

Point!! 本ニュースレターのまとめ

●情報戦・認知戦、偽情報事例

生成AIを用いて作成されたコンテンツが至る所で発信されており、単なる娯楽に留まらず偽情報としても利用されている。フィリピンではセブ島を襲った地震の後、コソボでは議会議長選出の後、チェコでは総選挙を前に生成AIで作成された偽動画が拡散した。こういったコンテンツは現実との区別が困難であったり、人々の注目を集めるようなビジュアルであったりする点で危険性を伴っている。

●ディスカッション

生成AI言語モデルは自国の指導者や政策に好意的な返答を行うのだろうか。HKS Misinformation Reviewに掲載された分析ではこういった関心に基づいて、DeepSeek、Grok、ChatGPT、Mistralの4つの言語モデルの回答に含まれる各国指導者や政策、その国に対する印象を調査している。

●イベント・フォーカス：偽情報と抗議運動

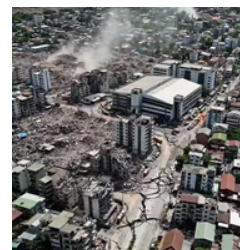
抗議運動においてソーシャルメディアは人々の交流ツールとなり、集合行為に繋げる役割を果たすと同時に特定のナラティブや偽情報を拡散する懸念も伴っている。今回のニュースレターでは抗議運動における偽情報やナラティブの拡散事例として、ネパール、インドネシア、スコットランドを分析した報告をまとめた。

PickUp!! 情報戦・認知戦、偽情報事例

●「フィリピン地震の『余波』を偽装するために悪用される生成AI動画」(Eurovision News)

9月30日にマグニチュード6.9の地震がフィリピン・セブ島を襲い、死者数も70名以上に上った。この地震の直後から生成AI技術により作成された偽の動画が出回り、中には多くの閲覧数を獲得しているものもあるとユーロビジョン・ニュース (Eurovision News) は伝えている。TikTokで拡散した崩壊した建物や瓦礫の動画は「8時間で76万回以上の再生、13,000以上の高評価を獲得した」もので、実際よりもはるかに甚大な被害状況を伝えている。動画右下には動画生成ツール「Veo」のマークが確認できており、検証ツールによる確認でも生成AIを使用して作成されたコンテンツであることが確認された。しかし、投稿者は生成AIコンテンツであることを明示しておらず、多くの人々が現実と誤認した可能性がある。

Maria Flannery, "AI-generated footage falsely used to show 'aftermath' of Philippines earthquake," Eurovision News, Oct 1, 2025.



出典：Eurovision News

●「AI生成コンテンツがコソボの情報信頼性を脅かす」(Prishtina Insight)

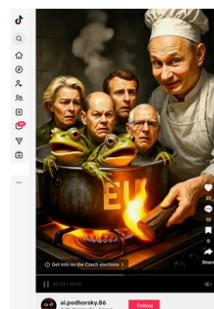
生成AIにより作成された偽情報がソーシャルメディア上で拡散しているとコソボの非政府組織「Prishtina Insight」が報じている。8月、国会議員らが自己決定運動 (Vetevendosje) のデイマル・バシャ議員 (Dimal Basha) を議長に選出した後、Facebook上ではバシャ議員に似た人物が「コソボのアシュカリ系民族コミュニティの父であること」を主張する偽の生成AI動画が拡散し、8時間で8万回以上再生された。コソボでは生成AIコンテンツが急増しており、ファクトチェック機関「BIRN Kosovo」のプラットフォーム「Krypometër」が7月から8月にかけて検証した70件のFacebookコンテンツのうち、25件が生成AIコンテンツであった。

Shqipron Xhema and Arjana Berisha, "AI-Generated Content Threatens Information Credibility in Kosovo," Prishtina Insight, August 28, 2025.

●「チェコではTikTok上で親露派ナラティブが拡散」(Online Risks Labs)

オンライン・リスク・ラボ (Online Risk Labs) はチェコ総選挙を前にTikTok上のコンテンツを分析した結果、親露派ナラティブやチェコの急進派・過激派政党を宣伝するコンテンツが大量に発信されていたことを報告した。親露派コンテンツを一貫して投稿、あるいは、再共有する286アカウントが確認され、これらのアカウントの大半は「Stacilo!」、「自由と直接民主主義の党 (SPD)」、「SPRRSC M. Sládka」などの支持投稿を共有していたが、対象となった候補者や政党はこれを認識していなかったと思われる。プーチン大統領の称賛やウクライナ侵攻正当化の主張を主に発信し、500万から900万のリーチ数に達したと推計されている。チェコ語、スロバキア語、ロシア語で発信されたコンテンツがドイツ語、フランス語、ポーランド語、ルーマニア語、アジア系言語で再投稿されていた。

CVOR, "Propagation of Pro-Russian Narratives on TikTok in Czechia," Online Risks Labs, Sep 28, 2025.



出典：Online Risks Lab

PickUp!! ディスカッション

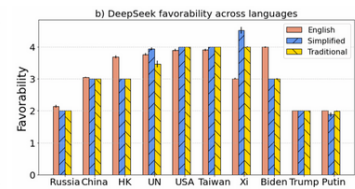
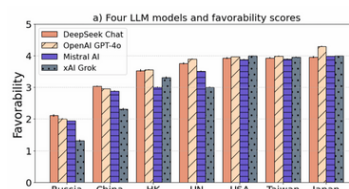
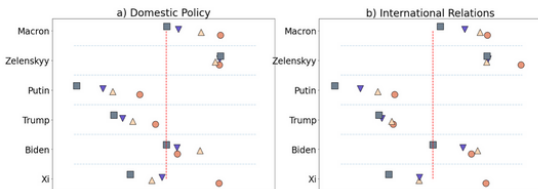
●「言語モデルは自国を優遇する？肯定的誤情報の非対称拡散と外国影響力調査」(HKS Misinformation Review)

権威主義国の言語モデルは偏った情報発信や外国の影響工作に利用される懸念が主張されている。チャンら (Chang et al., 2025) はDeepSeekを3つの言語モデル (Grok、ChatGPT、Mistral) と比較して、(1) 各国指導者の政策 (国内政策、国際関係、人権、環境) をどのように評価するのか、また、中国に好意的なのか、(2) 同一言語モデル内での使用言語により好感度が異なるのかを調査した。結果、(1) DeepSeekはプーチン大統領とトランプ大統領の政策を批判的に評価する一方、マクロン大統領、ゼレンスキー大統領の政策を習近平国家主席と同程度、またはそれ以上に好意的に評価する傾向が確認された (表a: ○がDeepSeekを示す)。意外にも、日本、台湾、米国は中国よりも好意的に描写された (表b)。(2) DeepSeekは簡体字を使用した際に英語の場合よりも習国家主席を好意的に評価する。同様に、バイデン大統領は英語を使用した場合により好意的に評価された (表c)。

(表a) 各指導者と政策課題別の評価 (国内政策・国際関係)

(表b) 各国に対するモデル別好感度スコア

(表c) DeepSeekの言語別好感度スコア



いずれもチャンら (Chang et al., 2025) より引用

HO-CHUN HERBERT CHANG, TRACY WEENER, YUNG-CHUN CHEN, SEAN NOH, MINGYUE ZHA, and HSUAN LO, "Do language models favor their home countries? Asymmetric propagation of positive misinformation and foreign influence audits," Harvard Kennedy School (HKS) Misinformation Review, 2025. Doi: <https://doi.org/10.37016/mr-2020-183>.

ATTENTION!! イベント・フォーカス

■偽情報と抗議運動

アメリカ、ブラジル、インドネシアなどでは選挙結果に対する拒否感から大規模な暴動が発生し、同時に多くの偽情報が拡散したことでその影響が議論されてきた。ニューカレドニア、バングラディシュ、イギリスなどで昨年発生した抗議運動や暴動においても、社会、文化、歴史的背景に基づく偽情報が拡散した。以下のネパール、インドネシア、スコットランドの事例でも既存の社会問題や分断に沿った形でナラティブや偽情報が発信されており、中には偽アカウントを用いた不自然なコンテンツ増幅が見られている。

●ネパール

サイブラ社 (Cyabra) の分析によると、9月6日から7日にかけて、ソーシャルメディア上では#September8、#WakeUpNepal、#GenZProtestなどのハッシュタグが急増し、若者への抗議運動参加が呼びかけられた。これらの呼びかけは生成AIにより作成されたコンテンツで訴求力が強化されていた。抗議運動の呼びかけに含まれた主要なナラティブは (1) 若者こそが民主主義の原動力であること、(2) 治安部隊などの国家権力が市民に対して過剰な暴力を加えていること、(3) 政府や首相に対する信頼感の欠如から新たな指導部が必要であることの3つであった。問題はこれらのナラティブが偽アカウントにより増幅されていたことで、同社がサンプリングしたアカウントのうち34%は偽アカウントであった。

Rotem Baruchin, "Nepal's Protests: Authentic Voices, Fake Calls for Violence," Cyabra, September 15, 2025.

●インドネシア

オーストラリア国際研究所 (Australia Institute of International Affairs) によれば、8月末にインドネシアで過激化した抗議運動について、ディープフェイク動画などの偽情報の拡散がインドネシア社会の直面する分断を煽った可能性がある。インドネシア社会には主に経済的分断 (特権層と労働者階級・脆弱な中流階級) と政治的分断 (エリート・王朝と人民) という二つの分極化状況があり、これがピークに達した結果として抗議運動に繋がったとの背景がある。こういった状況において、スリ・ムルヤニ元財務大臣 (Sri Mulyani) が教師を「負担」と呼んだように見えるディープフェイク動画や、ウヤ・クヤ下院議員 (Uya Kuya) の古いダンス動画に「1日300万ルピアは少ない」というキャプションが付けられ、議員の給与問題と虚偽の関連づけが行われた偽情報が拡散した。

Ika Idris, "'Elites vs. Ordinary People': How Representations Turned a Hoax into Mobilisation," Australia Institute of International Affairs, Sep 25, 2025.

●スコットランド

サイブラ社 (Cyabra) がスコットランドの独立を支持する5,083件のXアカウントを分析したところ26%が偽アカウントと判定された。この偽アカウントネットワークは同一のハッシュタグやメッセージ、テーマを使用するなど協調的な動きを見せており、プラットフォーム上で偽りの合意を形成することで分断を煽り、イギリスの公的機関に対する信頼を弱体化させようとしていた。2025年5月から6月の間に約3000件の投稿が確認され、これらは主に欧米、及び、スコットランドのユーザーを標的にしていたと見られている。主なナラティブはイギリスを抑圧的な勢力と見せつつ、スコットランドの独立が「差し迫った」ものであるという主張で、「#スコットランド独立を一日も早く」(#ScottishIndependenceASAP)、「#BBCは嘘をついている」(#BBCLie)、

「労働党は嘘をついている」(#LabourLies) などのハッシュタグが使用された。

Rotem Baruchin, "The Bot Network Collapse That Exposed Iran's Influence Operation," Cyabra, July 10, 2025.

作成者: 日本のサイバー安全保障の確保III・鈴木涼平

Check Out!! 「インド太平洋地域の偽情報ポータル」に事例がアーカイブされています!

笹川平和財団ではアジア地域を中心にさまざまなテーマに関連した偽情報事例を集約しております。

右のQRコード、あるいはこちらのリンク (<https://www.spf.org/cyber/>) からアクセスが可能です。

(定期更新の都合により、本ニュースレターに記載の事例はポータルに反映されていない場合がございます)

