AI, Deepfakes, and the Future of Financial Deception

Perry Carpenter
KnowBe4

Prepared Remarks for:
SEC Investor Advisory Committee
Retail Investor Fraud in America Panel Session
March 6, 2025

Addressing The Rise of AI-Enabled Fraud

Thank you for the opportunity to address the Securities and Exchange Commission on this critical issue. As Chief Human Risk Management Strategist at KnowBe4 and author *of FAIK: A Practical Guide to Living in a World of Deepfakes, Disinformation, and AI-Generated Deceptions*, I bring over two decades of expertise in studying deception and social engineering. From this perspective, I can clearly see that we have reached a critical inflection point: AI is fundamentally transforming financial fraud—making sophisticated scams more accessible, scalable, and convincing than at any point in history.

The threat is immediate and pervasive. According to research from Deloitte, 25.9% of executives say their organizations have experienced one or more deepfake incidents.[1] And a study cited in CFO Magazine reports even higher numbers, stating that 92% of companies have experienced financial loss due to a deepfake.[2] Such statistics highlight the rapidly escalating threat landscape.

As an example, in early 2024, a finance employee in Hong Kong was deceived into wiring $25 million to criminals after participating in a deepfake video conference with executives who appeared authentic but were entirely synthetic.[3] And the threats aren't only targeting organizations, they are using a wide range of lures – from fake kidnapping ploys[4] to romance scams[5] – to defraud ordinary citizens.

These examples represent just a few of the emerging threats we face. AI is a force multiplier for fraudsters, making sophisticated deception accessible to a broader spectrum of malicious actors,

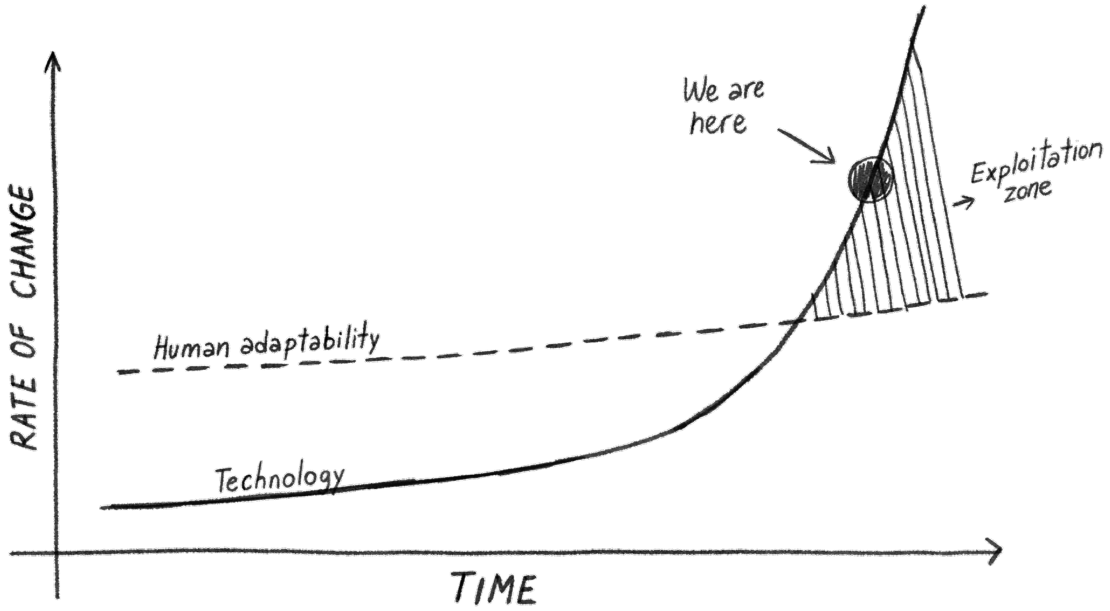
[1] "Generative AI and the fight for trust". Deloitte. May 2024.
https://www2.deloitte.com/content/dam/Deloitte/us/Documents/Advisory/us-generative-ai-and-the-fight-for-trust.pdf
[2] "92% of companies have experienced financial loss due to a deepfake." CFO Magazine. November 6, 2024.
https://www.cfo.com/news/most-companies-have-experienced-financial-loss-due-to-a-deepfake-regula-report/732094/
[3] "Hong Kong Clerk Defrauded of $25 Million in Sophisticated Deepfake Scam." SecureWorld. Feb 13, 2024.
https://www.secureworld.io/industry-news/hong-kong-deepfake-cybercrime
[4] "Beware of Virtual Kidnapping Ransom Scam." National Institutes of Health.
https://ors.od.nih.gov/News/Pages/Beware-of-Virtual-Kidnapping-Ransom-Scam.aspx
[5] "Deepfake romance scam raked in $46 million from men across Asia, police say." CNN. October 15, 2024.
https://www.cnn.com/2024/10/15/asia/hong-kong-deepfake-romance-scam-intl-hnk/index.html

thereby challenging the SEC's core mission: investor protection, market integrity, and capital formation.

The Exploitation Zone: Where Technology Outpaces Awareness & Safety

We are currently in what I term the "Exploitation Zone"—a concept I adapted from the work of Alphabet's Astro Teller. The *exploitation zone* is the ever-widening chasm between technological advancement and human adaptability.[6] Here's how I describe it:

- Technology progresses exponentially, while our collective capacity to adapt—cognitively, socially, and regulatorily—advances slowly and much more linearly.
- This model shows that the gap between technology and human adaptability is expanding at an accelerating rate.
- Cybercriminals and scammers take advantage of the widening gap, exploiting a lack of awareness, emotional triggers, and weak technological controls.



This vulnerability is particularly perilous for financial markets because:

1. Financial decisions are inherently emotional, rendering investors susceptible to manipulation (a.k.a. social engineering).
2. Most individuals lack the technical acumen to detect sophisticated AI-driven deceptions.
3. The stakes within our financial systems are exceptionally high.

[6] "FAIK: A Practical Guide to Living in a World of Deepfakes, Disinformation, and AI-Generated Deceptions." Perry Carpenter. October 2024. https://thisbookisfaik.com/

Tools once exclusive to elite state actors are now accessible to virtually anyone with an internet connection. As a result, even unsophisticated perpetrators can craft compelling, targeted scams at scale.[7]

The Twin Motivations: Money and Minds

Nearly all deception falls into two categories: financial fraud (money) or influence campaigns (minds).

Money-Motivated Deception: AI as a Fraud Multiplier

AI is rapidly diminishing the skill barrier for fraud, facilitating the execution of high-level scams. Here are a few examples:

- **Deepfake Executive Fraud:**
  Malicious actors have cloned the voices and appearances of executives during live video calls, deceiving employees into transferring substantial funds. This represents an evolved, AI-enhanced variant of the traditional Business Email Compromise.
- **AI-Enhanced Investment Scams:**
  Schemes such as "pig butchering" now employ AI to generate convincing financial documents, counterfeit trading platforms, and personalized communications. What's particularly concerning is how AI enables relationship-building at unprecedented scale – one operation can now conduct thousands of individualized scams concurrently.
- **Romance Scams with Financial Motives:**
  These scams have become more sophisticated, utilizing AI to generate contextually appropriate responses around the clock. They can even engage in real-time deepfake video interactions, increasing credibility and amplifying their emotional manipulation.

Minds-Motivated Deception: AI as a Societal and Market Manipulation Tool

Imagine this scenario –a deepfaked statement from the Federal Reserve Chair circulates online, falsely announcing an emergency rate hike or regulatory shift. By the time the truth is verified, markets have already reacted.

The scenario I outlined above shows that, beyond direct fraud, AI can be a powerful weapon to manipulate financial sentiment, disseminate disinformation, and erode public trust:

- **Artificial Consensus Creation:**
  AI-powered bots can fabricate false impressions of investor or public consensus about particular issues or securities. This manufactured consensus can trigger momentum in markets or public opinion.

[7] This is a result of what researchers at Harvard describe as the "Jagged technological frontier." See: Navigating the Jagged Technological Frontier: Field Experimental Evidence of the Effects of AI on Knowledge Worker Productivity and Quality." September 22, 2023. https://www.hbs.edu/ris/Publication%20Files/24-013_d9b45b68-9e74-42d6-a1c6-c72fb70c7282.pdf

- **Targeted Erosion of Confidence:**
  Strategic deployment of synthetic media can quickly destabilize specific companies, financial markets[8], or democratic processes. As social engineering expert Rachel Tobac describes, even innocuous AI-generated content, like videos depicting long ATM queues or bank closures, can trigger bank runs or market panics.[9]
- **The Liar's Dividend:**
  Research shows how AI-generated disinformation systematically undermines trust in media, financial institutions, regulatory bodies, and other pillars of society, creating what is called "the Liar's Dividend" – a situation where even authentic evidence can be dismissed as potentially fake.[10]

While the risks posed by AI-driven deception are significant, we are beginning to see promising defensive measures emerge. Recognizing the urgent need for identity verification safeguards, some financial institutions are deploying a range of methods and protocols specifically designed to counter deepfake impersonation.[11] At the same time, media literacy programs have demonstrated success in helping investors recognize and question suspicious content before acting. For instance, a study published in the Proceedings of the National Academy of Sciences demonstrated that a one-hour, self-directed digital media literacy intervention significantly improved older adults' ability to discern fake from true news, increasing accuracy from 64% to 85%.[12] However, these efforts alone are not enough—without regulatory support, scalable industry-wide adoption of AI-specific fraud detection measures will lag behind evolving threats. The next step is to ensure these countermeasures become the norm rather than the exception.

Technical Reality: Key Capabilities That Enable Deception

The increasing sophistication of AI-driven deception isn't just theoretical—it's powered by tangible advancements in multimodal synthesis, automation, and hyper-personalization. Understanding these technical enablers is critical to formulating effective countermeasures.

In my research and experiments, I've observed three key technical developments that have fundamentally transformed deception capabilities:

1. **Multimodal Synthesis:**
   Today's AI can seamlessly generate authentic text, voice, and video, enabling fraudsters to create cohesive, interactive deceptive environments. At the 2024 DEFCON hacker

[8] "Fake Pentagon explosion photo caused a real dip in the stock market." Mashable. May 22, 2023. https://mashable.com/article/ai-deepfake-image-pentagon-explosion-hoax
[9] "OpenAI's New Product Makes Incredibly Realistic Fake Videos" ScienceFriday interview with Rachel Tobac. February 23, 2024. https://www.sciencefriday.com/segments/sora-ai-video/
[10] "Deepfakes, Elections, and Shrinking the Liar's Dividend." Brennan Center for Justice. January 23, 2024. https://www.brennancenter.org/our-work/research-reports/deepfakes-elections-and-shrinking-liars-dividend
[11] "Deepfakes in the FinancialSector: Understanding the Threats, Managing the Risks." FS-ISAC. October 2024. https://www.fsisac.com/hubfs/Knowledge/AI/DeepfakesInTheFinancialSector-UnderstandingTheThreatsManagingTheRisks.pdf
[12] "A digital media literacy intervention for older adults improves resilience to fake news." NIH National Library of Medicine. April 9, 2022. https://pmc.ncbi.nlm.nih.gov/articles/PMC8994776/

conference, I proved that AI-powered scambots are just as effective (if not more effective) as seasoned social engineers.[13] During that conference, I was able to ethically test an AI scam system I built capable of conducting real-time phone conversations using cloned voices, complete with contextual awareness. This system was able to quickly build rapport, impersonate trusted individuals, and trick targets into divulging information or taking actions.

2. **Multisystem Integration Vulnerabilities:**
   As I discovered and documented in my research, AI components generally operate in modular silos without awareness of how they're being used when integrated with other technologies. An LLM generating investment advice has no awareness it's feeding a voice synthesis service that's conducting fraudulent calls. This creates blind spots where no single component is responsible for the resulting harm.
3. **Personalization at Scale:**
   AI systems can analyze vast amounts of public data to identify and target individuals with precision. Systems like this can create highly individualized custom LLM-generated text, combined with other synthetically generated or pre-existing media, exploiting personal vulnerabilities to hijack thoughts and manipulate actions.

These technical capabilities create unique challenges that require both targeted regulatory responses and broader societal adaptation. The following recommendations address this complex threat landscape.

Recommendations for Regulatory Action: Advancing AI-Focused Financial Protections

As AI continues to reshape financial markets, fraud, and regulatory oversight, the SEC has taken commendable steps to address the risks and opportunities it presents. The establishment of the Cyber and Emerging Technologies Unit (CETU), the integration of AI-driven market surveillance, and proposed regulations for AI use in financial decision-making demonstrate a proactive approach to safeguarding market integrity.

To build upon these efforts and further enhance protections, I offer the following recommendations:

1. **Establish an AI Financial Fraud Task Force within CETU**
   The SEC's Cyber and Emerging Technologies Unit (CETU) is a crucial step in addressing financial threats posed by emerging technologies, including artificial intelligence. To further enhance its efforts, the SEC could form a dedicated AI Financial Fraud Task Force within CETU. This task force would focus on monitoring AI-driven financial fraud schemes, issuing industry-specific guidance, and collaborating with financial institutions to strengthen AI risk mitigation strategies. Given the rapid evolution of AI in fraud and market manipulation, a specialized task force would provide targeted expertise and regulatory oversight to combat these emerging risks.

[13] "[PROVED] Unsuspecting Call Recipients Are Super Vulnerable to AI Vishing." KnowBe4 blog. August 16, 2024. https://blog.knowbe4.com/proved-unsuspecting-call-recipients-are-super-vulnerable-to-ai-vishing

2. **Further Strengthen AI Transparency and Combat "AI-Washing"**
   The SEC has taken an active stance against misleading AI claims ("AI-washing"), where companies exaggerate or misrepresent their AI capabilities to investors. To build on this effort, the SEC could implement clearer AI disclosure requirements, ensuring firms provide verifiable, evidence-backed statements about AI-driven products and services. Additionally, encouraging AI watermarking, digital signatures, or blockchain-based verification for financial disclosures and earnings calls would help prevent AI-generated misinformation from misleading investors.
3. **Mandate AI-Specific Fraud Detection Requirements for Financial Institutions**
   While the SEC has prioritized cybersecurity risk management, there is a growing need to address AI-driven fraud techniques. Financial institutions should be required to adopt AI-specific fraud detection mechanisms—such as deepfake detection, behavioral analysis, and anomaly monitoring for high-value transactions. AI is already being used to enhance fraud techniques at scale, and codifying AI-specific detection requirements would align with the SEC's broader mission to combat financial crimes enabled by emerging technologies.

As AI-enabled fraud scales, organizations must recognize that employees are both prime targets and first lines of defense. Just as businesses have invested in phishing awareness training, they must now prepare employees to detect AI-generated deception, from deepfake fraud to synthetic identity scams.

Recommendations for Businesses

Businesses must take a proactive approach to protecting their employees, customers, and financial assets. Combating AI-enabled deception requires a layered strategy—blending training, advanced detection tools, and strong incident response plans. The following steps will help organizations reduce risk and build resilience against emerging AI threats.

1. **Train Employees to Recognize AI-Driven Deception**
   AI-powered fraud isn't just a technical problem—it's a human problem. After all, people—not systems—are the targets. That's why security awareness programs must evolve alongside the threat landscape. Organizations should integrate AI threat awareness into security training, just as they did with phishing and social engineering. Employees must learn to spot deepfake scams and synthetic media attacks that target both businesses and individuals.
2. **Deploy AI-Specific Fraud Detection & Authentication Measures**
   Traditional security tools are not designed to detect AI-generated fraud. Businesses must implement AI-powered defenses, including deepfake detection, behavioral anomaly monitoring, and biometric authentication with liveness checks. High-risk transactions and executive communications should be protected by multi-factor verification and AI-driven anomaly detection to flag suspicious activity in real time.
3. **Develop AI-Informed Incident Response Plans**
   The rise of AI-driven scams means businesses must be prepared for deepfake-enabled fraud, synthetic identity attacks, and AI-enhanced phishing schemes. Organizations should update incident response plans to include rapid fraud assessment procedures,

forensic AI analysis capabilities, and clear internal escalation paths. Employees, finance teams, and security personnel must know how to identify, verify, and respond swiftly to AI-enabled deception before financial or reputational damage occurs.

Recommendations for Individuals, Communities, and Educators

As AI-driven scams and misinformation continue to shape the digital landscape, individuals, communities, and educators play a critical role in building resilience. Beyond relying on regulations and technology-based solutions, equipping people with the right knowledge and skills is essential for fostering informed decision-making and collective defense against AI-enabled manipulation. The following recommendations focus on media literacy, adversarial thinking, and AI awareness to empower individuals and communities to navigate an increasingly complex information ecosystem.

1. **Strengthen Media & Information Literacy**
   To combat misinformation and disinformation—both AI-generated and traditional—individuals and communities need stronger media literacy skills. This includes learning to evaluate sources, recognize bias, and verify claims using critical thinking frameworks such as the SIFT method[14] (Stop, Investigate, Find trusted coverage, Trace to the original) or similar approaches. Educators should incorporate fact-checking exercises into curricula, while communities can encourage peer-driven initiatives to strengthen trust in reliable information.
2. **Teach Adversarial Thinking**
   Understanding how attackers think is critical to effective defense. Organizations and individuals should regularly practice "threat modeling"—analyzing new technologies not just for their intended uses but for potential vulnerabilities and misuse scenarios. Educators can integrate adversarial thinking exercises into curricula to help students anticipate and counter real-world digital threats.
3. **Build AI Literacy for Everyday Decision-Making**
   As AI tools become more embedded in daily life, individuals and communities must understand not just how AI deceives, but how it functions overall. AI literacy should include how generative AI works, its strengths and weaknesses, and how it influences decision-making in areas like finance, healthcare, and online interactions. Educators can integrate AI-awareness activities, while communities can provide workshops or discussions to help people navigate AI-generated content with confidence.

Key Takeaways

1. **AI fundamentally alters risk profiles for fraud:** Traditional fraud detection frameworks are insufficient against AI-enhanced threats.
2. **Technology outpaces adaptability:** The exploitation zone is widening faster than our natural adaptation. Combatting this requires proactive intervention.

[14] "Evaluating Resources and Misinformation." University of Chicago Library. https://guides.lib.uchicago.edu/c.php?g=1241077&p=9082322

3. **Dual threat landscape:** Both direct financial fraud (money) and market manipulation (minds) present systemic risks to investor protection and market integrity.
4. **Defense necessitates collaboration:** Effective response requires coordinated efforts among regulators, financial institutions, and the public.
5. **Existing tools aren't perfect—but we can't wait for perfect:** While today's detection and mitigation technologies have limitations, they are still critical in combating AI-driven fraud. A multi-layered approach—combining technical solutions, regulatory frameworks, corporate training, and public awareness—is essential to staying ahead. Organizations must equip employees with the knowledge and skills to recognize AI-driven threats, just as they do with phishing and social engineering. This is an arms race, and waiting for flawless solutions will only allow threats to evolve unchecked.

The Time to Act is Now

The challenges outlined above are not hypothetical—they are already here. AI deception is scalable, accessible, and evolving faster than our defenses. But hope is not lost. Criminals have means, motive, and opportunity: but, so do we.

Throughout history, financial markets have successfully adapted to technological disruptions—from the telegraph to high-frequency trading. The key difference today is the compressed timeframe and democratized access to powerful tools that can be easily weaponized for deception. With thoughtful regulation that works in concert with industry innovation and public education, we can navigate the challenge while preserving the integrity and efficiency of our financial systems.

The exploitation zone will always exist—but the amount of harm it causes will depend on our ability to recognize and manage it. Now is the time to act. Through education, policy, and technical defenses, we can reduce the gap.

Thank you for your consideration of these critical issues.

Perry Carpenter
Chief Human Risk Management Strategist
KnowBe4, Inc.