気になるこの用語 第89回

消費生活相談の周辺用語を取り上げ、やさしく解説します。

小木曽 健 Ogiso Ken
国際大学グローバル・コミュニケーション・センター　客員研究員
講演やメディア出演を通じ、ネットで絶対に失敗しない方法を伝えている。全国の企業・学校などで2,000回以上の講演。著書に『13歳からの「ネットのルール」誰も傷つけないためのスマホリテラシーを身につける本（コツがわかる！ジュニアシリーズ）』（メイツ出版、2020年）ほか多数

AIによる情報戦のリスク

前回はネットやSNSにおける誹謗中傷についてお話ししました。今回は、最近急速に発達しているAIについて、今SNSで何が起きているのか、AIは私たちの社会にどんなリスクをもたらしているのか、具体的にお伝えしたいと思います。

災害デマという地獄

2024年の元日に発生した令和6年能登半島地震（以下、能登半島地震）。多くの方が被災したこの震災は、「災害デマ」が劇的に変化したターニングポイントでもありました。

そもそも災害デマとは、大地震や豪雨など社会を混乱させるレベルの災害が発生したときに、その混乱に乗じて虚偽の情報を拡散させる行為のこと。

東日本大震災では「倒れたサーバーに挟まれて動けない」というウソの投稿が拡散し問題となりました（災害デマの「はしり」とも言われています）。熊本地震の「ライオンが逃げた」デマ投稿は有名ですよね。今や大きな災害が起きる度に、毎回何かしらのデマが拡散する状況です。中には違法性を問われ、逮捕されたような悪質なケースもあったのですが、とはいえ、これまでの震災デマはあくまで愉快犯的な動機によるもの。悪質ではあっても、その動機は実に幼かったのです。ところが……。

愉快犯からの変質

2024年の能登半島地震は、「X」が投稿の閲覧数に対して報酬を支払うプログラム、つまり投稿がバズれば利益を得られるというしくみをスタートさせて以降、最初に発生した大地震でした。また、世の中に広く生成AIが普及し、誰もが使えるようになったタイミングでもありました。そんな中で起きた災害だったのです。あの時、ネットでは何が起きていたのか、時系列で見ていきましょう。

地震発生後ほどなくして、被災した現地から「家から出られない」「消防につながらない」といった救助要請が投稿され始めます。もちろんその大半は、生命の危機に瀕した「本当の」救助要請です。自宅の住所番地を記載していた人もいました。それだけ切迫した状況だったということです。

そんな投稿を目にした多くの人たちは、心配し、その投稿を善意で拡散していきます。ここまでは問題なかったのですが、この様子を眺めていた一部の不届き者がこう考えたのです。

> 「いま救助要請を投稿すると、簡単に注目を集められる。これはチャンスだ。閲覧数を稼ぎ『X』から報酬を得られるぞ」

そして現地の「適当な住所」「架空の住所」を添えた救助要請を投稿するという「災害デマ活動」を開始したのです。

もしこれら虚偽の救助要請を信じて、救急隊やレスキュー隊が救助に向かってしまったら、本当に助けを求めている人たちに向けられたはずのリソースが無駄になりかねません。人命に直結する許しがたい行為、報酬欲しさに他人の命すら踏み台にする非常に悪質な行為であり、もう論外なのですが、実はこの時、一部の海外SNSアカウントがAIを使って、日本に向けて数多くの「災害デマ」を投稿していたと言われています。そのしくみはこうです。

すべてをAIで自動化

SNS投稿で利益を得ようとしている連中は世界中にいます。一部はAIを導入し、全世界のネッ

トの動向を常に注視。地球のどこかで誰かの投稿がバズれば、AIが即座に検知して、その投稿がどんな理由でバズっているのか分析し、その騒ぎに便乗すべく行動を開始します。能登半島地震でもそれが行われていたようなのです。

AIが「日本で何かバズっているな」と感知して、今日本で、日本語で救助要請を投稿すると拡散される可能性が高いのか、と分析。虚偽の救助要請など「バズりそうな内容」の投稿を生成し始めたと言われています。

以前なら、生身の人間が世界中のバズり投稿をチェックするなど不可能でした。また言葉の壁（特に日本語は難しいので）が、悪だくみの越境をそれなりに防いでいたのですが、「悪人＋AI」という最悪のコラボレーションにより、世界中から流暢（りゅうちょう）な日本語のデマが押し寄せてくるという、もう地獄のような事態がスタートしてしまったのです。

コストと不信感

ここ1年ほど、SNSなどで見かける詐欺メッセージの日本語（海外からのウソ投資話やロマンス詐欺など）が、以前と比べてかなり自然なものに変化していることにお気付きでしょうか？試しにチャットで話しかけてみたのですが、まるでネイティブの日本人とやり取りしているような、とても自然な会話でした。恐らくここにもAIが活用されているはずです。AIの発展は、違法行為・迷惑行為の国境を取っ払ってしまったとも言えます。

少し前に、大きな話題となっていた学校でのいじめ動画や暴行動画が「拡散している様子」も、当然ながらAIは認識しているでしょう。「日本では、いじめの暴露動画を投稿するとバズる」という知見を与えてしまった以上、近い将来、ニセのいじめ動画によるバズりねらい投稿が出現してもおかしくはありません。「いじめの暴行シーン」をゼロから生成するなどたやすいこと。架空のいじめ動画で閲覧数を稼ぐ連中が現れても不思議ではないのです。

こういったニセ情報が社会にはびこると、誰もがすべての情報を疑う必要に迫られ、社会全体に膨大なコストが生まれます。情報に対する不信感も増大する。これらはもう、すでに始まっているのです。

SNSのショート動画で「赤ちゃんと子犬」「事故やアクシデント」みたいな動画はバズりやすいので、いまやSNSはAI作成による「その手のニセ動画」であふれています。

だから面白そうなコンテンツを見つける度に「これは本物なのかな」と気になり、コンテンツを心の底から楽しめなくなってしまった……そんな方もいらっしゃると思います（少なくとも私はそうです）。これこそまさに社会全体に与えるコストなのですが、でもコストですんでいるうちはまだマシなのです。

目的はカネじゃない

2025年10月号でもお伝えしましたが、最近、日本国内で行われた選挙に対して、海外からAIを使った介入があったのでは、と言われています。AIが社会分断を加速させそうな投稿を見つけては意図的にバズらせたり、フェイクを作成しては拡散させていたという……。ただ問題なのは、その目的がカネではないという点です。

日本の社会を混乱させる目的で、特定の国家により行われていたのだろう、というのが複数の専門家による見立てであり、実際、流暢な日本語で「沖縄は日本から独立せよ」という主張を繰り返していた「自称」日本人アカウントが、実は某国から投稿されていたということがバレてしまい、そのアカウントが瞬時に消えた、なんてこともありました。

AIによる国境を越えた情報戦は、もう始まっています。ですが、本連載でこれまでお伝えしてきたテクニックを活かし、AIによる攻撃を無効化（あおるような真偽不明の情報は気にせずに無視、拡散もしない）すれば、AIだろうが対国家だろうが私たちが負けることはありません。

今起きていることを知り、おのおのが備える。これは誰もが取り組める、また今日から始められる「悪意を持ったAI」との戦い方です。