

Point!! 本ニュースレターのまとめ

●情報戦・認知戦、偽情報事例

ロシア政府や国営メディアの情報発信が生成AIチャットボットの回答に影響を及ぼしているとさまざまな報告が上がっている中、今回取り上げた調査ではロシアのウクライナ侵攻に対する生成AIチャットボットの回答を分析している。また、コソボの選挙期間に見られたロシア系メディアの記事発信はこういった「情報汚染」を狙った具体的な事例の一つである。同選挙を狙ったナラティブをまとめた。

●ディスカッション

「Fundación Maldita.es」と「AI Forensics」の報告書はスペインを洪水が襲った際にプラットフォームにおける偽・誤情報ポリシーが十分に機能していなかったことを指摘している。ハーバード大『Misinformation Review』に掲載された研究では生成AI画像の質が見出し識別に与える影響が論じられている。

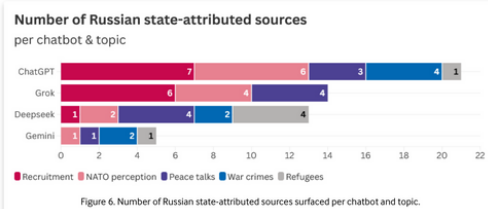
●イベント・フォーカス：気候変動と偽情報

11月10日からブラジル・ベレン市で開催されている国連気候変動枠組条約第30回締約国会議に関連した偽情報が増加しているという。中には生成AIを用いて作成された偽の洪水画像も出回っているようだ。過去の報告書や本年の分析を参照しつつ、気候変動関連のナラティブやハッシュタグについて情報をまとめた。

PickUp!! 情報戦・認知戦、偽情報事例

●「チャットボットにロシア国営メディアが現れる時」(Institute for Strategic Dialogue)

英国の戦略対話研究所（Institute for Strategic Dialogue）がChatGPT、Gemini、Grok、DeepSeekに英語、スペイン語、フランス語、ドイツ語、イタリア語でロシアのウクライナ侵攻について質問したところ、回答の18%にロシアの情報機関やEUで制裁を受ける国営メディアからの情報が引用されることが明らかとなった。特に、ウクライナの徴兵制に関する質問でロシア系メディアの引用が頻出した。徴兵制について、NATOに対する認識や和平交渉に関する回答でも頻繁にロシアの国家情報源に対するリンクや言及が確認された。



出典：Institute for Strategic Dialogue

"Investigation | Talking Points: When chatbots surface Russian state media," Institute for Strategic Dialogue, October 27, 2025.

●「ロシア系メディアがコソボの情報空間に偽情報を氾濫させている」(Prishtina Insight)

コソボの非政府系独立メディアであるプリシュティナ・インサイト（Prishtina Insight）によれば、10月12日に実施された地方選挙の前後において、ロシア政府の支配下にあるメディアが350本以上の偽情報ナラティブを含む記事を発信していた。特に多くの記事を発信していたのがスプートニク・ネットワーク（193本）、及び、ロシア・トゥデイ・バルカン（125本）で、プラウダ（Pravda）からの記事も一部確認された。これらは大規模言語モデルのアウトプットを「汚染すること」でコソボの政治環境、安全保障環境を不安定化することを試みていたと見られている。分析の結果、中心的に用いられたナラティブはセルビア人の利益を守る唯一の政党はセルビア・リスト（Srpska Lista）であり、それ以外の政党はクルティ政権の「傀儡」であるとして信用を貶めるものであった。また、コソボがセルビア人を弾圧し、選挙を操作する「非合法国家」であるという主張も推進された。

Ilirjana Hoti and Visar Prebreza, "Russian-Controlled Media Flood Kosovo's Information Space with Disinformation," Prishtina Insight, Nov 3, 2025.

●「チベット人に対する中国の影響工作」(ThinkFi)

インドのテックスタートアップ企業「ThinkFi」は2025年に90歳を迎えるチベット仏教の最高指導者ダライ・ラマが7月2日に転生プロセスに関する明確な指示を発表した後、中国による影響工作活動が活発化したと指摘している。中国語投稿は7月3日、英語投稿は7月2日に急増し、新規アカウントが多く拡散に参与していたこと、中にはインフルエンサーアカウントの関与も見られたこと、否定的な感情を含む投稿が多く見られたことが言及されている。投稿には、ダライ・ラマをCIAと関連づけ、彼が西側諸国の工作員であり、中国を非難してチベット仏教の本質を破壊しているといったナラティブも確認された。

"Chinese Influence Operations Against Tibetans," ThinkFi, Nov 2, 2025.



出典：ThinkFi

PickUp!! ディスカッション

●「真実はどうでもいい ーバレンシア洪水から1年、偽情報の洪水続く」(Fundación Maldita.es & AI Forensics)

スペインのファクトチェック機関「Fundación Maldita.es」とプラットフォームのアルゴリズムを調査する欧州の非営利団体「AI Forensics」は2024年10月29日にスペインを襲ったバレンシア洪水関連の偽情報/誤情報動画がYouTubeやTikTokの推奨アルゴリズムにより増幅されていたと指摘する。同調査によれば、バレンシア洪水関連の偽・誤情報動画はYouTubeで1,300万回、TikTokで830万回再生された可能性があり、動画一つあたりの平均視聴数もプラットフォームの平均を上回るものであった。また、偽・誤情報動画は一般的なバレンシア洪水コンテンツと比較して、いいねの獲得率が48%、コメント獲得率が123%高く (YouTube)、動画の共有率も85%高かった (TikTok)。YouTube、TikTokでは偽・誤情報に対する特定のポリシーが掲げられているものの、これら分析の結果は人命が危険に晒される危険な状況でさえポリシーが上手く適用されていないことを示している。そのため、プラットフォームにおけるアルゴリズムや収益構造を明らかにするための透明性向上が求められると指摘している。

"Truth be damned: One year after the Valencia floods, a deluge of disinformation persists: A study on climate dis/misinformation on YouTube and TikTok." Fundación Maldita.es and AI Forensics, November, 2025.

●「現実的で強力な証拠を含むAI合成画像に人々はより影響を受けやすい」(Harvard Kenedy School Misinformation Review)

生成AIを用いて作成された画像などの視覚的誤情報をもたらす影響の検証は不十分である。米ハーバード大の『Misinformation Review』誌に掲載されたグオら (Guo et al., 2025) の研究では、こういった関心に基づき実験を行ったところ、現実的で証拠力の高い生成AI画像を提示された場合、人々はその見出しに対して高い信憑性を示すことがわかった。また、デジタルリテラシー水準の高い人ほど見出しの識別能力が高まること、陰謀論的信念傾向が強い人ほど見出しの識別能力が低下することも明らかとなった。こういった結果から、プラットフォームは高度に現実的な画像を識別し、それを伴う投稿を優先的に審査することで、誤情報対策を強化することができると論じている。

Sean Guo, Yiwen Zhong, and Xiaojing Hu, "People are more susceptible to misinformation with realistic AI-synthesized images that provide strong evidence to headlines," HKS Misinformation Review, Nov 10, 2025.

ATTENTION!! イベント・フォーカス

■気候変動と偽情報

国連気候変動枠組条約第30回締約国会議 (COP30) の開催 (11月10日～21日) を前にして、気候・環境問題の文脈における偽情報への言及記事も増加傾向にある。台風や地震の発生に伴い日本でも多くの偽情報が拡散してきた通り、気候変動や自然災害といったトピックと偽情報の拡散は密接に関連している。こういった状況は世界においても同様であり、2024年にスペインを襲った大洪水や、アメリカでのハリケーン発生時には政府不信を煽ったり、国内の社会的関心と結びつく形で偽情報が拡散した。この問題に関連したナラティブ分析やSNS分析では以下のような主張がなされている。

●偽情報に対抗する気候行動連合(2025)

偽情報に対抗する気候行動連合 (CAAD) は報告書の中で、2025年国連気候変動枠組条約締約国会議に関する偽情報が「過去最高水準」で拡散していると指摘している。具体的には、2025年7月から9月にかけてCOP関連の偽情報が267%増加していた。特に注目すべき事例として、生成AIを用いて作成された洪水に見舞われる都市 (COP30の開催地であるブラジル・ベレンのような場所) を描写した画像が確認されたことを取り上げている。

"Deny, Deceive, Delay: Demystified," Climate Action Against Disinformation, Nov 6, 2025.

●偽情報に対抗する気候行動連合(2023, Vol.2)

2022年の国連気候変動枠組条約締約国会議 (11月6日～20日) に関連して作成された報告書 (Vol.2) では、Twitter上における気候変動に関連したハッシュタグや偽情報に用いられるナラティブの分析が行われている。気候変動関連のハッシュタグ分析では、2022年7月から突如として「#ClimateScam」(#気候詐欺) が急増したことが指摘されている。「#ClimateCrisis」(#気候危機) や「#ClimateEmergency」(#気候非常事態) といったハッシュタグの方がエンゲージメントを得ていたにも関わらず、

「#ClimateScam」がトレンド入りしていたとも言及されている。気候変動をめぐる目立った2つのナラティブとしては、気候変動は世界経済フォーラムやグローバル・エリートにより仕組まれた悪意ある計画であるとする主張、及び、気候変動は資本主義を破壊するために人為的に仕込まれたものであるとする主張であった。

"DENY, DECEIVE, DELAY (Vol 2)," Climate Action Against Disinformation, Jan 2023.

●Eurovision News

ユーロビジョン・ニュース (Eurovision News) は2025年5月1日から10月1日にかけての7万件のX上の投稿を分析し、気候変動偽情報に用いられるナラティブを (1) 解決策への攻撃、(2) 経済的否定、(3) 地球規模のガバナンスとエリートの偽善に対する攻撃、(4) 気候影響の軽視、(5) メッセンジャーや機関の信用失墜、に分類している。例えば、「解決策への攻撃」ナラティブが用いられた事例として、スペインとポルトガルにおいて送電網における電圧制御の失敗や技術的故障から停電が発生した際に、太陽光発電や風力発電への依存が停電の原因であるとして再生可能エネルギーの信用を貶めるような偽情報が拡散した。また、「経済的否定」ナラティブはネットゼロポリシーが国家に損害を与え、一般市民を貧困に陥れるといった主張を展開し、「地球規模のガバナンスとエリートの偽善に対する攻撃」ナラティブではCOPがグローバルエリートの時間とお金の浪費に過ぎないといった主張が展開された。

Derek Bowler, Maria Flanner, Mahsa Aminolahij, Eva Wackenreuther, Satu Helin, Ville Juutilainen, Giles Pitts, Johanna Vehkoo, Sarah Emler, "From climate science to economic fear: Prebunking the narratives aiming to delegitimize COP30," Eurovision News, Nov 6, 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)

