ScamGPT: GenAI and the Automation of Fraud

A Primer

Lana Swartz, Alice E. Marwick, and Kate Larson

DATA&
SOCIETY

ScamGPT: GenAI and the Automation of Fraud

by Lana Swartz, Alice E. Marwick, and Kate Larson

Contents

Executive Summary

The rapid rise of generative AI (GenAI) has sparked growing concern about its potential harms. While issues like plagiarism, data privacy, and the erosion of creative work have rightly drawn attention, GenAI is also fueling a surge in scams and misinformation at a global scale. Scamming has become a sophisticated, globalized industry — and like any industry, it is quickly integrating AI to increase scale, speed, and believability. Researchers are still in the early stages of exploring the connections between GenAI and scams, and many questions remain about the scale of the problem and what kinds of interventions are most effective. This primer maps what we currently know about GenAI's role in scams, the communities most at risk, and the broader economic and cultural shifts that are making people more willing to take risks — and more vulnerable to deception.

This primer is divided into three parts. In Part 1, we introduce GenAI technology, examine how it scales up existing scam tactics and enables new types of scams, and offer an overview of common scam types. Part 2 explores specific communities and groups that are disproportionately impacted and describes the larger sociotechnical landscape that renders people vulnerable to scams. Finally, Part 3 outlines the three major categories of scam response — personal, corporate, and regulatory — and evaluates their strengths and weaknesses. We conclude that:

- **Generative AI serves as a force multiplier for existing scam tactics.** From automating the generation of content and code, to selecting targets and compiling background research, GenAI tools are transforming the speed, sophistication, and personalization of scams — lowering the barrier to entry for criminal networks that previously lacked the technical expertise to conduct scams.
- **Scams exploit social vulnerabilities and thus require social solutions in addition to technical ones.** Declining trust in institutions, the inherent "scamminess" of the contemporary internet, and shifting economic norms blur the line between legitimate opportunities and scams, leaving even tech-savvy individuals vulnerable.
- **Marginalized communities that are already vulnerable are more likely to feel the impacts of AI-enhanced scams.** Those who have been pushed into informal systems of finance and labor fundamentally live in less secure environments, and thus are more exposed to risk. From gig workers relying on unregulated platforms to immigrants navigating unfamiliar systems, marginalized communities often operate without formal safety nets. AI-enhanced scams exploit their insecurities by mimicking legitimate opportunities, preying on financial desperation, or targeting linguistic and cultural connection points.
- **There is an urgent need for further research, particularly around effective scam prevention.** We must identify successful strategies to defend against AI-enhanced scams and design methods of public education that avoid paternalism or shaming. We need a clearer understanding of the AI-powered cybersecurity measures being proposed by corporations. Finally, more research is needed on AI-enhanced scams in non-Western contexts, where challenges and impacts may differ substantially.

Introduction

A woman receives an unsolicited text message on WhatsApp from a man with an alluring profile picture. She responds that he must have the wrong number, and as they chat, she finds him to be warm and patient. She checks his Instagram profile, which shows him on exotic vacations and driving fancy cars. For weeks, they talk about their families, careers, and travel experiences, even video chatting once or twice. He builds trust, slowly introducing her to an "investment opportunity" that helped him become financially independent. But behind the smiling profile photo and flirtatious messages is not a wealthy expat or savvy investor. It's a trafficked worker in a scam compound — possibly in Cambodia or Dubai — forced to carry out "pig butchering" scams for 12 to 18 hours a day. Workers in the compound use AI-powered translation tools to communicate fluently with targets around the world, AI-generated deepfakes to pose as good-looking romantic prospects, and large language models (LLMs) to tailor messages to each victim's interests and emotional triggers. The phone, scripts, and app interfaces are all provided by the criminal syndicate running the compound. If the captive worker fails to extract money, he faces beatings, starvation, or worse. He's as much a victim as the woman he's scamming — but neither of them knows it.[1]

Romance scams are not new, but deepfakes and LLMs are. Along with other forms of generative AI (GenAI), they are changing how scams are created, targeted, and executed. Scammers have quickly adopted GenAI technologies, whether mass market applications or jailbroken versions sold on the dark web.[2] International crime syndicates use these to supercharge existing scams, dramatically expanding both the number of targets and the profits they generate. Meanwhile, regulators and policymakers are scrambling to understand the extent of GenAI capabilities and protect the public from being scammed.

We define scams as a specific type of fraud where a victim is deceived into willingly sending money or sharing personal information. Scams can have a devastating impact on individuals and small businesses, and this impact is increasing.

In 2023, an estimated $43.6 billion was lost to consumer scams globally.[3] In the US, the Federal Trade Commission reports that Americans lost $10 billion to fraud in 2023, a 14 percent increase from the previous year.[4]

The increase in scam-related economic losses, as well as the blistering hype around GenAI tools, seems to predict a terrifying golden age of scams. The vast majority of financial tools are now digital services, networked together through obscure corporate vendors and interfaced with a mix of websites, apps, and email messages. And it is exactly these forms of digital media that GenAI are so adept at creating: quickly, cheaply, and hallmarked by generic corporate style. Given this accelerating exploitation, it is worth asking what forms of resistance might slow it, even if no single solution is likely to stop it completely. By examining how scammers are changing and accelerating their methods, we hope to show that our best defense is not only technical but social — a mix that will require a constellation of cultural shifts, corporate interventions, and effective legislation.

Part 1: How is generative AI increasing the scope and scale of scamming?

Scams have always been a fundamental part of technological innovation: whenever a new communication technology emerges — be it telegraph or email — people find ways to use it to deceive others for profit.[5] When technology changes, it creates new vulnerabilities that scammers and hackers can exploit. The people who maintain our everyday communicative and financial infrastructures are tasked with perpetually identifying ways to fend off attacks. GenAI is no exception. While business and industrial leaders tout its potential to boost efficiency and profits, and internet users tinker with chatbots and image generators, countless individuals who make their living organizing and executing scams are quietly working GenAI into their techniques, unleashing them worldwide.

AI-enhanced scams are not merely financial or technological crimes; they also exploit social vulnerabilities, from short-term susceptibilities, like when a person is traveling, to structural issues, as when a person's job is precarious. Therefore, they require social solutions in addition to technical ones. These solutions need to be flexible and responsive to evolving technology. As scammers refine or change their methods to outmaneuver public awareness, law enforcement, and other systems of protection, they are drawn into a constant cat-and-mouse game with the change and maintenance of sociotechnical systems.[6]

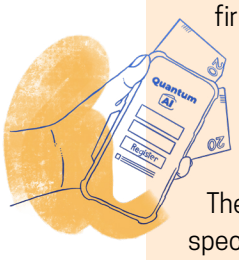
Scams can have a devastating impact on individuals and small businesses, and this impact is increasing.

For example, while initial coin offerings (ICOs) of cryptocurrencies were intended to be a radical form of crowdfunding and an innovative financial venture, the vast majority of ICOs turned out to be scams, preying on investors' optimism that the right coin purchased at the right time would make them rich.[7] Despite the claims of crypto-proponents, no technical fix to the blockchain or encrypted wallets can undo the effects of speculative hype and celebrity-endorsed schemes.

This primer highlights a broad range of scams, most of which already incorporate GenAI, and some that may not. We include the latter because understanding the dynamics of traditional scams is essential for grasping how generative AI expands their scale, automates their execution, and blurs the lines between old and new tactics. As AI becomes increasingly accessible, these scams are likely to evolve into AI-enabled threats. We distinguish between scams that target a wide range of potential victims (e.g., phishing emails sent to millions in the hopes that a small percentage will be tricked into divulging sensitive information) and those that hone in on a small number of high-net-worth targets and apply sophisticated techniques to extract a large financial payoff. The latter require more time and energy to pull off, often involving repeated psychological manipulation and sustained interaction with the target. The press often covers these scams, which involve large sums of money but ultimately have a small number of victims.[8] In contrast, scams that prioritize breadth over depth may involve much smaller sums of money or information and thus seem mundane, but their impact on individuals is far greater. While such scams rarely make the news, we see them as equally, if not more, important as their higher-profile equivalents.

A tale of two deepfakes

In 2023, deepfake videos of Elon Musk promoting a so-called cryptocurrency trading platform called Quantum AI circulated on Facebook, promising "high returns and minimal risk." The videos used AI-generated voice cloning dubbed over repurposed footage from Musk's recent podcast and conference appearances. One video, styled as a newscast, featured an Australian news anchor with an artificially generated voice.[9] These videos, which received tens of thousands of views, directed users to fraudulent websites that connected them to apparent "brokers" urging them to invest.[10] Those who "invested" funds were shown fake account balances but were ultimately unable to withdraw anything; their virtual wallets were empty.[11] By leveraging deepfake technology, scammers reached a vast audience with minimal effort. This type of AI-enabled scam, operating as a form of mass media, profits from a wide reach despite a low click-through rate.



In contrast, other AI-enhanced scams are hypertargeted. An employee at the British design firm Arup received an email that he initially thought was suspicious, directing him to a video call. When he joined the call, he was relieved to see the company's CEO and several other executives. Unbeknownst to him, scammers had used face-swapping technology and voice cloning to emulate higher-ups. Following their instructions, he transferred HK$200 million (approximately $25 million USD) across 15 transactions.[12] These laser-focused scams require extensive research to identify and manipulate specific targets, making them high-effort but potentially high-reward.

How has GenAI changed the game?

Scams are not a new phenomenon. But GenAI has the potential — and is already being used — to enhance existing scams and generate new ways to scam. The term "generative AI" broadly refers to artificial intelligence technologies that can create novel content by ingesting and learning from preexisting data. These include LLMs which generate text — like generative pretrained transformers (GPT) — as well as generative adversarial networks (GANs), which are commonly used to generate images and other media. Both LLMs and GANs are trained on vast quantities of data; notably, the industry is often silent about which data sources have been used to train GenAI models, how the data was acquired, and how models will be trained going forward.[13] This means that the public's everyday interactions on social media, online marketplaces, and dating apps — precisely the kinds of interactions scammers seek to mimic — are likely being used to train LLMs. Over the past few years, GenAI has exploded into public view and availability, offering users a level of sophistication and capability that previously belonged in the realm of science fiction.[14]

GenAI applications can produce a wide variety of content that is, at first glance, indistinguishable from human outputs. GPTs (e.g., ChatGPT, Bard) can produce grammatically and syntactically correct text and synthesize and present information in an accessible format. Image generators (e.g., DALL-E, Stable Diffusion) can create images that mimic photographs or graphics based on input prompts. Video generators (e.g., Meta Movie Gen, Runway) can create animated or seemingly live-action video based on text prompts or still photos.[15] Falsified images or videos of actual people are known as deepfakes and are already being used to facilitate and are already being used to facilitate fraud.[16] Voice cloning apps can use a few minutes of audio to create a passable reproduction of a person's voice, which can then be manipulated to verbalize written text.[17] Together, these tools can produce human-like writing, credible text translation, functional programming code, photos and videos of scenes that never happened, and audio of recognizable voices saying things their owners never said.

Generative AI exacerbates existing scam tactics by expanding scale, accelerating speed, and increasing efficiency. The technology serves as an intensifier and force multiplier for scams, broadening their reach and impact. In crypto scams, for example, scammers can use GenAI to create a slew of fake investment sites, generate scam tokens and trading bots, and produce deepfake videos to suggest celebrity involvement, all spread by bots on social media networks.[18]

Generative AI exacerbates existing scam tactics by expanding scale, accelerating speed, and increasing efficiency.

Commercial GPTs are generally packaged with some safeguards against using them for fraud (although these can be hacked[19]); grey market AI products like WormGPT and FraudGPT have no such protections and can be used to plan and execute an entire phishing scam, from writing code for a fake login page to drafting content for phishing emails that link to it.[20] GenAI can quickly create vast amounts of content (e.g., hundreds of websites, images, videos, and social media posts) that reinforce each other, which makes scams seem more legitimate and reach more viewers with lower effort.[21] Deepfake videos and voice cloning have been shown to make scams more convincing.[22] AI can also be used to profile targets and conduct reconnaissance for social engineering attacks.[23]

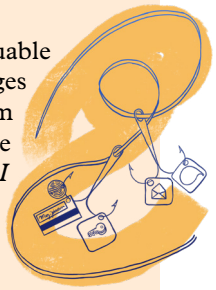
Generative AI lowers barriers to entry for scammers. It can translate text smoothly, enabling scammers to target speakers of other languages.[24] Scammers can purchase detailed guides on exploiting GenAI to run so-called AI pimping scams on Instagram,[25] while networks of online vendors sell jailbroken LLMs and prepackaged AI-generated content to make it easy to set up social engineering scams. Against this backdrop, those willing to make scams a business have created a sophisticated, global criminal shadow technology industry. In Southeast Asia, for example, criminal networks that did not previously have the sophistication or capacity to run profitable online scams have moved into scamming through the use of GenAI tools.[26] According to a 2024 study by the Organized Crime, Corruption, and Reporting project, large professionalized call centers in Israel, Georgia, Spain, and Bulgaria employed hundreds of workers to carry out investment scams, defrauding more than 30,000 people worldwide and stealing nearly $300 million over three years.[27] The United Nations reports that "cyber-enabled fraud operations have taken on industrial proportions, with independent and scattered fraud gangs being replaced by larger, consolidated criminal groups often operating under the guise of industrial and science and technology parks as well as casinos and hotels."[28]

Generative AI has also given rise to new types of scams. For example, voice cloning has enabled virtual kidnapping scams, where targets receive a phone call that sounds like a family member or friend who is being held at gunpoint until a ransom is paid via wire transfer. The voice-cloned loved one, meanwhile, has no idea that they are being used as a pawn to extract money from someone close to them. The widespread industrialization of scamming operations, combined with automated processes and accessible AI voice cloning technology, have created an environment where it is profitable to target victims even for relatively small sums of money. What once would have been a complex and laborious scam can now be deployed on a mass scale. One victim told *The New Yorker* that the most upsetting part of finding out about her own virtual kidnapping was the relatively small sum ($750) that was stolen from her son, reflecting: "I still can't believe that's all I was worth."[29]

AI-enhanced scams are dynamic, as scammers use evolving tools and existing exploitation strategies which adapt to and reshape each other in unpredictable ways.

Common types of AI-enabled scams[30]

AI amplifies existing scams



- **Social engineering** scams exploit victims' trust to gain money or valuable information. These include phishing scams, in which emails, text messages (also known as *smishing*), or voice calls (*vishing*) mimic those from legitimate sources, like a bank or government office, and entice people to divulge personal information. *While phishing is well-known, GenAI applications like GPTs can write persuasive text that is grammatically correct, tonally appropriate, and personalized to the recipient, making it much harder to identify as a scam.*

- **Business email compromise** is a type of scam where a legitimate-seeming email from a business account is used to persuade an employee to share sensitive information or authorize a purchase, or to impersonate the company in order to scam customers or other recipients. This could appear as a fraudulent invoice, a bogus assignment from a supervisor, or a fake IT request for login details, among others. *AI makes it possible to write more convincing emails by including specific company details or even mimicking the writing style of an executive.*[31]

- **Disinformation** and **misinformation** campaigns strategically spread false content for profit, ideological reasons, or harm.[32] *GenAI makes it easier to craft convincing content (including photos and videos), lowers the cost of spreading false information, and works at a much larger scale.*[33] *Even when it is not explicitly used to spread misinformation,* ***AI slop*** *(low-quality AI-generated content) can flood social media sites, making it difficult for users to know what is real or who to trust.*[34]

AI lowers barriers to entry

- **Financial grooming** or **pig butchering** scams describe a version of catfishing where a scammer (or team of scammers) develops an online romantic relationship with the target to gain their trust and then begins extracting money, often through investments in cryptocurrencies.[35] (The colorful term "pig butchering" refers to the victim being "fattened up" over time.) Eventually, the invested money is stolen from the victim's wallet and the scammer disappears. Sometimes the target is asked to pay fees or taxes to retrieve their money before they discover that it is irretrievable.[36] *AI deepfakes let perpetrators convincingly appear as ideal romantic partners and enable near-flawless translation, making it easier for "fraud factories" to use human trafficking and forced labor to target people in the West.*[37]

- **Harpoon whaling** scams target a high-net-worth individual or corporate executive, using research to gain their trust and convince them to invest large sums of money in fraudulent ventures.[38] Similarly, **spear phishing** targets an employee who has the authority to make large financial transactions. *These sophisticated schemes can use AI to identify possible targets, and later use voice cloning and live video deepfakes to convince the target to invest funds or make wire transfers that are ultimately irretrievable.*

AI enables new types of scams

- **Sextortion** involves threatening targets with the release of deepfake pornographic photos or videos unless they pay to bury them.
- **Voice cloning** scams use AI to create a realistic replica of a person's voice from only a short audio clip.[39] This technology has engendered new types of scams, like virtual kidnapping and false pleas for money from relatives. Scammers use this to enhance nearly every other type of scam we discuss here with an audio component, such as harpoon whaling and vishing. Scam targets are far more likely to trust familiar voices, whether family members or well-known public figures.[40]
- **Deepfake videos** fuel a wide range of scam types. These videos are produced or doctored using GenAI, and used for social engineering, financial grooming, sextortion, and more. Some scammers also have access to realistic face cloning during live video, as demonstrated by high-profile harpoon whaling cases.

Part 2: Who is vulnerable to scams and why?

When Lilah Jones and her husband found their dream home in the Chicago area shortly after their wedding, she began navigating the home-buying process while simultaneously coordinating the logistics of moving her mother into the new home with her family. Jones, a sales professional at Google, was no stranger to fast-moving projects, and she stayed in close contact with her real estate agent, mortgage company, and the title company to confirm everything was going to plan. When she received an email from the title company three days before closing with instructions for how to wire the down payment and closing costs, she called the company to verify its legitimacy, but did not hear back by the end of the day. Not wanting to delay the closing, she followed the steps in the email carefully.

When Jones and her husband arrived at the closing, a representative from the title company informed them that the money had never arrived. Jones pulled up the instructions she had been sent, from a very closely spoofed email address, so similar to an actual employee's that it was extremely difficult to spot. Courtesy of a highly targeted business email compromise scheme, the couple had wired $130,000 to a scammer. Their carefully saved down payment was gone.[41]

Fraud prevention awareness campaigns often focus on the elderly and people with limited technical skills; it is not a coincidence that a widely used fraud prevention hotline in the US is staffed by the American Association of Retired Persons (AARP).[42] But as Jones discovered, nearly anyone can be the victim of a scam. The factors that render us susceptible to scams are deeply embedded in our cultural, technological, and economic landscapes, cutting across demographic lines and leaving many at risk who might think themselves unlikely targets.

There is still very little research on who is most vulnerable to AI-enabled scams — partly because these tactics are so new, and partly because they build on the same tactics as older forms of fraud. However, the patterns of vulnerability remain the same.

Who is impacted by scams?

While older adults are certainly susceptible to scams, the FTC reports that younger adults (ages 18–59) are actually more likely to have lost money to fraud, based on consumer reporting. Age plays a role in the type and medium of scamming; younger adults were more likely to lose money to online shopping fraud (often through social media ads), bogus cryptocurrency investment scams, and job scams. Consumers over 60, who reported losing more money overall to scams, were more likely to be the victims of tech support, prize, sweepstakes, and lottery scams. They were also far more likely to have been scammed by phone.[43] While scammers might exploit seniors' lack of technological savvy, tendency to trust strangers, isolation, and higher financial resources,[44] younger adults may be at risk due to the ease with which scams can appear legitimate on social media and video games,[45] as well as their need to find employment.[46] The impact of scams is escalating among teens as well; a 2024 study by the online ID verification company Social Catfish, which analyzed scam reporting data, found that teens and children experienced a nearly 2,500 percent increase in money lost to scams from 2017 to 2022, more than any other age group.[47] Notably, the impact of scams is likely being underreported across all age groups, as many victims avoid reporting due to shame or embarrassment.[48]

Similarly, gig workers are disproportionately impacted by scams due to the precarious nature of their work and their reliance on tech platforms with inconsistent governance. Care workers who rely on online labor platforms have a complicated relationship with these platforms when it comes to protection from scams; some feel that platforms do a good job flagging and removing scammers, while others believe that platforms make them more vulnerable by oversharing personal information and ineffectively vetting potential clients. These workers seek out each other's support in online forums to identify and avoid scams, an effort which obligates them to engage in unpaid labor for each other's benefit.[49]

The factors that render us susceptible to scams are deeply embedded in our cultural, technological, and economic landscapes, cutting across demographic lines and leaving many at risk who might think themselves unlikely targets.

Individuals relying on informal systems and workarounds constantly navigate a less secure sociotechnical environment. Economically disadvantaged households in the US struggle to protect their personal information online, since many rely on public Wi-Fi networks or shared computers to access the internet.[50] Likewise, unbanked or underbanked individuals

lack access to protective measures like credit card charge reversals or bank fraud prevention, leaving them with fewer options for recourse if they fall victim to a scam.[51]

Immigrants are so frequently targeted that US Citizenship and Immigration Services publishes a list of common immigration scams. These range from broad scams affecting all immigrants — such as government impersonation and visa lottery fraud — to schemes targeting particular communities, like phishing scams soliciting fees from Ukrainian immigrants or identity theft scams preying on recent arrivals from Afghanistan.[52]

Minoritized groups are targeted in spaces, both physical and digital, where they gather and transact. In 2023, American users of the Chinese app WeChat were flooded with advertisements to "invest" in household goods and electronics at inflated prices in exchange for a future payout, a scam that preyed on users' sense of community and shared language.[53] Chinese nationals worldwide have found themselves in the crosshairs of a police impersonation scam, where scammers threaten them with extradition to China for fabricated crimes and extort them for "bail" money.[54] Vietnamese, Taiwanese, Chinese, and Indian diasporic communities in the US have been impacted by misinformation spread in their native languages through preferred communication platforms like WhatsApp and LINE. These harms go beyond individual or even community-level experiences, straining family relationships, religious communities, and other vital social structures.[55]

AI-enhanced scams harm not only their targets, but in many cases, the victims of human trafficking who are forced to conduct them. A *WIRED* review of law enforcement proceedings found that, in the last five years, Chinese-linked criminal organizations trafficked 200,000 people from at least 60 countries to Myanmar, Cambodia, Laos, and elsewhere. Once trafficked, *WIRED* reports, victims are often isolated from the outside world, working under duress in guarded facilities where they are forced to make fraudulent calls, create fake profiles, and chat with victims. These trafficking operations are spreading to other regions, including the Middle East, Eastern Europe, Latin America, and West Africa.[56] AI-enabled scams are not only a new threat to online safety but an integral part of a global trafficking crisis, where vulnerable individuals are exploited both digitally and physically.

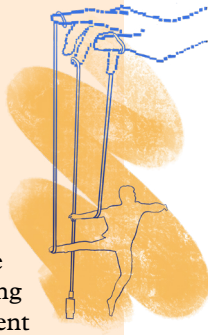
What makes us vulnerable to scams?

In the US, declining trust in institutions, paired with a high level of uncertainty about the future, has given rise (especially among young adults) to a "YOLO (you only live once) economy." In this economy, high-risk activities like investing in meme stocks and cryptocurrency, sports betting, entrepreneurship in multi-level marketing (MLM) companies, passive income-generating schemes, and other ventures that offer quick but uncertain financial gains thrive.[57] Some of these schemes are scams, and some are just high risk. However, all of them open participants up to more overt scams, because they sit at or beyond the margins of regulation. These kinds of gray area financial opportunities are becoming more commonplace, and, as they continue, more people will be exposed to a fuller range of scams and scam-like situations.

Economic uncertainty is not the only factor that makes ordinary people vulnerable to scams. Our relationships, commerce, and workplaces are increasingly networked and global, rendering information more exposed and multiplying potential points of contact between scammers and targets. Most professions require workers to use some level of technology in the workplace, and many are increasingly forced to use AI applications in the course of their jobs.[58] Within these infrastructures, employees at all levels can be lured into sending money or sharing sensitive corporate information through business email compromise schemes.[59] Romance scammers browse information on social media profiles, dating sites, and search engines to compile details on their targets, specifically targeting widows and divorced women in the hopes of exploiting their emotional vulnerability.[60] Voice cloning scams rely on knowledge about which of the target's loved ones they will be anxious to help.[61] Moreover, the decreasing quality of online tools, termed "enshittification"[62] or "platform decay," makes it harder for victims to find accurate information; for example, scammers have managed to rank fake helplines at the top of Google search results for major domestic airlines' customer service.[63] In this environment, individuals' ability to discriminate between legitimate and illegitimate uses of GenAI is a crucial but vastly underappreciated new professional skill.

Anatomy of a scam

Victims, and their loved ones, are often left wondering how they could have fallen for a scam. But scammers' tactics can be powerful and persuasive. Recognizing these tactics and how scams are conducted can aid in identifying effective interventions and designing public awareness campaigns that destigmatize being scammed.



- Scammers lure victims in with a **compelling hook**, whether through a seemingly trustworthy message or a fabricated emergency. Unlike traditional scams, AI-generated messages lack obvious red flags like spelling errors, making them even harder to spot.

- They then move into **trust engineering**, building a false sense of trust with the victim through personalized communication and deceptive cues that make the scam appear legitimate. Chatbots, AI-generated photos, and face swapping enable ongoing conversations that feel personal and credible. Fraudulent websites, complete with fake testimonials and professional design, make scams appear authentic. Personalized messages lower skepticism, using contract language, technical jargon, and professional-looking invoices to maintain a facade of legitimacy.

- Once engaged, victims are **pressured** into providing money or sensitive information, often believing they are making a rational decision. Scammers use emotional manipulation techniques to manufacture urgency — through fabricated crises or lucrative, time-sensitive opportunities — to push victims into immediate action. AI-generated content can create a heightened sense of imperative, making victims feel as though delaying their response would result in dire consequences.

- Finally, scammers employ a well-crafted **exit strategy** — often shifting blame onto the victim and discouraging them from seeking help, or simply vanishing by shutting down a string of fake websites and/or accounts. Scammers use GenAI to craft convincing excuses or false business closure stories, generate realistic fake authorities to discourage victims from reporting, and quickly shut down websites or accounts and delete digital evidence.

Some criminologists argue that the internet itself is a criminogenic environment — a setting that fosters or encourages criminal behavior — and that AI is a criminogenic technology, facilitating and augmenting crime.[64] Before AI tools were widely accessible to scammers, internet users with higher technological skills were better able to detect and avoid scams.[65] However, the accelerating pace of technological innovation and the broadening availability of GenAI create an environment in which scams are easier to conduct and increasingly difficult for individuals to detect. Not only are deepfake videos and cloned voices convincing, but the general public is largely unaware of how sophisticated these technologies have become, or that they are being used in financial scams.[66]

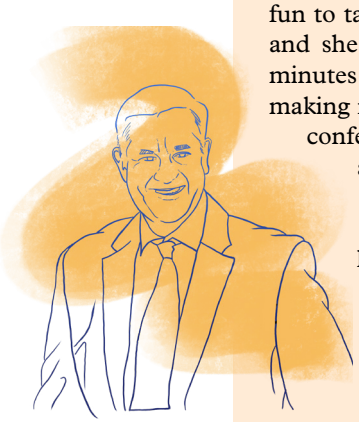
Scams can be flashy, but most are deeply mundane

AI-enhanced scams in the headlines often involve millions of dollars or complex deepfake technology. In reality, most scams are much more banal. These fictional scenarios illustrate some of the quieter, more common AI-enhanced scams.

Maria, 28, marketing professional

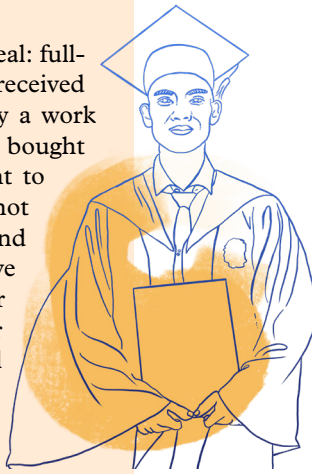
Maria received an email from PayPal that looked and felt authentic, informing her that someone was trying to access her account. When she clicked on the "Change your password" button, it took her to a scam site that perfectly mimicked PayPal, where she entered her login information. Scammers stole her password, logged into her account, and accessed her bank and credit card information. Maria is working with her bank, credit card vendor, and PayPal to change her account numbers, refute charges, and attempt to recover the roughly $800 that was stolen, but she does not know if or when she'll get it back. *This scam was powered by* ***AI-generated phishing emails*** *— LLMs trained on official PayPal communications were used to craft a message with flawless grammar, tone, and formatting, making it nearly indistinguishable from a real PayPal notification.*

Darryl, 55, sales manager



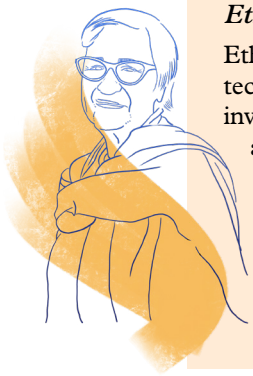
After his divorce was finalized, Darryl met a woman named Sarah on a dating site. She was fun to talk to, attractive, and seemed genuinely interested in him. They messaged regularly, and she eventually suggested they move to video chat. The call was brief — just a few minutes — but it reassured Darryl that she was real. A few days later, Sarah told him she was making money on the side by investing in cryptocurrencies and invited him to try it. Darryl confessed that he didn't know much about crypto, so she sent him a link to a trading app and even gave him tips. His initial investments made money, so it felt safe. But when he tried to withdraw his earnings, the app asked for $250 in taxes and froze his funds even after he paid. When he asked Sarah for help, she disappeared. He realized that he had been the target of a romance scam, and both his modest investments and the fees he paid were unrecoverable. *This scam used* ***AI-generated deepfake video****, allowing scammers to create a live, real-time video call with a fake persona that mimicked human facial expressions and speech. The video gave Darryl false confidence that he was speaking to a real person, making the emotional connection feel authentic.*

Ryan, 23, recent college graduate



Ryan received a personalized email inviting them to apply for a job that seemed ideal: full-time entry-level remote work for a well-known consulting firm. They applied and received an offer quickly. Even better, the company sent them a check for $3,500 to buy a work laptop, stipulating that Ryan needed to wire back any remaining money. Ryan bought a computer for $2,000 and planned to return the extra $1,500. When they went to cash the check, their bank informed them that it was counterfeit. When Ryan did not respond to the recruiter's requests for $1,500, the recruiter demanded that Ryan send the full $3,500. The recruiter became increasingly aggressive, threatening to involve law enforcement and destroy Ryan's online reputation. Ultimately, the scammer gave up. Ryan was able to return the computer for a refund and resumed their job search, shaken by how legitimate the job posting seemed, the level of personal detail the scammers used to craft the offer, and how plausible the threats seemed to be. *This scam was enhanced by* ***AI-enabled personalization*** *— scammers used AI tools to scrape data from Ryan's social media and online presence, generating a job offer tailored to their interests, education, and goals, making it feel eerily real.*

Ethel, 80, retiree



Ethel recently told her daughter she was frustrated by the high cost of Best Buy's Geek Squad tech support. When her daughter checked Ethel's email, she found dozens of fraudulent invoices from "Best Buy," which Ethel had diligently paid. Ethel had always been cautious about scams, looking for grammar errors and broken logos — but these emails looked legitimate. Her daughter has been unable to recover the $5,000 Ethel paid for fake tech support and Ethel is now leery of checking her email, even though she uses it to keep in touch with her friends and grandchildren. *Scammers used* ***AI-assisted branding and design tools*** *to create realistic invoices and logos, helping them bypass traditional red flags like poor formatting or typos that Ethel had learned to watch for.*

Part 3: How do we defend against AI-enhanced scams?

Defending against scams can feel like a game of whack-a-mole; scamming evolves in tandem with technological development. Even so, it is still worth asking how we can mitigate the damage caused by scamming, particularly to vulnerable individuals and communities. Practitioners have proposed a range of solutions to reduce the impact of AI-enhanced scams. They fall into three general categories: 1) individual solutions (e.g., public awareness raising, personal vigilance); 2) corporate accountability (e.g., cybersecurity, increased liability for platforms and AI developers); and 3) governmental and legal responses (e.g., improved regulatory frameworks, updated tools for prosecutors). We believe that the most effective consumer protection efforts will involve a combination of these approaches.[67] However, research on detecting, defending against, and preventing AI-enhanced scams is still in its early stages. There is little data available on the impact or effectiveness of specific interventions, and an urgent need to identify the most feasible and scalable solutions.

Defending against scams can feel like a game of whack-a-mole; scamming evolves in tandem with technological development.

Individual solutions

Raising public awareness has long been a key element of scam defense. Governmental bodies and community organizations work to educate the public about how to recognize scams, avoid becoming a victim, report fraudulent activity, and share information within their families and local communities.[68] Some scholars recommend that cybersecurity education start as young as elementary school,[69] while the FTC offers resources to help teens avoid scams as they take their first independent steps on digital platforms.[70]

In response to the rise of voice cloning scams, consumer protection agencies and journalists often recommend that families create a private family "password" or phrase that can be used to verify a voice on the phone purporting to be a loved one.[71] Reporters urge readers to be deeply skeptical; to understand that photos, voices, and video can be realistically faked; to avoid revealing personal information over the phone; and to report any suspicious activity or scamming to the authorities.[72]

This type of public education has value, but to be effective, it must be delivered without blaming or stigmatizing victims or sounding paternalistic. One pioneering example of community engagement around scam prevention is *Performances to Reduce Online Scams*, a research and theater project at Virginia Tech University, which culminated in a series of interactive theater performances titled "This is Not a Scam." Drawing from interviews with individuals targeted by scammers, the show weaves dramatized stories with questions posed to the audience, inviting attendees to contribute their own experiences to a non-hierarchical dialogue between performers and audiences. Researchers found that victims often fail to report scams due to shame and embarrassment. Rather than lecturing or questioning ("Why didn't you suspect anything?" "How could you have done that?"), victims require compassion and empathy from family members.[73] A similar project in Malaysia created an interactive role-playing game. Pretending to be linguistic experts working to bring down scammers, players learn to identify the language and techniques they use.[74] Information about scam prevention can also be disseminated through popular media; organizations like the USC Annenberg Norman Lear Center's Hollywood, Health & Society project consults on television shows to make health and safety storylines more accurate.[75]

There is very little evidence pointing to the efficacy of any of these solutions. Moreover, relying on individual efforts is a form of *responsibilization* — that is, placing blame on victims, rather than addressing root causes or holding financial or technological companies responsible.

Corporate solutions

Efforts to improve public awareness are inherently limited, especially as GenAI makes illicit activity more difficult to detect. There is little public consensus on the broader responsibility of companies that develop AI technologies or host digital platforms used by scammers, let alone holding them responsible for safeguards and fraud prevention tools. Moreover, there are open questions around which tools and techniques are effective at deterring scammers; who should develop, own, and maintain these technologies; and how to mitigate potential negative impacts on users, especially vulnerable communities and individuals.

In these conversations, AI is frequently touted as a way to mitigate scams. For example, several startups are developing plugins for video chat platforms like Zoom, which can detect whether the person you are talking to is using deepfake technology.[76] Machine learning can be used to analyze historical datasets to identify patterns of fraudulent behavior, which can be leveraged in real time to flag possible scams.[77] AI can potentially identify possible victims of hypertargeted scams,[78] scam-like messaging,[79] and e-commerce fraud.[80] However, most of these techniques are still in the early stages, and many focus on low-hanging fruit — such as detecting credit card fraud — which cannot prevent scams that use crypto or wire transfer.[81] Moreover, they do not address the complex social vulnerabilities that make people susceptible to scams. Many victims of romance scams, for example, do not even admit that they have been scammed, sometimes continuing to send money to scammers even after being warned by friends or banks.[82] In addition, these techniques are only as good as the AI tools themselves; AI models are error-prone and, in fact, LLMs themselves can easily be scammed.[83] It is important to keep humans in the loop to prevent the potential harms of AI, such as hallucination and reproducing biases from training data.[84]

Other technologies being developed or implemented to prevent scams include industry-specific tools like scam-proof ID verification for banks.[85] Researchers and policymakers have called for new and accessible techniques for watermarking voices, images, and videos, and identifying deepfakes and AI-generated media.[86] Finally, there are requests for the companies developing GenAI technologies to be more transparent about how they are acquiring data and training their models, and to disclose the kinds of guardrails they are creating to prevent misuse by scammers.[87] These efforts, however, can be a double-edged sword: transparency about scam prevention procedures can help scammers figure out how to skirt guardrails.

Any discussion of corporate cybersecurity and fraud prevention efforts must consider the knock-on effects of these technologies on user welfare. Traditional fraud prevention techniques already cause unintended harm, especially to vulnerable populations.[88] The rise of AI-enhanced scams will likely be the impetus for companies across a wide swath of industries — financial, e-commerce, fintech, social media, and so forth — to rethink and update their scam prevention procedures. These rules will shape systems used by millions of people every day. While some of them may indeed limit the impact of scams, without careful consideration of their impact, they will also exacerbate existing harms to vulnerable communities.[89] To balance these varied risks, companies must implement frameworks that explicitly weigh fraud prevention against accessibility impacts, applying stricter scrutiny to measures that could disproportionately burden marginalized users. Further, researchers and

policymakers must evaluate any proposed security protocols both for effectiveness and for their disparate impacts across populations, prioritizing solutions that maintain access while providing protection.

Regulatory, legal, and policy solutions

Around the globe, legislators, policymakers, and regulators are familiarizing themselves with AI-enhanced scams to develop new frameworks for pursuing, stopping, and prosecuting scammers, targeting the criminal networks engaged in scams, and holding companies accountable for boosting cybersecurity. Effective scam prevention requires greater collaboration among all stakeholders.

Prosecutors in the US and United Kingdom are actively discussing new legal frameworks for crimes involving deepfakes, including increasing the offense level of deepfake-based wire fraud (US)[90] and making creative use of the Fraud Act 2006 (UK).[91] In 2024, the US Federal Trade Commission finalized a rule banning business and government impersonation fraud, and proposed an update to that rule that would make it unlawful for a company "to provide goods or services that they know or have reason to know is being used to harm consumers through impersonation." If passed, the proposed rule changes would hold corporations accountable for deepfakes created or shared on their platforms.[92] In Southeast Asia, the UN Office on Drugs and Crime has called for region-wide intelligence sharing and collaboration and up-to-date legislative and regulatory frameworks, with enforcement mechanisms that can address the professionalization of scamming and the vast ecosystem of scams.[93]

Effective scam prevention requires greater collaboration among all stakeholders.

Effective enforcement against AI-enhanced scams requires more than just updated frameworks; we need robust reporting structures that capture who is being targeted, how specific scams operate, and where interventions are most urgent. Right now, reporting mechanisms are fragmented, underutilized, or not designed to capture the nuances of AI-enabled fraud. Regulators and prosecutors need high-quality data to track emerging threats, understand demographic impacts, and build stronger cases against criminal enterprises. Without them, enforcement will remain reactive. Future research should investigate the feasibility of creating and maintaining these reporting structures.

Scams in popular culture

The UK-based phone company O2 has "seeded" a particular number into popular phone lists often used for scamming. If a scammer dials that number, they'll hear the cheerful voice of an older woman named Daisy, who prattles on at length about her new kitten as she struggles to follow the scammer's instructions. A video shows scammers shouting at her in frustration as she blithely chatters away. Daisy is not, in fact, a woman or a cat owner; she's a scam-baiting AI program designed to engage scam callers for up to 40 minutes at a time.[94]

Scams and counter-scams are a growing part of popular culture. High-profile scammer Anna Delvey has been the subject of a Netflix miniseries and BBC podcast, as well as a contestant on the televised reality competition Dancing with the Stars.[95] The r/scams subreddit community, where users share cautionary tales, encourage each other to report fraudulent activity, and boast about thwarting scammers, currently has over 950,000 members.[96] Documentaries about lying entrepreneurs, scheming influencers, and meme stocks flood streaming services, podcast platforms, and movie theaters. As scams proliferate, so does their mystique, along with public interest in both scams and counter-scam tactics. Popular culture can be a useful site to raise awareness about how scams evolve and how to prevent them, but this depends on accurate media representations, rather than glamorized or stereotypical portrayals of scams, scammers, and victims.

What don't we know? What do we need to know?

Further research into how GenAI enhances and intensifies scams, and how we can mitigate its negative impacts, could strengthen consumer protection efforts, support the development of better regulatory and policy frameworks, and facilitate collaboration among multiple stakeholders in combating scams.

Some open questions for research include **foundational questions** like:

- What kind of social, cultural, economic, and technological conditions make us vulnerable to scams, including AI-enhanced scams?
- What long-term social impacts will AI-enhanced scams have on trust and online behavior?
- What can the history of the relationship between scams and technological change teach us about AI-enhanced scams?
- What do we know about AI-enhanced scams outside of Western democracies? Who is impacted by scams in these understudied contexts? How will mitigation efforts and awareness-raising campaigns need to evolve to be successful in these contexts?

Open questions also include those about how to **mitigate the threat of GenAI scams**, such as:

- What kinds of scam prevention efforts are the most effective? In which contexts are they effective? Are those efforts scalable?
- Can we reduce people's vulnerability to scams by addressing larger systemic issues?
- What skills and competencies can help people resist AI-enhanced scams in their personal and professional lives?
- How can cybersecurity tools be designed to minimize or eliminate negative impacts for vulnerable users?
- What design interventions might be effective? For example, given that frictionless payment systems facilitate scamming, what are potentially effective ways to reintroduce some level of friction to protect against scams?
- What legal and state measures can be taken against the criminal enterprises responsible for scams, especially those engaging in human trafficking? What can we learn from research on human trafficking to reduce the vulnerability of people forced to engage in these schemes?

And questions about **accountability and responsibility for reducing the harm of AI-enhanced scam**, such as:

- What level of accountability do corporations bear for scams conducted via their apps or digital ecosystems? Can and should regulators compel the companies whose products facilitate large amounts of scamming activity to take action against scammers?
- Which party (e.g., bank, individual, merchant) is responsible for bearing the financial impact of a scam? How might shifting this responsibility impact fraud prevention, consumer protection, and institutional accountability? What alternative models could distribute the financial burden more equitably?
- What kinds of policy and regulatory interventions might be the most effective and viable to prevent AI-enhanced scams?

Conclusion

The question of how to solve AI-enhanced scams is not just a question about GenAI, or even cybersecurity. In order to develop effective solutions to protect individuals against the rise in scamming, we need to understand the social and cultural vulnerabilities that render us increasingly susceptible to scams. These vulnerabilities are rooted in economic precarity, a culture of normalized risk, and unfettered technological expansion with insufficient regulation. The lines between an obvious scam and a legitimate, albeit questionable, business model are growing increasingly blurred. When everything seems risky, then, in a way, nothing seems risky. Technological innovation alone cannot solve for an environment in which scamming has begun to feel like the norm.

It is worth noting that GenAI is not the only technology that has enabled scamming at its current scale. It is merely one of an array of infrastructures that, taken together, permit frictionless scamming. These include standardized text messaging formats, cryptocurrencies, global social media networks, and others, all of which have legitimate use cases, and can be misused and exploited for financial gain or to spread misinformation.

How can we reduce the harm caused by these scams to the point where their effects are minimized and containable?

The suppression of spam email is often hailed as one of the internet's early success stories. Though it once threatened to make email unusable, spam was largely tamed through the development of increasingly sophisticated email filters, regulation like the CAN-SPAM act, and email users becoming more adept at recognizing and managing unsolicited messages.[97] Today, while some spam emails still sneak through, the vast majority is relegated to junk folders, nearly invisible.[98] This success did not come from eliminating spam altogether, but from mitigating its impact so effectively that it became a manageable nuisance rather than a disruptive force, through a mix of technical advancements, regulation, and education. As we turn our attention to the growing threat of AI-enhanced scams, the question becomes: What would a comparable success story look like? How can we reduce the harm caused by these scams to the point where their effects are minimized and containable?

AI-powered scams present a far more complex regulatory, technical, and social problem than spam, which will likely require more sophisticated and coordinated solutions across multiple domains. AI scams are always evolving and take place over many different kinds of channels. Systemic, institutionalized risks — such as economic precarity and the high-risk online economy it fosters, including legal sports betting, meme coins, and MLMs — are deeply entrenched and unlikely to disappear anytime soon

Given this, our task will be to create systems and tools that protect the most vulnerable populations, designing a future in which AI-enhanced scams cause the least possible harm. In this context, combating AI-enhanced scams is not just an issue of consumer protection; it is also a question of work skills, and more broadly, skills for navigating the digital age. From improving public discernment of artificially generated content to fostering resilience against manipulation, addressing these risks will require a multifaceted approach that will likely include education, judicious technological innovation, and policy reforms.

Acknowledgments

Thank you to Charley Johnson and the members of the Public Technology Leadership Coalition for their helpful feedback early in the process. Thanks to all the members of the Data & Society research team for feedback on our first draft. Thanks to our expert reviewers, Claire Maiers and Josephine Wolff, for their careful reading and commentary. Thank you to the Research Management team, especially Kiara Childs, Patrick Davison, and Siera Dissmore, for helping this project come to life and expertly editing it. Thank you also to Surbhi Chawla for layout; Gloria Mendoza for illustration; Chris Redwood, Alessa Erawan, Eryn Loeb, and Sona Rai for production and promotion; and Iretiolu Akinrinade for network engagement.

Suggested Citation

L. Swartz, A. Marwick, and K. Larson, ScamGPT: GenAI and the Automation of Fraud, Data & Society Research Institute, May 2025. DOI: 10.69985/VPIB8807

Endnotes

1 Dr. Arda Akartuna, "Are Pig Butchering Scammers Using AI? Here's What the Latest Trends Show," Elliptic, July 14, 2024, https://www.elliptic.co/blog/are-pig-butchering-scammers-using-ai-heres-what-the-latest-trends-show; Joshua Keating, "How Cyberscams Are Drawing China into Myanmar's Civil War," Vox, January 18, 2024, https://www.vox.com/world-politics/2024/1/18/24041696/cyberscams-myanmar-china-pig-butchering; Lily Hay Newman, "The Pig Butchering Invasion Has Begun," *Wired*, September 30, 2024, https://www.wired.com/story/pig-butchering-scam-invasion; Cezary Podkul, "What's a Pig Butchering Scam? Here's How to Avoid Falling Victim to One," *ProPublica*, September 19, 2022, https://www.propublica.org/article/whats-a-pig-butchering-scam-heres-how-to-avoid-falling-victim-to-one; Matt Burgess, "The Real-Time Deepfake Romance Scams Have Arrived," *Wired*, April 18, 2024, https://www.wired.com/story/yahoo-boys-real-time-deepfake-scams.

2 Renée DiResta and Josh A. Goldstein, "How Spammers and Scammers Leverage AI-Generated Images on Facebook for Audience Growth," *Harvard Kennedy School Misinformation Review*, August 15, 2024, https://doi.org/10.37016/mr-2020-151.

3 *2024 Global Financial Crime Report: Insights at the Intersection of Financial Crime Data & Real Survivor Stories* (Nasdaq Verafin, 2024), https://www.nasdaq.com/global-financial-crime-report.

4 Federal Trade Commission, "As Nationwide Fraud Losses Top $10 Billion in 2023, FTC Steps Up Efforts to Protect the Public," press release, February 8, 2024, https://www.ftc.gov/news-events/news/press-releases/2024/02/nationwide-fraud-losses-top-10-billion-2023-ftc-steps-efforts-protect-public.

5 In his wonderful book on the telegraph, *The Victorian Internet*, Tom Standage quotes Chicago police inspector John Bonfield in 1888: "It is a well-known fact that no other section of the population avail themselves more readily and speedily of the latest triumphs of science than the criminal class … The educated criminal skims the cream from every new invention, if he can make use of it." Tom Standage, *The Victorian Internet: The Remarkable Story of the Telegraph and the Nineteenth Century's Online Pioneers* (Berkley Books, 1998), 105.

6 Cassandra Cross, "Using Artificial Intelligence (AI) and Deepfakes to Deceive Victims: The Need to Rethink Current Romance Fraud Prevention Messaging," *Crime Prevention and Community Safety* 24, no. 1 (March 1, 2022): 30–41, https://doi.org/10.1057/s41300-021-00134-w.

7 While it may be difficult to disentangle a failed ICO from a scam ICO, *The Wall Street Journal* found that of the 1,450 ICOs they analyzed, 271 seemed to, as the newspaper put it, "show hallmarks of fraud." Shane Shifflett and Coulter Jones, "Buyer Beware: Hundreds of Bitcoin Wannabees Show Hallmarks of Fraud," *Wall Street Journal*, May 17, 2018, https://www.wsj.com/articles/buyer-beware-hundreds-of-bitcoin-wannabes-show-hallmarks-of-fraud-1526573115. An industry report from the firm Satis declared that 80% of ICOs were scams. Sherwin Dowlat, *Cryptoasset market coverage initiation: Network creation* (Satis Group Crypto Research, 2018). For a sociotechnical analysis, see Lana Swartz, "Theorizing the 2017 Blockchain ICO Bubble as a Network Scam," *New Media & Society 24*, no. 7 (July 9, 2022): 1695–1713, https://doi.org/10.1177/14614448221099224.

8 Of course, if a business is scammed out of a lot of money, those costs may eventually be passed on to the customers.

9 Lulu Graham, "Elon Musk Used in Fake AI Videos to Promote Financial Scam," RMIT University FactLab, August 15, 2023, https://www.rmit.edu.au/news/factlab-meta/elon-musk-used-in-fake-ai-videos-to-promote-financial-scam.

10 Steven Lindburg, "Quantum AI Review 2025: Scam Elon Musk Trading App Exposed!" *ScamCryptoRobots.com*, January 18, 2025, https://scamcryptorobots.com/quantum-ai-review-scam/. In the comments on this website, you can read testimonials of people who lost hundreds or even thousands of dollars from the scam.

11 There is virtually no recourse for those who "invest" in Quantum AI and similar scams. The Quantum AI websites direct targets to faux crypto exchanges or brokers who take the money and disappear. Typically, these brokers do not take credit cards and instead insist on wire transfer, cryptocurrency, or other forms of payment that are difficult or impossible to reverse. And because they are not hosted by app stores, there is no larger platform to hear complaints.

12 Kathleen Magramo, "British Engineering Giant Arup Revealed as $25 Million Deepfake Scam Victim," *CNN*, May 17, 2024, https://www.cnn.com/2024/05/16/tech/arup-deepfake-scam-loss-hong-kong-intl-hnk/index.html.

13 Ling Zhu and Laurie Harris, *Generative Artificial Intelligence and Data Privacy: A Primer*, CRS Report No. R47569 (Congressional Research Service, 2023), https://crsreports.congress.gov/product/pdf/R/R47569/5.

14 Emilio Ferrara, "GenAI against Humanity: Nefarious Applications of Generative Artificial Intelligence and Large Language Models," *Journal of Computational Social Science 7*, no. 1 (April 1, 2024): 549–69, https://doi.org/10.1007/s42001-024-00250-1.

15 Zhu and Harris, *Generative Artificial Intelligence and Data Privacy*.

16 Mark Jones and Henry Watkinson, "Deepfakes and Fraud: An Ever-Increasing Risk," *Solicitors' Journal* 167 (2024): 17.

17 Naroa Amezaga and Jeremy Hajek, "Availability of Voice Deepfake Technology and Its Impact for Good and Evil," in *Proceedings of the 23rd Annual Conference on Information Technology Education*, 23–28. SIGITE '22, New York, NY, USA: Association for Computing Machinery, 2022, https://doi.org/10.1145/3537674.3554742.

18 Dr. Arda Akartuna, *AI-Enabled Crime in the Cryptoasset Ecosystem* (Elliptic, 2024), https://www.elliptic.co/resources/ai-enabled-crime-in-the-cryptoasset-ecosystem.

19 Dijana Vukovic Grbic and Igor Dujlovic, "Social Engineering with ChatGPT," in *2023 22nd International Symposium INFOTEH-JAHORINA (INFOTEH)*, 1–5, 2023. https://doi.org/10.1109/INFOTEH57020.2023.10094141.

20 The availability of "dark LLMs" changes constantly; products are rebranded, open-source models are fine-tuned, and, of course, some sites purporting to sell "dark LLMs" are scams themselves. See Arthur Erzberger, "WormGPT and FraudGPT – The Rise of Malicious LLMs," **Trustwave**, August 8, 2023, https://www.trustwave.com/en-us/resources/blogs/spiderlabs-blog/wormgpt-and-fraudgpt-the-rise-of-malicious-llms.

21 Younghoo Lee and Ben Gelman, "The Dark Side of AI: Large-Scale Scam Campaigns Made Possible by Generative AI," *Sophos News*, November 27, 2023, https://news.sophos.com/en-us/2023/11/27/the-dark-side-of-ai-large-scale-scam-campaigns-made-possible-by-generative-ai.

22 Francesca Visser and Niamh McIntyre, "Doctored Footage and Hijacked Accounts: Anatomy of a Deepfake Scam Network," *The Bureau of Investigative Journalism*, June 5, 2024, https://www.thebureauinvestigates.com/stories/2024-06-05/anatomy-of-a-deepfake-scam-network.

23 "The Role of AI in Social Engineering," Zvelo, November 8, 2023, https://zvelo.com/the-role-of-ai-in-social-engineering

24 Ferrara, "GenAI against Humanity."

25 Jason Koebler, "Inside the Booming 'AI Pimping' Industry," *Wired*, November 28, 2024, https://www.wired.com/story/ai-pimping-industry-deepfakes-instagram.

26 See the fantastic recent investigative reporting by the *Economist's* Sue-Lin Wong for the podcast *Scam Inc* (https://www.economist.com/audio/podcasts/scam-inc), as well as the March 2025 Scam Empire project, led by the Organized Crime and Corruption Reporting Project, Swedish Television (SVT), and 30 other media partners from multiple countries (https://www.occrp.org/en/project/scam-empire).

27 "Everything You Need to Know About 'Scam Empire,'" Organized Crime and Corruption Reporting Project (OCCRP), March 5, 2025, https://www.occrp.org/en/project/scam-empire/scam-empire-everything-you-need-to-know-about-these-massive-investment-scams; "Scam Empire: Inside A Merciless International Investment Scam," OCCRP, March 5, 2025, https://www.occrp.org/en/project/scam-empire/scam-empire-inside-a-merciless-international-investment-scam.

28 United Nations Office on Drugs and Crime, "Transnational Organized Crime and the Convergence of Cyber-Enabled Fraud, Underground Banking and Technological Innovation in Southeast Asia: A Shifting Threat Landscape" (Bangkok, Thailand: Regional Office for Southeast Asia and the Pacific, United Nations, October 2024).

29 Charles Bethea, "The Terrifying A.I. Scam That Uses Your Loved One's Voice," *The New Yorker*, March 7, 2024, https://www.newyorker.com/science/annals-of-artificial-intelligence/the-terrifying-ai-scam-that-uses-your-loved-ones-voice.

30 Eren Kurshan, Mehta Dhagash and Tucker Balch, "AI versus AI in Financial Crimes & Detection: GenAI Crime Waves to Co-Evolutionary AI," in *Proceedings of the 5th ACM International Conference on AI in Finance*, 745–51, ICAIF '24, New York, NY, USA: Association for Computing Machinery, 2024, https://doi.org/10.1145/3677052.3698655; Christina Ianzito, "AI Fuels New, Frighteningly Effective Scams," *AARP*, April 3, 2024, https://www.aarp.org/money/scams-fraud/info-2024/ai-scams.html; Nick Heynen, "AI-Powered Scams: How to Protect Yourself in 2024," UW–Madison Information Technology, September 11, 2024, https://it.wisc.edu/news/ai-powered-scams-how-to-protect-yourself-2024.

31 GholamReza Zandi, Nor Azam Yaacob, Mazilena Tajuddin, and Nik Khadijah Nik Abdul Rahman, "Artificial Intelligence and the Evolving Cybercrime Paradigm: Current Threats to Businesses," *Journal of Information Technology Management* 16, no. 4 (October 16, 2024): 162–70, https://jitm.ut.ac.ir/article_99505_8c8a0737229c105a18f-73b7c2a61a9b3.pdf; "What Is Business Email Compromise (BEC)?" Microsoft Security, 2024, https://www.microsoft.com/en-us/security/business/security-101/what-is-business-email-compromise-bec.

32 Deen Freelon and Chris Wells, "Disinformation as Political Communication," *Political Communication* 37, no. 2 (March 3, 2020): 145–56.

33 Josh, A. Goldstein, et al., *Generative Language Models and Automated Influence Operations: Emerging Threats and Potential Mitigations* (Stanford Internet Observatory, OpenAI, and Georgetown's Center for Security and Emerging Technology, 2023), https://doi.org/10.48550/arXiv.2301.04246; Tiffany Hsu and Stuart A. Thompson, "Disinformation Researchers Raise Alarms About A.I. Chatbots," *New York Times*, February 8, 2023, https://www.nytimes.com/2023/02/08/technology/ai-chatbots-disinformation.html

34 DiResta and Goldstein, "Spammers and Scammers."

35 Jagadeesh Chandraiah and Sean Gallagher, "Sha Zhu Pan Scam Uses AI Chat Tool to Target iPhone and Android Users," *Sophos News*, August 2, 2023, https://news.sophos.com/en-us/2023/08/02/sha-zhu-pan-scam-uses-ai-chat-to-target-iphone-and-android-users.

36 Burgess, "The Real-Time Deepfake."

37 Julia Dickson and Lauren Burke Preputnik, "Cyber Scamming Goes Global: Unveiling Southeast Asia's High-Tech Fraud Factories," *Center for Strategic and International Studies: Critical Questions*, December 12, 2024, https://www.csis.org/analysis/cyber-scamming-goes-global-unveiling-southeast-asias-high-tech-fraud-factories; Lauren Burke Preputnik, Julia Dickson, Aly Senko, and Andrew Friedman, "Cyber Scamming Goes Global: Sourcing Forced Labor for Fraud Factories." *Center for Strategic and International Studies: Critical Questions*, December 12, 2024, https://www.csis.org/analysis/cyber-scamming-goes-global-sourcing-forced-labor-fraud-factories.

38 Craig Gibson and Josiah Hagen, "The Future of Whaling Attacks: AI-Powered Harpoon Whaling," *TrendMicro*, August 4, 2023, https://www.trendmicro.com/vinfo/us/security/news/cybercrime-and-digital-threats/the-future-of-whaling-attacks-ai-powered-harpoon-whaling.

39 The amount of audio needed is variable; so called "zero-shot" systems claim that they only need a few seconds, while older models need up to an hour of clearly enunciated speech. The more training data, the more realistic the audio generated.

40 NYC Department of Consumer and Worker Protection, "Tips on Artificial Intelligence Scams," NYC.gov, 2024, https://www.nyc.gov/site/dca/consumers/artificial-intelligence-scam-tips.page.

41 Nasdaq Verafin, "2024 Global Financial Crime Report."

42 "Report Fraud to the AARP Fraud Watch Network Helpline," AARP, Accessed January 20, 2025, https://www.aarp.org/money/scams-fraud/helpline.html.

43 "Who Experiences Scams? A Story for All Ages," Federal Trade Commission Consumer Protection: Data Spotlight, December 8, 2022, https://www.ftc.gov/news-events/data-visualizations/data-spotlight/2022/12/who-experiences-scams-story-all-ages.

44 Scott E. Augenbaum, "Keeping Seniors Safe from Cyber Scams," *Geriatric Nursing*, November 19, 2024, https://doi.org/10.1016/j.gerinurse.2024.10.081.

45 Federal Trade Commission, "Who Experiences Scams?"; Soren Kristiansen and Aksel Vassard Jensen, "Victimization in Online Gaming-Related Trade Scams: A Study Among Young Danes," *Nordic Journal of Criminology* 24, no. 2, 1–17, https://doi.org/10.18261/njc.24.2.6.

46 "'I Was Vulnerable': Artificial Intelligence Work-from-Home Job Scams Targeting Victims," *Boston 25 News*, May 29, 2024, https://www.boston25news.com/news/local/i-was-vulnerable-artificial-intelligence-work-from-home-job-scams-targeting-victims/LLV2FQUE2ZHGPOM72K7IBE6LHY.

47 Kelsey Havemann, "Bank Community Engagement: Protecting Teens from Financial Scammers," *ABA Banking Journal*, American Banking Association, October 2, 2024, https://bankingjournal.aba.com/2024/10/bank-community-engagement-protecting-teens-from-financial-scammers.

48 Erin Miller, "Why Victims of Fraud Are Hesitant to Come Forward," *WTKR*, November 7, 2023, https://www.wtkr.com/news/problem-solvers/why-victims-of-fraud-are-hesitant-to-come-forward.

49 Julia Ticona, "Red Flags, Sob Stories, and Scams: The Contested Meaning of Governance on Carework Labor Platforms," *New Media & Society* 24, no. 7 (July 2022): 1548–66, https://doi.org/10.1177/14614448221099233.

50 Jessica Vitak, Yuting Liao, Mega Subramaniam, and Priya Kumar, "'I Knew It Was Too Good to Be True,'" *Proceedings of the ACM on Human-Computer Interaction* 2, no. CSCW (November 1, 2018): 1–25, https://doi.org/10.1145/3274445.

51 Dana Fowle, "The High Cost for Fraud and Fees for the 'Unbanked,'" *Fox 5 Atlanta*, July 11, 2023, https://www.fox5atlanta.com/news/how-the-unbanked-can-protect-themselves; Szymon Morytko, "How to Protect the Underbanked from Fraud and Scams," FICO Blog, April 11, 2024, https://www.fico.com/blogs/how-protect-underbanked-fraud-and-scams.

52 "Common Scams," US Citizenship and Immigration Services, February 21, 2023, https://www.uscis.gov/scams-fraud-and-misconduct/avoid-scams/common-scams.

53 Tiffany Smedley, "Investment Scam Targeting WeChat Groups," *Federal Trade Commission Consumer Alerts*, May 18, 2023, https://consumer.ftc.gov/consumer-alerts/2023/05/investment-scam-targeting-wechat-groups

54 Elaine Chong and Ed Main, "Scammed by the Fake Chinese Police," *BBC*, July 6, 2024, https://www.bbc.com/news/articles/c6p2lq0qk41o.

55 Sarah Nguyễn, et al. "Studying Mis- and Disinformation in Asian Diasporic Communities: The Need for Critical Transnational Research Beyond Anglocentrism," *Harvard Kennedy School Misinformation Review*, March 24, 2022, https://doi.org/10.37016/mr-2020-95.

56 Matt Burgess, "Pig Butchering Scams Are Going High Tech," *Wired*, October 12, 2024, https://www.wired.com/story/pig-butchering-scams-go-high-tech.

57 Kevin Roose, "Welcome to the Yolo Economy," *New York Times*, April 21, 2021, https://www.nytimes.com/2021/04/21/technology/welcome-to-the-yolo-economy.html; Christian Nyemcsok, Hannah Pitt, Peter Kremer, and Samantha L. Thomas, "Young Men's Perceptions about the Risks Associated with Sports Betting: A Critical Qualitative Inquiry," *BMC Public Health* 22, no. 1 (April 30, 2022), https://doi.org/10.1186/s12889-022-13164-2; As Hailey Stauffer-Person writes in her article on multi-level marketing schemes, "In the United States, the line between financial fraud and legal business opportunity is blurry at best." Haile Stauffer-Person, "Artful Deception: The Multilevel Marketing Industry's Use of Smoke and Mirrors to Hide Their Pyramid-Shaped Truths," *Lewis & Clark Law Review 27*, no. 1 (2023): 355–94.

58 Aiha Nguyen and Alexandra Mateescu, *Generative AI and Labor: Power, Hype, and Value at Work* (Data & Society, December 4, 2024), https://datasociety.net/library/generative-ai-and-labor.

59 Microsoft Corporation, "What Is Business Email Compromise (BEC)?" Microsoft Security, 2024, https://www.microsoft.com/en-us/security/business/security-101/what-is-business-email-compromise-bec.

60 Yushawu Abubakari, "Modelling the Modus Operandi of Online Romance Fraud: Perspectives of Online Romance Fraudsters," *Journal of Economic Criminology* (November 16, 2024), https://doi.org/10.1016/j.jeconc.2024.100112.

61 Pranshu Verma, "They Thought Loved Ones Were Calling for Help. It Was an AI Scam," *Washington Post*, March 5, 2023, https://www.washingtonpost.com/technology/2023/03/05/ai-voice-scam.

62 Cory Doctorow, "Potemkin AI," *Pluralistic*, January 21, 2023, https://pluralistic.net/2023/01/21/potemkin-ai/#hey-guys.

63 Kevin Collier, "Phone Numbers for Airlines Listed on Google Directed to Scammers," *NBC News*, July 18, 2023, https://www.nbcnews.com/tech/tech-news/phone-numbers-airlines-listed-google-directed-scammers-rcna94766.

64 See for example: Mark Button and Cassandra Cross, "Technology and Fraud," in *Routledge Handbook of Technology*, Crime and Justice, ed. M.R. McGuire and Thomas J. Holt (Routledge, 2017), 91–103, https://doi.org/10.4324/9781315743981-5; Max Taylor, "Criminogenic Qualities of the Internet," *Dynamics of Asymmetric Conflict* 8, no. 2 (September 12, 2015): 97–106, https://doi.org/10.1080/17467586.2015.1065082; Keith J. Hayward and Matthijs M. Mass, "Artificial Intelligence and Crime: A Primer for Criminologists," *Crime, Media, Culture* 17, no. 2 (August 1, 2021): 209–33, https://doi.org/10.1177/1741659020917434.

65 Sally M. Gainsbury, Matthew Browne, and Matthew Rockloff, "Identifying Risky Internet Use: Associating Negative Online Experience with Specific Online Behaviours," *New Media & Society* 21, no. 6 (December 11, 2018): 1232–52, https://doi.org/10.1177/1461444818815442.

66 Anna Cooban, "This Bank Says 'Millions' of People Could Be Targeted by AI Voice-Cloning Scams," *CNN*, September 18, 2024, https://www.cnn.com/2024/09/18/tech/ai-voice-cloning-scam-warning/index.html.

67 Wai Yie Leong, Yuan Zhi Leong, and Wai Sang Leong, "The Intersection of Scammers and Artificial Intelligence," in *2024 International Conference on Consumer Electronics — Taiwan* (ICCE-Taiwan, 2024), 539–40, https://doi.org/10.1109/ICCE-Taiwan62264.2024.10674334.

68 "How to Avoid a Scam," *Federal Trade Commission Consumer Advice*, July 2023, https://consumer.ftc.gov/articles/how-avoid-scam; Heynen, "AI-Powered Scams."

69 Kiori Edwards, "From Classroom to Cloud: Why Cybersecurity Education Is Crucial for Tomorrow's Learners," Cybersecurity Undergraduate Research Showcase, November 18, 2024, https://digitalcommons.odu.edu/covacci-undergraduateresearch/2024fall/projects/8; Weiru Chen, Yuming He, Xin Tian, and Wu He, "Exploring Cybersecurity Education at the K-12 Level," *SITE Interactive Conference* (Association for the Advancement of Computing in Education [AACE], October 26, 2021): 108–114, https://www.learntechlib.org/primary/d/220175/.

70 Jim Kreidler, "Scam Proof the Young People in Your Life," *Federal Trade Commission Consumer Advice*, May 17, 2023, https://consumer.ftc.gov/consumer-alerts/2023/05/scam-proof-young-people-your-life.

71 Internet Crime Complaint Center, "Criminals Use Generative Artificial Intelligence to Facilitate Financial Fraud," Federal Bureau of Investigations, public service announcement, December 3, 2024, https://www.ic3.gov/PSA/2024/PSA241203; Ianzito, "AI Fuels New, Frighteningly Effective Scams."

72 Ianzito, "AI Fuels New, Frighteningly Effective Scams."

73 Mary Crawford, "'This Is Not a Scam!!' It's a Theatrical Tool for Scam Awareness," *Virginia Tech News*, February 28, 2023, https://news.vt.edu/articles/2023/02/clahs-this-is-not-a-scam-performance.html,

74 Nursyaidatul Kamar Shah, et al., "Linguistic Detective: Gamified Storyboard for Scam Detection," *International Journal of Research and Innovation in Social Science* VIII, no. X (November 4, 2024): 708–13, https://doi.org/10.47772/ijriss.2024.8100059.

75 USC Annenberg Norman Lear Center, "Hollywood, Health and Society," accessed February 11, 2025, https://hollywoodhealthandsociety.org.

76 Reece Rogers, "Real-Time Video Deepfake Scams Are Here. This Tool Attempts to Zap Them," *Wired*, October 15, 2024, https://www.wired.com/story/real-time-video-deepfake-scams-reality-defender/.

77 Oluwabusayo Adijat Bello and Komolafe Olufemi, "Artificial Intelligence in Fraud Prevention: Exploring Techniques and Applications Challenges and Opportunities," *Computer Science & IT Research Journal* 5, no. 6 (June 27, 2024): 1505–20, https://doi.org/10.51594/csitrj.v5i6.1252.

78 Gibson and Hagen, "The Future of Whaling Attacks."

79 Xue Wen Tan, Kenneth See, and Stanley Kok, "ScamGPT-J: Inside the Scammer's Mind, A Generative AI-Based Approach Toward Combating Messaging Scams," *ICIS 2024 Proceedings*, December 15, 2024, https://aisel.aisnet.org/icis2024/humtechinter/humtechinter/14.

80 Stripe, Inc, "A Primer on Machine Learning for Fraud Detection," Stripe.com, December 15, 2021, https://stripe.com/guides/primer-on-machine-learning-for-fraud-protection.

81 Ludivia Hernandez Aros, Luisa Ximena Bustamante Molano, Fernando Gutierrez-Portela, John Johver Moreno Hernandez, and Mario Samuel Rodríguez Barrero, "Financial Fraud Detection through the Application of Machine Learning Techniques: A Literature Review," *Humanities and Social Sciences Communications* 11, no. 1 (September 3, 2024): 1–22, https://doi.org/10.1057/s41599-024-03606-0.

82 Matt Alderton, "It Can Be Agony When a Loved One Is a Scam Victim — But Refuses to See It," AARP, August 20, 2024, https://www.aarp.org/money/scams-fraud/victims-in-denial/; Mark Button and Cassandra Cross, *Cyber Frauds, Scams and Their Victims* (Routlege, 2017); Cassandra Cross, "'I Knew It Was a Scam': Understanding the Triggers for Recognizing Romance Fraud," *Criminology & Public Policy* 22, no. 4 (2023): 613–37, https://doi.org/10.1111/1745-9133.12645.

83 Udari Madhushani Sehwag, et. al, "Can LLMs Be Scammed? A Baseline Measurement Study," *arXiv E-Prints* (October 1, 2024), https://doi.org/10.48550/arXiv.2410.13893.

84 "When AI Gets It Wrong: Addressing AI Hallucinations and Bias," MIT Sloan Teaching & Learning Technologies, November 12, 2024, https://mitsloanedtech.mit.edu/ai/basics/addressing-ai-hallucinations-and-bias; Leonardo Nicoletti and Dina Bass, "Humans Are Biased. Generative AI Is Even Worse," *Bloomberg*, June 9, 2023, https://www.bloomberg.com/graphics/2023-generative-ai-bias.

85 MoneyLIVE, "Combating the Next Wave of AI Fraud in Banking," accessed December 2, 2024, https://moneylive-insights.com/webinars/combating-the-next-wave-of-ai-fraud-in-banking; FinCEN, "FinCEN Alert on Fraud Schemes Involving Deepfake Media Targeting Financial Institutions," FinCEN Alert, Washington, DC: US Treasury Financial Crimes Enforcement Network, November 13, 2024, https://fincen.gov/sites/default/files/shared/FinCEN-Alert-DeepFakes-Alert508FINAL.pdf.

86 Gilad Gressel, Rahul Pankajakshan, and Yisroel Mirsky, "Discussion Paper: Exploiting LLMs for Scam Automation: A Looming Threat," in *Proceedings of the 3rd ACM Workshop on the Security Implications of Deepfakes and Cheapfakes* (New York, NY, USA: Association for Computing Machinery, 2024), 20–24, https://doi.org/10.1145/3660354.3660356.

87 DiResta and Goldstein, " Spammers and Scammers"; Erzberger, "WormGPT and FraudGPT."

88 Yi Ting Chua, et al., "Identifying Unintended Harms of Cybersecurity Countermeasures," in *2019 APWG Symposium on Electronic Crime Research (eCrime)* (Pittsburgh, PA, USA: IEEE, 2019), 1–15, https://doi.org/10.1109/eCrime47957.2019.9037589; Virginia Eubanks, *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor* (St. Martin's Press, 2018); Kevin De Liban, *Inescapable AI: The Ways AI Decides How Low-Income People Work, Live, Learn, and Survive* (Techtonic Justice, 2024), https://www.techtonicjustice.org/reports/inescapable-ai.

89 Kevin De Liban and Alice Marwick, "AI Can't Solve Government Waste — and May Hurt Vulnerable Americans," *Tech Policy Press*, December 10, 2024, https://www.techpolicy.press/ai-cant-solve-government-waste-and-may-hurt-vulnerable-americans.

90 Andrew W. Eichner, "Artificial Intelligence and Weaponized Illusions: Methodologies for Federal Fraud Prosecutions Involving Deepfakes," *American University Law Review* 73 (2024): 1319–66.

91 Jones and Watkinson, "Deepfakes and Fraud."

92 "FTC Proposes New Protections to Combat AI Impersonation of Individuals," Federal Trade Commission, press release, February 15, 2024, https://www.ftc.gov/news-events/news/press-releases/2024/02/ftc-proposes-new-protections-combat-ai-impersonation-individuals.

93 United Nations Office on Drugs and Crime, "Transnational Organized Crime."

94 Alana Wise, "A Phone Company Developed an AI 'Granny' to Beat Scammers at Their Own Game," *NPR*, December 10, 2024, https://www.npr.org/2024/12/10/nx-s1-5220362/daisy-ai-granny-o2-fraud-spam-prevention.

95 Vicky Baker, "The Legal Battles behind Anna Delvey's Dancing With The Stars Debut," *BBC*, September 16, 2024, https://www.bbc.com/news/articles/cvgd2y5e23jo.

96 "/r/Scams," Reddit, accessed January 20, 2025, https://www.reddit.com/r/Scams.

97 "CAN-SPAM Act: A Compliance Guide for Business," Federal Trade Commission, accessed April 15, 2025, https://www.ftc.gov/business-guidance/resources/can-spam-act-compliance-guide-business.

98 Emilio Ferrara, "The History of Digital Spam," *Communications of the ACM* 62, no. 8 (July 24, 2019): 82–9, https://doi.org/10.1145/3299768.

Data & Society is an independent nonprofit research and policy institute, studying the social implications of data-centric technologies and automation. We recognize that the same innovative technologies that may benefit society can also be abused to invade privacy, provide new tools of discrimination, foreclose opportunity and harm individuals and communities. We believe that technology policy must be grounded in empirical evidence, and serve the public.

www.datasociety.net | @datasociety

Layout by Surbhi Chawla
Illustrations by Gloria Mendoza

MAY 2025