ニューノーマル処方箋（第55回）

AIの「禁止事項」とは何か？AI事業者ガイドラインを読む

2024.07.31

＜目次＞

生成AIが詐欺に利用されている

簡単な命令を入力するだけで、その命令に沿った文章や画像などを自動で生成する「生成AI」（Generative AI）を、日々の業務に使用しているビジネスパーソンも多いでしょう。文章や画像だけでなく、簡単なプログラムコードも生成できるため、幅広く活用されています。

このように生成AIを活用するシーンが増えると、AIが犯罪などに悪用される機会も増えることにつながります。

2023年には、サイバー犯罪向けの生成AIツール「WormGPT」が誕生しました。WormGPTは、マルウエア関連のデータを学習した生成AIで、例えばビジネスメール詐欺に使用される、ユーザーをだますような偽情報のメール本文を自動で生成します。

WormGPTは2024年1月に終了していますが、今後も似たような生成AIが生まれ、犯罪に悪用される可能性は十分に考えられます。生成AIが社会にもたらすリスクは、どうすれば抑えられるのでしょうか？

そのヒントとなるのが、経済産業省と総務省が2024年4月19日に公開した「AI事業者ガイドライン（第1.0版）」です。この資料では、AIに関わる事業者が守るべき事項が記されており、AIを活用する際のリスクや、そのリスクに対してどのように対処すべきかといった行動規範が示されています。

AIがやってはいけないこと「偽情報を作る」

AI事業者ガイドラインには、AIを扱う企業が「やってはいけないこと」「留意すべきこと」が明記されています。

例えばWormGPTのように、AIで「偽情報」を作り出すことも、やってはいけないことの1つです。生成AIによって、誰でも真実・公平であるかのように装った情報を生成できるようになったため、偽情報や誤情報が社会を不安定化・混乱させるリスクは

常に潜んでいます。

ガイドラインの別紙では、生成AIによって生み出された偽情報が、社会に悪影響を及ぼした例を紹介しています。例えばアメリカでは、弁護士が審理中の民事訴訟にて、資料の作成にChatGPTを利用した結果、実際には存在しない判例を引用した事例があったといいます。

ディープフェイク（AIによって生成された偽画像・偽動画）も、偽情報の1つです。アメリカでは2023年5月に「国防総省付近で爆発が起きた」という偽画像がSNSで拡散され、その結果、ニューヨーク株式市場のダウ平均株価が一時100ドル以上下落しました。

AIが人間の意思決定や感情を不当に操作してはいけない… 続きを読む