# 生成AI時代のフェイクニュースの広がり

平 和博 ●桜美林大学 教授

**生成AIの爆発的普及で、フェイクニュースの脅威は真偽の境界を揺るがす新たな時代に突入した。イスラエルとハマスの軍事衝突では「AIフェイク」拡散の一方、「本物」が「フェイク」とされる騒動も起きた。**

## ■「生成AIフェイク」が株価を下げる

「#ペンタゴン（米国防総省）のビル近くで爆発」。「CBKニュース」と名乗るX（旧ツイッター）アカウントが、建物脇で大きな煙の上がる画像を投稿したのは2023年5月22日午前8時42分（現地時間）だった。

同様の投稿が同日午前10時すぎにかけて、ブルームバーグ通信とは無関係の「ブルームバーグフィード」を名乗るXアカウントや、ロシア国営メディア「RT」のアカウントなどからも拡散されていった。

だが画像の建物は、国防総省のビルとは異なり、しかも一部が溶けたように見えるなど、不自然な点が目立った。これは、生成AIで作られたフェイク画像だった。この騒動の影響で、米株式市場のS&P500種株価指数は、一時0.3％の下落を見せたという。

2022年11月末のChatGPTの登場をきっかけに、熱狂的な生成AIブームが世界を覆った。ただそのインパクトは、メリットと同時に大きなリスクにもなり得るとの懸念を呼び起こした。その一つが、フェイクニュース作成・拡散などへの悪用だ。

ChatGPTの登場から半年後、2023年5月のG7広島サミットでは、生成AIの活用と規制の取り組み「G7広島AIプロセス」を掲げた。その成果の一つ、経済協力開発機構（OECD）が9月にまとめたレポートによると、生成AIのリスクとして参加7か国が一致して挙げたのは「偽情報／情報操作」、すなわちフェイクニュース問題だった。

生成AIの脅威として懸念されるのは、フェイクニュース作成・拡散の大規模化、低コスト化、巧妙化、リアルタイム化だ。

これまでも紛争や災害などの非常時に、フェイクニュースが大量に拡散し、情報の混乱を引き起こしてきた。生成AIの登場で、フェイク作成のハードルが一気に下がり、リスクのレベルが格段に跳ね上がることが懸念されている。

3月20日には、オンラインの調査報道機関「ベリングキャット」の創設者、エリオット・ヒギンズ氏が、「ドナルド・トランプ前米大統領逮捕」という架空の生成AI画像をXに連続投稿。リアルな画像は680万回を超す再生数を集め、騒動となった。その週末には、シカゴの建設作業員が「バレンシアガの純白のダウンコートを着たローマ教皇」の生成AI画像をフェイスブックなどに投稿し、やはり騒動になった。

そして前述の「ペンタゴン爆発」のフェイク画像騒動は、一時的とはいえ現実社会の株価に影響を与えた。では、生成AIの普及によって、軍事

紛争では何が起きるのか。それを現実に示したのが、イスラエルとハマスの軍事衝突だ。

## ■イスラエルとハマスの軍事衝突

10月7日朝、ハマスによるイスラエルへの大規模攻撃によって始まった軍事衝突でも、大量のフェイクニュースが拡散した。

「ウクライナがハマスに武器を売却」「ハマスの攻撃はイスラエルか西側陣営による『偽旗作戦』だった」――そんなフェイク投稿がソーシャルメディアなどに噴出する。その中には、「ジョー・バイデン米大統領が徴兵を発表」とするディープフェイクス（AIで作ったフェイク動画）や、「瓦礫に挟まれ助けを求める幼児」「イスラエル人が難民キャンプに避難」などの生成AIによるフェイク画像も含まれていた。

ただ、それを上回る勢いで拡散したのは、「リサイクルフェイク」とも言うべき過去の画像や動画の使い回しだ。

同様の傾向は、2022年2月から始まったロシアによるウクライナ侵攻でも見られた。同年3月半ばには、「降伏宣言」を口にする偽の「ウクライナのヴォロディミル・ゼレンスキー大統領」や、「和平合意宣言」を口にする偽の「ロシアのウラジーミル・プーチン大統領」のディープフェイクスが拡散した。しかし、圧倒的多数のフェイク画像や動画はリサイクルフェイクだった。

## ■フェイク対策、Xの後退

イスラエルとハマスの衝突で改めて浮き彫りになったのは、イーロン・マスク氏が2022年10月末に買収した後のX（2023年7月24日に「ツイッター」からブランド名を変更）のフェイクニュース対策の後退ぶりだ。マスク氏は買収後に8割にのぼる大規模リストラを実施。フェイクニュース対策部門も解体した。さらに、従来は審査を経て表示し、一定の信頼性担保の役割があったアカウントのチェックマークも、月額980円の有料ユーザーであれば誰でも表示されるようなった。

また2023年5月には、欧州連合（EU）の違法有害情報対策のためのプラットフォーム規制法「デジタルサービス法（DSA：Digital Services Act）」と連動する自主ガイドライン「行動規範」から離脱した。EUが9月に発表した報告では、フェイクニュース（偽情報）の発見率はXが最も高かった。

EUはこれに先立つ8月25日、前述のDSAを、域内の月間アクティブユーザーが4500万人を超すグーグル、フェイスブック、Xなどの超大規模プラットフォームに適用開始したばかりだった。

イスラエルとハマスの衝突をめぐるフェイクニュース氾濫を受けて、EUは10月10日にX、翌11日にはフェイスブックなどを運営するメタに相次いで対策強化を要請。さらにXに対しては12月18日、DSAに基づく正式調査を開始している。

## ■「嘘つきの分け前」の広がり

生成AIの広がりによるインパクトは、本物と見分けがつかない画像や動画の氾濫だけではない。本物が「生成AIフェイク」として否定されるリスクも突き付ける。

衝突開始から5日後の10月12日、イスラエル首相府は、ハマスに殺害されたとする4枚の乳児の遺体画像をXに投稿。同日、イスラエルを訪問したアントニー・ブリンケン米国務長官に、ベンヤミン・ネタニヤフ首相が見せたものだとし、表示数は780万回を超えた。

そのうちの1枚の焼死体の画像が、波紋を広げた。生成AI検知ソフトが、これを「AIによる生成」と判定した画像とともに、「フェイク画像だ」との主張がXなどで拡散。表示数は2200万回を超した。判定ソフトを開発した米ベンチャー企業は、画像の乳児の名札の部分にモザイク処理がさ

れていたことに、ソフトが反応したとし、「判定結果は決定的なものではなかった」と釈明した。

同様のケースは他にもあった。イスラエルのインフルエンサーが10月20日、「ハマスの指導者たちが贅沢な暮らしを満喫している」との書き込みとともに、プライベートジェット搭乗の様子などを示す4枚の画像をXに投稿した。

投稿画像には、AI生成画像に特有の不自然な部分が目立ち、「AI生成フェイク」との批判が広がる。だが4枚の画像は、すでに2014年にはメディアに掲載されていた本物だった。ただし、投稿したインフルエンサーは、画質の低い元画像を、AIを使った高精細化処理をした上で投稿していた。これが原因でAI処理特有の不自然さが表出。AI検知ソフトでも「AI生成」の判定が出る状態だった。

TBSの報道番組「サンデーモーニング」は、2023年11月5日に、この4枚の画像を「生成AIで作られたフェイク画像」と誤って放送。翌日、訂正・謝罪した。

テキサス大学法学部部長のロバート・チェスニー氏とバージニア大学教授のダニエル・シトロン氏は2018年、ディープフェイクスの副作用として、「実際に起きた本当のことに対して、嘘つきが説明責任を容易に回避できるようにしてしまう」と指摘。「嘘つきの分け前（liar's dividend）」と呼んだ。そのリスクは、生成AIの広がりの中で、さらに深刻なものとなっている。

## ■生成AIが作る「コンテンツ工場」

生成AIはウェブサイトの粗製乱造にも使われている。米ウェブ評価サイト「ニュースガード」は6月26日に公表した調査で、生成AIを使って自動生成した低品質のウェブサイト、“コンテンツファーム（工場）”が急増し、世界的企業などの広告費を飲み込んでいる実態を明らかにした。

ChatGPTなどの生成AIは、人間と見分けのつかない自然な文章を量産できる。その機能を使って、メディアを偽装したサイトを立ち上げ、フェイクニュースや低品質な自動生成コンテンツを次々に掲載し、広告収入を獲得しているという。

そこには、主な企業だけでも日本を含む141社にのぼる広告が掲載されていたという。さらに広告の9割以上が、グーグルによる「プログラマティック広告」で配信されていたとしている。調査によれば、ニュースサイトを擬した「ワールド・トゥデイ・ニュース」というサイトの場合、6月9日から15日までの1週間で約8600件、1日平均で約1200件もの記事を公開していた。

また米サイバーセキュリティ企業「レコーディッド・フューチャー」は12月5日、偽造ニュースサイトとボットアカウントのネットワークを組み合わせた親ロシアの影響工作キャンペーン「ドッペルゲンガー」で、生成AI使用が疑われる事例を確認した、と報告している。それによれば、米国を標的とした「エレクション・ウオッチ」というサイトの「バイデン政権、人道援助の遅れが深刻化する中東の危機と闘う」という記事が、生成AI検知ソフト「ZeroGPT」で66.06％の確率で「AI生成」との判定が出たと指摘している。

メタは8月に公表した「敵対的脅威レポート」の中で、「ドッペルゲンガー」について「2017年以降にわれわれが対処した中で、最大かつ最も積極的で持続的なロシアからの秘密の影響工作だ」としている。

## ■ラテン語やバスク語のフェイク

ニュースガードは9月11日、ハワイ・オアフ島で8月に発生した大規模な山火事をめぐる、中国発と見られるフェイクニュースの拡散ネットワークの存在を明らかにした。

調査では、フェイスブックやXなど14のプラッ

トフォームの85のアカウントを特定した。「MI6によると、この山火事は自然発生したものではなく、米国政府によって人為的に引き起こされたものです。(中略)『気象兵器』を極秘に開発していることが判明した」――中国語や英語など16言語で、そんな陰謀論を発信していたという。

メタは8月末、中国の大規模影響工作ネットワーク「スパモフラージュ」に関係する7700件のアカウント削除を発表している。陰謀論の発信アカウントは、このネットワークの一部だという。

この陰謀論拡散は、日本も標的になっていた。筆者が確認したところ、同様の陰謀論が、アメーバブログ、ピクシブ、楽天ブログなどを舞台に、新たに開設したアカウントで発信されていた。その中には、あわせて福島第一原発の処理水海洋放出を批判する投稿もあった。

使われていた言語は、日本語や英語のほか、イタリア語やラテン語もあった。また、1つのアカウントが、山火事とは別のテーマについて、インドの公用語の一つであるオリヤー語、エチオピアやケニアで使われるオロモ語、バスク語の3言語で投稿している例もあった。生成AIの翻訳機能を使って多言語発信をしていた可能性がある。

また国内では、7月から11月にかけて、岸田文雄首相が卑猥な言葉を口にするという、日本テレビのロゴを付けたディープフェイクスが、Xやニコニコ動画などに拡散した。

## ■「選挙の年」2024年への懸念

選挙をめぐるフェイクニュースは民主主義を揺るがす。ブラジルでは2023年1月8日、首都ブラジリアの連邦議会、大統領府、最高裁判所に、前年の大統領選の「不正選挙」を主張する前大統領、ジャイール・ボルソナーロ氏の支持者ら約5000人が乱入するという事件が起きた。これは2021年1月6日に、前年の米大統領選の「不正選挙」を主張した前大統領、トランプ氏の支持者らが連邦議会議事堂に乱入した事件を思わせた。

2024年は「選挙の年」だ。台湾総統選（1月）、インドネシア大統領選（2月）、ロシア大統領選（3月）、インド総選挙（4〜5月）、欧州議会選（6月）、米大統領選（11月）などが予定される。従来型のフェイクニュースの氾濫に加え、生成AI、ディープフェイクスによる拡散の大規模化、巧妙化などの懸念が高まる。

規制の枠組み作りも進む。G7広島サミット以来続いてきた広島AIプロセスは2023年12月1日、AIをめぐる「国際指針」「国際行動規範」を含む「包括的政策枠組み」が合意された。EUでは初の包括的な規制法「AI法案」が、2026年施行に向けて最終手続きに入っている。また、メタ、ユーチューブ、ティックトックなどのプラットフォームは、ユーザーによる生成AIコンテンツの表示義務化を相次いで打ち出す。

だが、この1年の生成AIの進化と普及を考え合わせると、変化のスピードとフェイクニュースの拡散はさらに加速することが見込まれる。社会に及ぼすインパクトも、はるかに大きなものになることは間違いない。

2024年は、生成AIと民主主義社会にとっての試金石になりそうだ。

インターネット白書 ARCHIVES

1996,1997,1998,1999,2000....

# ［インターネット白書ARCHIVES］ご利用上の注意

このファイルは、株式会社インプレスR&Dおよび株式会社インプレスが1996年～2024年までに発行したインターネットの年鑑『インターネット白書』の誌面をPDF化し、「インターネット白書ARCHIVES」として以下のウェブサイトで公開しているものです。

https://IWParchives.jp/

このファイルをご利用いただくにあたり、下記の注意事項を必ずお読みください。

- 記載されている内容（技術解説、データ、URL、名称など）は発行当時のものです。
- 収録されている内容は著作権法上の保護を受けています。著作権はそれぞれの記事の著作者（執筆者、写真・図の作成者、編集部など）が保持しています。
- 著作者から許諾が得られなかった著作物は掲載されていない場合があります。
- このファイルの内容を改変したり、商用目的として再利用したりすることはできません。あくまで個人や企業の非商用利用での閲覧、複製、送信に限られます。
- 収録されている内容を何らかの媒体に引用としてご利用される際は、出典として媒体名および年号、該当ページ番号、発行元などの情報をご明記ください。
- オリジナルの発行時点では、株式会社インプレスR&Dおよび株式会社インプレスと著作権者は内容が正確なものであるように最大限に努めましたが、すべての情報が完全に正確であることは保証できません。このファイルの内容に起因する直接的および間接的な損害に対して、一切の責任を負いません。お客様個人の責任においてご利用ください。

お問い合わせ先 | インプレス・サステナブルラボ | ✉ iwp-info@impress.co.jp