2023 年 2 月 27 日

Venture Solution

コラム記事

いま全世界で話題の「ChatGPT」
IT 業界だけではなく、多くの業界・企業から注目されている対話型 AI チャット bot です。
既に自社サービスに組み込み、業務効率化・お客様満足度向上と言った利用をしている企業も多く存在する反面、
サイバー犯罪者による悪用も懸念されると感じております。
そんな中で、ChatGPT が犯罪に悪用される可能性についての記事が掲載されておりましたので、ご紹介いたします。

もしサイバー犯罪者が“話題の対話型 AI”「ChatGPT」を犯罪に悪用したら

（Tech Target Japan 2023/2/27（木）23：35 配信より引用）
人工知能（AI）技術を活用したチャット bot（以下、AI チャット bot）「ChatGPT」に関心を寄せるのは、ライターなどのクリエイターだけではない。攻撃者も AI チャット bot に目を付けており、攻撃の手段として悪用する可能性がある。

ChatGPT が“最凶の犯罪ツール化”に

開発元の OpenAI は ChatGPT の機能を一部制限し、悪用を防ごうとしている。人種差別につながる言動ができないようにするなどだ。一方で同社への主要な出資元である Microsoft は、ChatGPT の活用範囲を広げようとしている。既に Microsoft は、自社の AI サービス「Azure OpenAI Service」に ChatGPT を組み込むことを表明済みだ。

セキュリティベンダーSophos のプリンシパルリサーチサイエンティスト、チェスター・ウィスニエフスキー氏は、ChatGPT が攻撃の実行をいかに容易にするかを研究している。ウィスニエフスキー氏によると、ChatGPT の進化はフィッシング（情報窃取を目的とした詐欺行為）を容易にする可能性がある。ChatGPT に潜む危険性を同氏に聞いた。

―― 攻撃者は ChatGPT に、どのような可能性を見いだしているのでしょうか。

ウィスニエフスキー氏　フィッシングをはじめとする攻撃の手段として、ChatGPT などの AI チャット bot を捉えた場合、大きく 2 つの可能性を見いだすことができる。1 つ目は技術的な可能性、2 つ目は社会的な可能性だ。

技術的な可能性とは「ChatGPT に、マルウェア（悪意のあるプログラム）開発といった悪事をさせることができる」といったことを指す。私が簡単に調べたところ、ChatGPT にマルウェア開発の手伝いをさせることは可能だと考えられる。

こうした技術的な可能性は、あまり重要だとは考えていない。あるマルウェアを検出した場合、そのソースコードを書いたのがイワンという名の男性なのか、キャロルという名の女性なのか、ChatGPT という名の AI チャット bot なのかは、さほど問題にならないからだ。マルウェアは誰が開発したのかを問わず、マルウェアであることに変わりはない。

影響が大きいと考えているのが、ChatGPT の社会的な可能性だ。社会的な可能性とは、いかに自然に会話ができるか、いかに上手に文章が書けるか、といったことを指す。ChatGPT を使えば、巧妙な文章を書くスキルがなくても、特定の企業や個人になりすますことが非常に簡単にできてしまう可能性がある。

※次回「ChatGPT の『まるで人』の自然さが生み出してしまった“次世代の脅威”とは？」は、ウィスニエフスキー氏が危惧する、ChatGPT の社会的な可能性を深掘りする。

TechTarget ジャパン

開発元である OpenAI も ChatGPT を利用した犯罪や悪用を防ぐため、一部機能の制限をかけていますが、
別のニュース記事では、ウイルス対策ソフトに検出されないマルウエアの生成に成功したとの報告もございました。
多くの可能性を持つ ChatGPT を正しく理解し、利用していくことが重要であると感じております。