旧フェイスブック）や米マイクロソフトなどIT（情報技術）大手も検出ソフトなどを開発し、悪質な偽動画の排除を強化する。政治家などの偽動画がはびこれば社会に混乱をきたすため、封じ込めを急ぐ。



偽画像・動画を見破るにはAIを駆使するしかない
<https://www.nikkei.com/article/DGKKZO61483110W2A600C2TEB000/>

オリジネーター・プロファイルで発信者情報を紐づける

<https://www.nikkei.com/article/DGKKZO61483110W2A600C2TEB000/>

マイクロソフトは微妙な色あせなどをもとに偽動画かどうかを判定するソフトウェアを開発。偽情報対策を推進する活動を通じ影響力の大きな報道機関などに提供する。米アマゾン・ドット・コムも多様な偽動画を見分けられるAIを開発し、論文で公表した。

マイクロソフトは偽動画判定ソフトを開発し、報道機関に提供
<https://www.nikkei.com/article/DGKKZO61483110W2A600C2TEB000/>

Originator Profile 技術研究組合とは

Originator Profile 技術研究組合（OP CIP）は、インターネット上のニュース記事や広告などの情報コンテンツに、発信者情報を紐付ける Originator Profile（OP）技術を開発研究します。

OP CIPは、信頼できる発信元からの情報だとインターネット利用者に表示することで、デジタル空間の信頼性向上を目指します。

注：技術研究組合（CIP）とは、複数の企業、大学、独立法人等が協同して試験研究を行うことにより、単独では解決出来ない課題を克服し、技術の実用化を図るために主務大臣の認可により設立される法人です

参画企業一覧

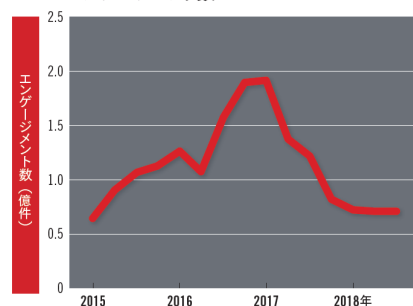


2023年10月時点

プラットフォーム事業者に求められること

- プラットフォーム事業者は、様々な偽情報が飛び交う場を提供している事業者として、改善に向けて常に努力していくことや、透明性の確保が求められる。特に、日本ローカルの透明性が重要。
- 既に有効と考えられている施策の積極的導入、ファクトチェック結果の効率的配信、生成AIを見破る技術の開発と実装など、実施できることは様々にある。
- メディア企業やファクトチェック組織などとの連携の継続・発展も必要。
- 経済的理由から偽・誤情報が拡散されるのを抑止する取り組みが必要である。日本語圏においても、プラットフォーム事業者と連携し、偽・誤情報を取り扱っているウェブサイトに広告収入が流れないような仕組みを構築していくことが肝要だ。
- 研究者と連携した研究や、研究者へのデータ提供などによって情報環境の研究を促進することも求められる。

図表3| フェイスブック上でのフェイクニュースに対するエンゲージメント数



偽・誤情報対策は十分な効果を発揮する

山口真一. (2019). フェイクニュースの正体と情報社会の未来. ダイヤモンドハーバードビジネスレビュー, 2019(1). 64-73.

選挙の誤った情報に関するポリシー



誤解を招くまたは虚偽が含まれる特定の種類のコンテンツで、深刻な危害を及ぼす可能性のあるものは YouTube では許可されていません。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など）が含まれます。

このポリシーに違反するコンテンツを見つけた場合はご報告ください。コミュニティガイドラインへの違反を報告する手順は [こちら](#) でご確認ください。1つのチャンネルに報告する動画やコメントが複数ある場合は、[そのチャンネルを報告](#)してください。

各社ポリシーを持っているため、厳格な運用も期待される

<https://support.google.com/youtube/answer/10835034?hl=ja>

ホーム > ニュース > 経済

偽情報対策のEU新規範、Googleやメタが署名...広告収入停止やディープフェイク監視求める

2022/06/17 12:56

[この記事をスクラップする](#) [f](#) [t](#) [r](#)

【ジュネーブ=池田晋一】欧州連合（EU）の執行機関・欧州委員会は16日、ネット上の偽情報対策を強化する新たな行動規範をまとめた。米Googleやメタ（旧フェイスブック）など30を超える企業・団体が署名したと発表した。コロナ禍やロシアのウクライナ侵攻で偽情報の影響力が懸念されており、対策を急ぐ。



従来は行動規範は2018年につくられた。偽情報の研究者がGoogleなどの巨大IT企業が持つデータを手しにくいといった課題があった。EUは、巨大IT企業に違法な情報への対応を義務づける「デジタルサービス法」の施行を目指しており、新しい行動規範と合わせて

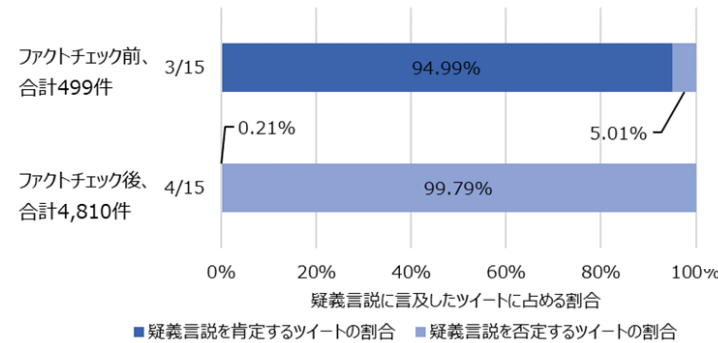
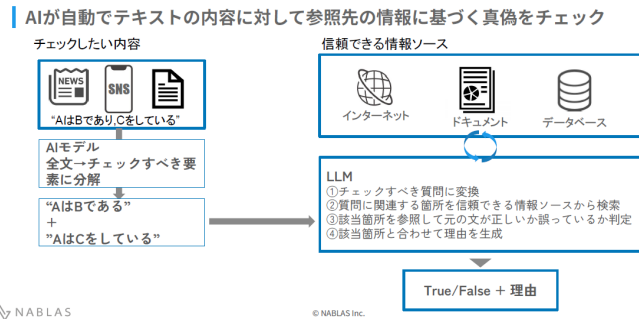
広告収入停止やディープフェイク監視が必要とされる

<https://www.yomiuri.co.jp/economy/20220617-OYT1T50067/>

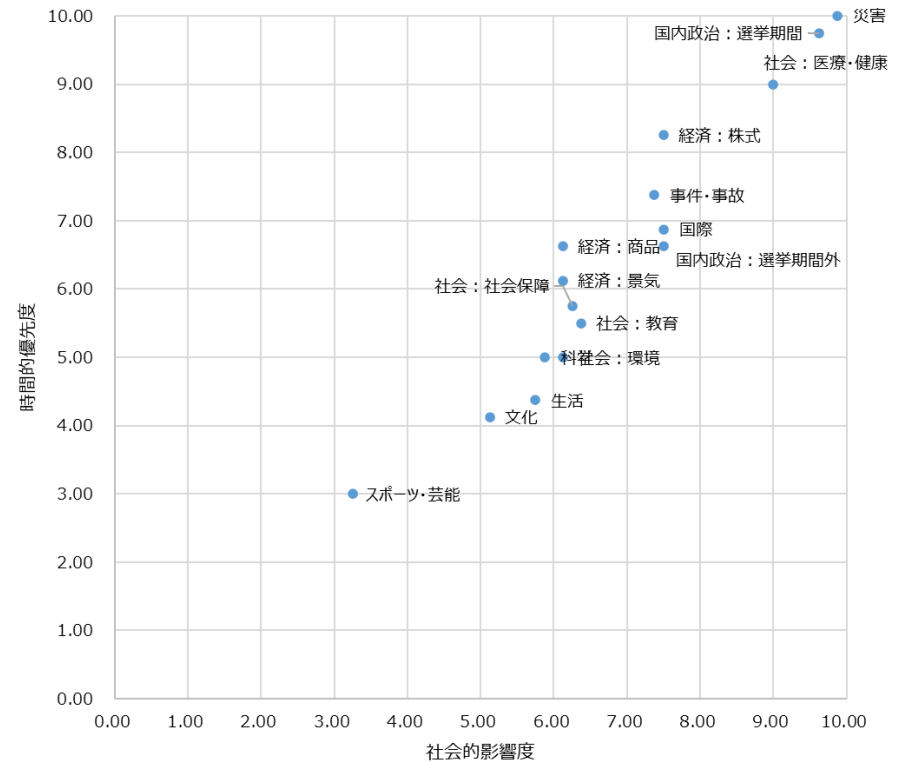
ファクトチェックの推進

- ファクトチェックには情報空間における発信主体を塗り替える力がある。
例えば、菅首相がワクチンを打ったと報道された2021年3月16日に、「打ったワクチンは偽物」という真偽不明情報を肯定するツイートは95%だった。それに対し、ファクトチェック記事の配信後は99.79%がファクトチェック結果を広めようとするものだった。
- 偽・誤情報は膨大であるため、優先順位の高いものからファクトチェックをし、事実を広めることが必要。
- AIによる偽・誤情報スクリーニングなど、技術を上手く活用した効率的なファクトチェックの推進が必要。

テキストファクトチェッカー



「菅首相が打ったワクチンは偽物」に関するツイート数
山口真一 (2022) 『ソーシャルメディア解体全書』、勁草書房



各分野における偽・誤情報の社会的影響度とファクトチェック時間的優先度

有識者へのインタビュー結果から、「時間的優先度」「社会的影響度」の2軸で調査し、散布図でまとめた。

山口真一・谷原史 (2022) 「わが国における偽・誤情報の実態の把握と社会的対処の検討 一政治・コロナワクチン等の偽・誤情報の実証分析一」

重要なのはステークホルダー間の連携

- 偽・誤情報対策に特效薬はない。根絶は不可能であるが、問題を改善していくことはできる。「自由・責任・信頼があるインターネット」を築くために、ステークホルダー間の連携が必須。
- 例えば、メディア・プラットフォーム事業者・業界団体・教育関係者・アカデミアなどが対等な立場で参画し、議論を重ねる会議体などが考えられる。ベストプラクティス共有、技術の共有、偽・誤情報傾向（内容・技術）の情報の共有といったことや、具体的な対策の議論・連携など、幅広い役割が期待される。
- 偽・誤情報問題は国内で完結しない。国際的な連携・情報共有・対策の実施が求められる。

Fake news: BBC and tech firms join forces to fight disinformation

© 7 September 2019



| What is so-called "fake news" and how do you spot it? (From 2018)

BBCは偽・誤情報に対抗するためにIT企業と連携すると発表

<https://www.bbc.com/news/technology-49615771>

Disinformation対策フォーラム

一般社団法人セーフインターネット協会（会長：中山明）以下、SIA）は、偽情報対策について多様なステークホルダーによる協力関係構築を図り、対話の枠組みを設けるべく、「Disinformation対策フォーラム」を2020年6月22日に設置致しました。

参加者一覧

■構成員（有識者） ※敬称略/順不同

沢田登志子 一般社団法人ECネットワーク 理事
穴戸常寿 東京大学大学院法学政治学研究所 教授
瀬尾保 スマートニュース メディア研究所 所長
西田亮介 東京工業大学リベラルアーツ研究教育院 准教授
藤代裕之 法政大学社会学部 教授
安野智子 中央大学 文学部 教授
山口真一 国際大学GLOCOM 准教授

■構成員（事業者） ※順不同

Facebook Japan株式会社
Google合同会社
ヤフー株式会社
Twitter Japan株式会社

■オブザーバー ※順不同

一般社団法人 日本新聞協会
日本放送協会
一般社団法人 日本民間放送連盟
総務省
消費者庁

Disinformation対策フォーラムも連携のために設立された

<https://www.saferinternet.or.jp/anti-disinformation/>

G7 広島 AI プロセス G7 デジタル・技術閣僚声明 (2023年9月7日)

14. 我々は、エビデンスに基づく政策議論を進める上で、OECD、GPAI、UNESCO 等の国際機関と協力してプロジェクトベースの取組を推進することを計画している。このようなプロジェクトベースの取組には、AI に対する信頼を高め情報環境を支援するために、外国からの情報操作を含む AI を活用した偽/誤情報を識別するための最先端の技術的能力に関する研究と理解を進めること等、OECD の生成 AI に関する G7 の共通理解に向けた報告書の中で G7 メンバーによって特定されたものが含まれる。

G7広島AIプロセスでも国際機関と協力してプロジェクトベースの取組を推進することが明記されている

<https://www.bbc.com/news/technology-49615771>



Existing Practices against Disinformation (EPaD)

In the Ministerial Declaration of the G7 Digital and Tech Ministers' Meeting held on April 29-30, 2023, we recognized "the importance of actions taken by a wide range of stakeholders, including social media platforms, civil society, the internet technical community, and academia, to address online information manipulation and interference and disinformation while respecting human rights, in particular the right to freedom of expression." Also, the Declaration stated: "We plan to collaborate in gathering and compiling existing practices by various stakeholders to address online disinformation as Existing Practices against Disinformation (EPaD), and we intend to publicize and present this report at the UN IGF 2023 in Kyoto." We are pleased with the publication of this EPaD, and it is expected to be referred to when tackling online disinformation.

G7各国等における多様な関係者による偽情報対策に関する既存プラクティス集がIGF2023で公表

<https://www.bbc.com/news/technology-49615771>

ご清聴ありがとうございました