AI 詐欺メールに騙されないために

～ハーバード実験が示した“2 人に 1 人が騙される時代”～

ライフデザイン研究部 主席研究員/テクノロジーリサーチャー 柏村 祐

1.「怪しいメール」は、もう過去の話である

一日に何十通もの迷惑メールが届く。多くの方がこの問題に悩まされているのではないだろうか。メールの振り分け機能を使っても、次々と別のアドレスから送られてくる。しかも、その中には国税庁のe-TAXやクレジットカード会社を装った「本物そっくり」のメールが紛れ込んでいる。かつての「当選おめでとうございます！」といった明らかに不自然な詐欺メールとは異なり、今や正規のメールと見分けがつかないほど巧妙化しているのだ。

そして、多くの方が「自分はそんなメールには騙されない」と思っているはずである。しかし、その自信は、もはや通用しない時代になりつつある。実際に、クレジットカード会社や銀行を装ったメールのURLにアクセスしてしまい、個人情報を入力した結果、不正送金の被害に遭うケースが後を絶たない。国税庁のe-TAXや大手クレジットカード会社の本物そっくりのサイトに誘導され、銀行口座やクレジットカード情報を盗まれるといった被害が報告されている。ウイルス感染、個人情報の漏洩、フィッシング詐欺による金銭被害などは、今や誰にでも起こりうる現実なのである。

では、なぜこれほど本物そっくりの迷惑メールが大量に送られてくるようになったのだろうか。そして、なぜ私たちはこれほど簡単に騙されてしまうのだろうか。

かつて、特定の個人や会社を狙い撃ちにする巧妙な詐欺メールは、専門的な知識を持つ犯罪者が、多くの時間をかけてターゲットの情報を調べ上げる「特別な手口」であった。そのため、私たちが日常的にそのようなメールを受け取ることは稀であった。ところが今、AIがその手間のかかる作業をすべて自動で行えるようになった。これにより、特定の人物を騙すための、完璧で自然な文章のメールを、瞬時に、かつ大量に作り出せるようになったのである。

このレポートでは、世界的に権威のあるハーバード大学の研究チームが実際に 101 名の人々を対象に行った実験結果を基に、「なぜAIは本物そっくりの大量の詐欺メールを送れるようになったのか」「なぜ私たちは騙されてしまうのか」を明らかにし、この新たな脅威から身を守るために私たちが取るべき対策を示したい。

2.データが語る衝撃の事実 ～AIは、ここまで人間を騙せる～

ハーバード大学の研究チームは、AI が作った詐欺メールが本当に危険なのかを確か

めるため、実際にメールを送って、どれくらいの人が騙されてしまうのかを調査した。その結果は、専門家の予想を超えるものであった。

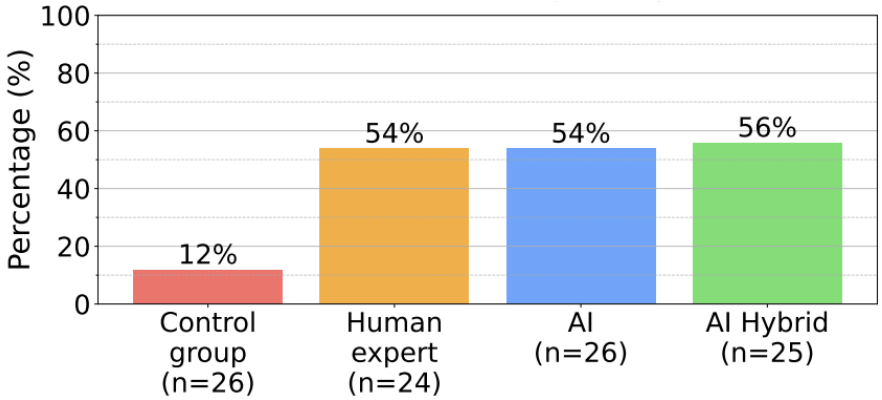
(1)成功率は「プロの人間」と全く同じ。2 人に 1 人が騙される現実

実験では、101 名の参加者をランダムに 4 つのグループに分け（Control group：26 名、Human expert：24 名、AI：26 名、AI Hybrid：25 名）、それぞれ違う種類のメールを送った。そして、「メールに書かれた危険なリンクを、ついクリックしてしまった人」の割合を比べた。

図表 1 をみると、一番左の「Control group（従来の迷惑メール）」では、クリックしてしまった人はわずか 12%、つまり約 8 人に 1 人である。ところが、真ん中の 2 つ、「Human expert（人間の専門家が作ったメール）」と「AI（AI が全自動で作ったメール）」では、どちらも 54%、つまり 2 人に 1 人がクリックしている。さらに右端の「AI Hybrid（AI が作り、人間が手直ししたメール）」では 56%と、わずかではあるが最も高い結果となった。

この結果が示すのは、AI が完全に自動で作ったメールと、詐欺のプロである人間が知恵を絞って作ったメールでは、騙されてしまった人の割合が全く同じ「54%」だったという点である。これは、もはや AI が人間と同じレベルで、人を巧みに信じ込ませる文章や状況を作り出せるようになったことを意味する。従来の迷惑メールとは比べ物にならない、「2 人に 1 人が騙されてしまう」ほどの危険なメールが、AI によって自動的に、大量に生み出される時代が始まったといえる。

図表 1　騙されてしまった人の割合



資料：Heiding, F. et al. "Evaluating Large Language Models' Capability to Launch Fully Automated Spear Phishing Campaigns: Validated on Human Subjects." November 2024

(2)あなたの SNS 投稿、AI は全て見ている

「なぜ AI はそれほど巧妙なメールが作れるのか？」と疑問に思うかもしれない。

その秘密は、AI の驚異的な「情報収集能力」にある。今回の実験で使われた AI は、ただ文章を作るだけではない。まず、ターゲットにされた人の名前を手掛かりに、イ

ンターネット上を検索する。そして、その人の SNS（Facebook や X など）の投稿、会社のホームページの自己紹介、過去のニュース記事といった、誰でも見られる公開情報を自動で集める。次に、集めた情報を分析する。「この人は何に興味があるのか」「どんな仕事をしているのか」「最近どんな活動をしていたのか」を読み取り、その人を騙すための「弱点」や「信じやすいポイント」をまとめた、詳細なプロフィールを作成するのである。

図表 2 は、AI の情報収集がどれほど成功したかを示している。

「OSINT 3（レベル 3）」は、その人を騙すのに十分な、正確で詳しい情報を集められたことを意味しており、全体の 88%がこのレベルに達した。「OSINT 2（レベル 2）」は、正しい人物の情報だが内容が限定的だったケースで 8%。「OSINT 1（レベル 1）」は、全くの別人の情報を集めてしまった失敗例で、わずか 4%であった。

つまり、AI は 100 人中 88 人について、その人を騙すのに十分なレベルの情報を、正確に集めることができたのである。

さらに注目すべきは、この表の右側に示された 2023 年との比較である。2023 年には「n/a（データなし）」となっており、当時はこのような AI による自動情報収集が実現できていなかった。また、メールの内容品質を示す「Content」の項目を見ると、さらに大きな変化が確認できる。2023 年には「Content 5（最高品質）」がわずか 25%だったのに対し、2024 年には 71%に急上昇している。わずか 1 年で、AI の能力が飛躍的に向上したことが明確に示されているのである。私たちが「友だちにだけ」と思って公開している情報も、AI にとっては格好の分析材料になっている。

図表 2　AI があなたの情報をどれだけ正確に集められるか

	AI-emails (2024)	AI-emails (2023)
OSINT 3	88%	n/a
OSINT 2	8%	n/a
OSINT 1	4%	n/a
Content 5	71%	25%
Content 4	25%	0%
Content 3	4%	0%
Content 2	0%	50%
Content 1	0%	25%

資料：図表 1 に同じ

(3)「これは自分宛ての特別なメールだ」と信じ込ませるワナ

最後に、実験に参加した人たちが「なぜ、そのメールを本物だと信じてしまったのか」を調べたところ、AI の注目すべき能力が明らかになった。

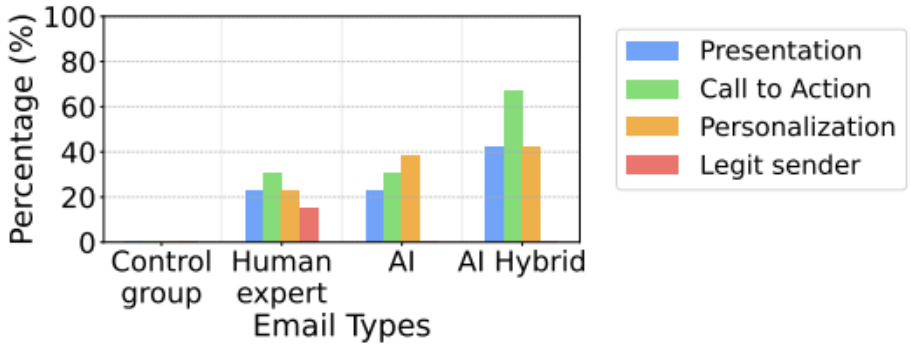
図表 3 は、メールを受け取った人が「なぜそのメールを信用してしまったのか」という理由を色分けして示している。青色は「Presentation （見た目や文章がしっかりしていた）」、緑色は「Call to Action（行動を促す呼びかけが魅力的だった）」、オレンジ色は「Personalization（自分個人に関係する内容だった）」、赤色は「Legit

sender（送信者が信頼できそうだった）」を表している。ここで注目すべきは、オレンジ色の「自分個人に関係する内容」の違いである。一番左の「Control group（従来の迷惑メール）」では、このオレンジ色がまったく見られない。つまり、誰も「自分に関係がある」とは感じなかったのである。ところが、「AI」と「AI Hybrid」のグループでは、オレンジ色の部分が大きく伸びており、約 40%の人が「自分個人に関係がある内容だったから信じた」と答えている。

さらに興味深いのは、「Human expert（人間の専門家）」のグループでは、オレンジ色が約 20%程度に留まっている点である。つまり、AI は人間の専門家の 2 倍も効果的に、「これは自分宛ての特別なメールだ」と相手に信じ込ませることに成功しているのである。

例えば、あなたが最近仕事で関わったプロジェクトの名前や、あなたの専門分野に触れた上で、「ぜひ、この件でご協力いただけませんか？」といったメールが届けば、疑うことは難しいだろう。AI は、「自分に関係のある話は、つい信じてしまう」という人間の心理を、効果的に利用してくるのである。

図表 3　メールを信じてしまった理由の違い



資料：図表 1 に同じ

3.AI詐欺メールから身を守るために、私たちが今すぐ始めるべきこと

第 2 章で見てきたように、AI による詐欺メールはもはや「自分は大丈夫」という自信が通用する相手ではない。AI は SNS などから私たちの個人情報を収集し、人間の専門家と同等、あるいはそれ以上に巧みな「あなただけの特別なメール」を作り出す。その結果、2 人に 1 人が騙されてしまうという現実がある。では、この新たな脅威に私たちはどう立ち向かえば良いのだろうか。特別な IT スキルは必要ない。重要なのは、メールに対する「考え方」と「接し方」を根本から変えることである。

AI は「自分に関係のある話は、つい信じてしまう」という人間の心理を突いてくる。だからこそ、以下の 5 つの行動を「新しい習慣」として身につけることが、自身と資産を守る盾となる。

第一に、「知っている相手」からのメールについても、まず疑うことである。これ

までは「知らない相手からのメールは怪しい」が常識だった。しかし、AIは仕事関係者や利用しているサービスの名前をかたって、極めて自然なメールを送ってくる。これからは「自分のことをよく知っているようなメールについても、まずは疑う」という逆転の発想が不可欠である。

第二に、メール内のリンクやボタンは「絶対にクリックしない」と心得ることである。これは最も重要で、効果的な防御策である。攻撃者は、対象者を本物そっくりの偽サイトに誘導し、情報を盗もうと待ち構えている。リンク先のサイトがどれだけ本物に見えても、ID、パスワード、クレジットカード番号、暗証番号などを安易に入力してはいけない。

第三に、確認は「公式ルート」から行うことである。メールの内容が気になった場合は、必ずそのサービスの公式アプリや、事前にブックマークしておいた公式サイトからログインして確認するのが望ましい。メールに添付されたリンクからではなく、いつも使っている公式サイトやアプリを開いて確認するという行動を徹底しよう。

第四に、SNSの公開設定を見直すことである。AIは、私たちが無意識に公開している情報をかき集めて攻撃の材料にする。勤務先、役職、最近訪れた場所、趣味など、個人が特定できるような情報の公開範囲は、見直す必要がある。

第五に、「多要素認証(MFA)」を設定することである。多要素認証とは、パスワードに加えて、スマートフォンに送られる認証コードや生体認証など、複数の方法で本人確認を行う仕組みのことだ。万が一、IDとパスワードが盗まれてしまった場合の「最後の砦」が多要素認証である。金融機関や主要なウェブサービスでは必ず設定してほしい。少し手間が増えると感じるかもしれないが、その手間があなたを金銭的な被害から守ってくれる。

図表 4　今すぐ実践したい 5 つの対処法

資料：筆者作成

AI による詐欺メールの脅威は、私たちの日常に深く静かに侵入してきている。しかし、必要以上に恐れることはない。重要なのは、「AI は私のことを、私が思っている以上に知っているかもしれない」という前提に立ち、メールとの付き合い方をアップデートすることである。今回紹介した「5 つの対処法」を実践するだけで、被害に遭うリスクを劇的に減らすことができる。技術の進化は、私たちの生活を豊かにする一方で、新たなリスクも生み出す。そのリスクを正しく理解し、賢く対処していくことこそが、「2 人に 1 人が騙される時代」を生き抜くための、最も確実な護身術なのである。