第 69 回：「生成 AI の悪用：セキュリティ危機と身の守り方」

掲載日： 2024 年 4 月 5 日
執筆者： 株式会社スクウェイブ
社長室マネージャー
沼澤　綾佳

昨今は生成 AI の進化発展が目覚ましく、DX 推進のために活用している企業・自治体も増えてきている。その最中、やはりというべきか、もうというべきか、バーチャル誘拐という生成 AI を利用した恐ろしい事件が起きていた。
2023 年 4 月、米国アリゾナ州に住む女性は、匿名の人物から突然電話を受け、15 歳の娘を誘拐したという内容を告げられ、さらに、身代金として 100 万米ドルを支払わなければ暴行を加えると脅されたという。女性は電話で聞いた声について「確かに娘のものであり、後ろで泣き叫び、懇願しているものと理解した。ただし、娘と電話越しで話をすることについて、誘拐犯は許可しなかった」と述べている。

電話口で娘の声を聞かされたこの女性にとって、生成 AI で作られた娘の声は、声はもちろん抑揚も話し方も娘のそれとそっくりそのままであったというから驚きである。実際は身代金を払う前に誘拐は起きていないことが判明し、事なきを得たが、AI の発展とともに騙す手口も多様かつ巧妙に変化していることを象徴する事件だったと考えている。これではビジネスメール詐欺のようにメールだけに気を付けていればよいということはなく、電話でも顔出しの Web 会議であってもそれが本当に本人なのか、見分けることは極めて難しくなる可能性がある。

こういう場合、何よりもまず、慌てて身代金を支払ってはならない。なりすましも起き得ることを念頭に置いたうえで、まずはメールよりも本人に通じる番号へ電話する、さらに電話よりもテレビ電話や直接会って確かめるなど、より確実な手段が望ましいだろう。
技術の進化、発展とともにいずれはこういったことが起きても不思議ではないと漠然と思ってはいたものの、著名人のみならず一般人も標的にされてしまっている。今後どのように身を守るべきなのか、今から真剣に向き合う必要があるだろう。

先述の例は大きなニュースであるものの、身の回りで生じるレベル感では、フリーメールアドレスに着信する詐欺メールが増加し、かつ巧妙化してきていると感じる。昔は日本語がおかしいメールが主流だったのだが、そういったおかしな日本語が減り、企業ロゴや見栄えも本物に寄せたようなメールになっている。ここにも生成 AI の学習による弊害があ

ると考える。

余談だが、なぜ詐欺メールの日本語がおかしいメールになるかというと、送ってきている大半が外国人であることと、外国人にとって日本語は非常に難しい言語であるためだ。漢字とひらがなとカタカナが混ざっているために、例えば「し」と「レ」のように似た文字の区別がつかないこともそうだが、漢字の読み方ひとつとっても難しい。「日」という同じ漢字であっても、「日々」、「日曜日」、「日本」や「祝日」といった全ての「日」の読み方が異なるという難しさがある。日本人にとっては当たり前に読めてしまう漢字ではあるが、外国人にとってはそれが音読みか訓読みかを判断することすら難しい事例は枚挙に暇がない。また、「大丈夫」といった、文脈によって意味が変わってしまう言葉も多い。そういったハイコンテクストな事情もあり、これまで海外の詐欺師にとって日本語で精巧な詐欺メールを作成することはさぞ高い壁であったことだろう。我々は日本語に守られていたともいえる。

しかし、生成 AI がその問題をクリアしつつあることも事実である。日本人が英語の翻訳を AI ツールに任せるのと同様に詐欺メールも AI 翻訳によって学習し、進化してきている。忙しい日々の中、疲れたあなたがメールをチェックする際、詐欺メールは日本語がおかしいものだけだと甘く考えているといつか本当に騙されてしまうかもしれない。ならば生成 AI だけでなく私たちも学習する必要があるだろう。

元日に能登半島地震が発生したが、発生直後のテレビでは被災地で接続できるフリーWi-Fi の案内があった。この善意の是非はひとまず置いておくが、きっと正常な判断ができればむやみにフリーWi-Fi に繋ごうとはしないだろう。公共の電波でフリーWi-Fi の情報を知らせるということは、それを狙ったハッカーが攻撃を仕掛けてくる可能性があり、大変危険だからだ。
では、果たして緊急時や疲れている時に、情報が錯綜する中で、慌てずに正しい判断ができるだろうか。実際に自分が疲れているのか、動揺しているかどうかは、自分では意外とわからないものである。それでも冷静に対応する必要があることを我々はこの 2024 年の元日に突き付けられてしまった。
生成 AI にただ怯えるのでなく、自分で実際に利用し、何ができるかを知るだけでも良いので、皆さんにもぜひ自分ができることから対策を始めて頂きたい。

スクウェイブとしても、セキュリティ意識レベルを可視化し、その後の教育ツールでもご支援ができるので、ぜひお気軽にお声掛け頂きたい。
情報セキュリティ意識レベル診断：https://www.k2wave.biz/msis-t
遊びながら学べる教育ツール CyberCross®： https://www.k2wave.biz/cybercross

ちなみに疲れているかの判断は当社のストレス診断からできる。
（https://k2wave.viewer.kintoneapp.com/public/39e0f08f0b47802d6f9cbd7603023abfcc1a58ebf743ca3101b33b011252164e/）。
通常のストレス診断とは異なり、様々な人間の表情を見てその感情を判断するものである。無自覚の疲労やストレスが溜まっていると人間の表情を読み取れないことを利用したものであるが、本当に疲れているとものの見事に間違える。仕事に対するモチベーション診断とセットになっており、面白いのでこちらもぜひ無料診断をお試しいただきたい。
モチベーションマネジメント：https://www.k2wave.biz/r-2-scm2