学べるゲラ

第131回
AI悪用　多業種で防げ

※2026年2月の毎日新聞記事を元にした文章です。校閲し、直すべきところを指摘してください。

生成ＡＩ（人工知能）を用いた偽の画像や動画「ディープフェイク」がこの数年飛躍的に精度を高め、企業や自治体もその驚異にさらされている。対抗技術を開発する企業では、自治体のほか異分野の組織を含めた幅広い連携も進んでいるという。通常のＡＩ判定ツールでは判断できないケースも出る中、偽情報の検知に特化したシステム開発が進められている。

モニターに映し出されたビデオ通話の画面で、突如者の顔が別人にすり替わると、赤い文字で「‼ＤｅｅｐＦａｋｅ‼」と警告が表示された。富士通が開発したなりすまし不正を検知する技術のデモンストレーションだ。

香港では２０２４年、企業の担当者がビデオ会議で最高財務責任者（ＣＥＯ）を装った相手にだまされ、２億ﾄﾞﾙ（約40億円）を振り込む事件が起こるなど、生成ＡＩを悪用したビデオ会議の不正は世界的に問題になっている。

話者の顔をリアルタイムで認識して違和感なく別人の顔になりすまし、標的となる人物の表情や顔の向きの変化にも追随できるため、見破るのは容易ではない。

そこで富士通は独自技術を開発。動画に含まれる細かな特徴を捉え、目や口、ひげなど顔のパーツレベルで見破ることを可能にし、人種や性別などで判定結果の偏りも抑えた。富士通データ&セキュリティ研究所の担当者は「ＡＩ開発に注力してきた企業だからこそ、攻

撃者側の考えを理解して対策を打ち出せる」と強調する。

更に24年からはＮＥＣや東京大、国立情報学研究所など、産学組織の９社と連携し、インターネットやＳＮＳ上の偽情報を見抜くシステムの開発を進めている。

ＮＴＴ東日本とＡＩスタートアップの「ＮＡＢＬＡＳ」（東京）も、ＳＮＳ上などのフェイクコンテンツの拡散を防ぐための自治体のシステムを開発した。

例えば、ＳＮＳに投稿された道路のひび割れやクマの出没などの画像や動画をシステムにアップロードすると、加工が疑われる箇所を色を変えて示す。その上で「フェイクの可能性が高い」「輪郭が不自然で背景と完全に一致していない」などと理由とともに表示する。

正規の発信元からの情報であると市民が判別できる技術も実装した。自治体職員がＳＮＳやウェブサイトに画像や動画をアップする際、電子証明書をそれらに埋め込むことで仮に一部が改ざんされて、ネット上で転載・拡散された後でも、改ざんを見破ることができる。

自治体が判別用システムを導入し、市民がそこにアクセスする形での運用を想定している。

25年８月から長野県伊那市で実証実験をしており、26年度以降の展開開始を目指す。実際に市民から送られてくる「クマを目撃した」とする画像や動画の真偽判定に活用されているという。

ただ生成ＡＩの精度向上で悪用される業種や領域は急速に広がっており、対策が追いついていないのが現状だ。

そのため異なる業界間の幅広い連携を通じ、新たな技術開発につなげようとする動きが進む。

富士通は25年12月、偽情報やＡＩリスクに対応する国際的なコンソーシアム「Ｆｒｏｎｔｉａ」を立ち上げた。参加組織はＩＴや金融、コンサルティング、研究機関など幅広く、欧州やイタリア、オーストラリアの組織も加わった。

富士通データ&セキュリティ研究

所の担当者は「複雑な社会問題となっている偽情報は、我々だけの力では解決が難しい。世界中の企業・組織が協力して解決策を生み出すことが重要だ」と意義を語る。

Ｆｒｏｎｔｉａに参加する東京海上ホールディングスのＡＩ部門の担当者は「生成ＡＩの悪用は、保険業の根幹が揺るがしかねない問題で強い危機感を持っている。偽情報の確認は保険査定でも今後必須になるだろう」と話す。保険業界では、自動車事故の証拠となる画像をＡＩで偽造されることなどで、保険金詐欺への悪用の懸念があるという。

フェイクニュースに詳しい笹原和俊・東京科学大教授（計算社会科学）は「偽情報の問題は詐欺などの実害にとまらない。『何を信じてよいのか分からない』という不確実さを社会に広げ、人々の不安を増幅させるおそれもある」と指摘。生成ＡＩによるディープフェイクでそのリスクが飛躍的に増大しているとし、各業界の垣根を越えた研究開発の更なる推進が急務だと訴える。