Prize Winner

A Retiree's Guide to Artificial Intelligence Risks and Mitigating Those Risks

Gregory Whittaker, FSA, FASSA

Any views and ideas expressed in the essay are the author's alone and may not reflect the views and ideas of the Society of Actuaries, the Society of Actuaries Research Institute, Society of Actuaries members, nor the author's employer.

INTRODUCTION

As noted by Tegmark [1], there are many competing definitions of intelligence including:

…the capacity for logic, understanding, planning, emotional knowledge, self-awareness, creativity, problem solving and learning.

Broadly, intelligence can be referred to as the ability to accomplish complex goals. Artificial intelligence (AI) in turn is the umbrella term for the algorithm-based technology that enables computers and machines to simulate intelligence. These include rules-based expert systems (a deterministic form of AI), machine learning (AI that learns more as more data is introduced but which has a performance plateau), and deep learning (a form of machine learning built on a network of computations similar to the human brain). As noted by Sun & Huo [2]:

deep learning $\subset$ *machine learning* $\subset$ *artificial intelligence*

Managing post-retirement risks has been well documented by the Society of Actuaries Research Institute (SOA) [3]. As noted by the SOA, the exposure of retirees to fraud and theft may increase as retirees control more assets, financial products become more complex, more retirees use computers, and scammers become more proficient.

On March 6, 2024, the Federal Bureau of Investigation (FBI) released their 2023 Internet Crime Report [4]. In 2023, the FBI's Internet Crime Complaint Center (IC3) registered 880,418 complaints from the American public with potential losses in excess of $12.5 billion. Not all complaints included an associated age range, but out of approximately $8.2 billion in potential losses where an age range was indicated, approximately $3.4 billion was in respect of Americans over the age of 60.

Whilst internet fraud has been in existence for decades, with phishing scams still the most prevalent, fraudulent schemes have become more sophisticated with the use of AI. It is the purpose of this article to highlight some AI tools that can be used to enhance fraudulent schemes targeting retirees and in turn to discuss some practical risk mitigation strategies that can be employed when confronted by these tools.

DEEPFAKES

The European Parliamentary Research Service [5] defines deepfakes as:

…manipulated or synthetic audio or visual media that seem authentic, and which feature people that appear to say or do something they have never said or done, produced using artificial intelligence techniques, including machine learning and deep learning.

A number of risks are posed by deepfakes as summarized in Table 1.

Table 1
RISKS ASSOCIATED WITH DEEPFAKES

Psychological Harm	Financial Harm	Societal Harm
Extortion	Extortion	News media manipulation
Defamation	Identity theft	Damage to economic stability
Intimidation	Fraud	Damage to the justice system
Bullying	Stock-price manipulation	Damage to the scientific system
Undermining trust	Brand damage	Erosion of trust
	Reputational damage	Damage to democracy
		Manipulation of elections
		Damage to international relations
		Damage to national security

Source: European Parliamentary Research Service.

As reported by the FBI IC3, losses due to investment scams are the most of any crime type tracked by that entity, with investment fraud losses estimated at $4.57 billion in 2023. Of that amount, an estimated $3.94 billion involved investment fraud with reference to a cryptocurrency.

A common scam uses social media advertisements including deepfake videos or images of celebrities or public figures, who claim to have made large profits from online trading platforms. This so-called “click bait” drives traffic to malicious websites, where an individual is asked to sign up to the trading platform. Once the form is submitted, the individual is contacted by the scammer and is requested to make a small up-front payment to access the platform. Details are provided to download a cryptocurrency app so that more funds can be invested. These apps initially show fake profits and scammers persuade investors to increase their investment. However, when an attempt is made to withdraw funds, individuals often find that they have been locked out of their accounts and their money has disappeared.

Various mitigation strategies are required to combat deepfakes as shown in Table 2. These can be broadly classified on an individual level as discussed by Singh and Dhiman [6] and on a technological level or policy level as discussed by Al-Khazraji et al. [7].

Table 2
MITIGATION STRATEGIES FOR DEEPFAKES

Individual Level	Technological or Policy Level
Understanding the existence of deepfakes	Advancing deepfake detection technology
Recognizing signs of manipulated content	Building robust authentication systems
Applying critical thinking skills	Strengthening social media platform policies
Verifying authenticity before content sharing	Legal and policy frameworks
Supporting authentic sources of information	Promoting AI literacy among the elderly

On March 8, 2024, the European Parliament approved the Artificial Intelligence Act [8]. Importantly, Article (70b) requires deepfakes to be labelled as such:

Further to the technical solutions employed by the providers of the system, deployers, who use an AI system to generate or manipulate image, audio or video content that appreciably resembles existing persons, places or events

and would falsely appear to a person to be authentic ('deep fakes'), should also clearly and distinguishably disclose that the content has been artificially created or manipulated by labelling the artificial intelligence output accordingly and disclosing its artificial origin.

It remains to be seen whether this will form a template for other countries around the world. As noted by Professor Mark Lemley of Stanford University [9]:

Generative AI is developing at a stunning speed, creating new and thorny problems in well-established legal areas, disrupting long-standing regimes of civil liability—and outpacing the necessary frameworks, both legal and regulatory, that can ensure the risks are anticipated and accounted for.

VOICE CLONING

As described by the Federal Trade Commission [10]:

You get a call. There's a panicked voice on the line. It's your grandson. He says he's in deep trouble – he wrecked the car and landed in jail. But you can help by sending money. You take a deep breath and think. You've heard about grandparent scams. But darn, it sounds just like him. How could it be a scam? Voice cloning, that's how.

From January 2020 to June 2021, the FBI's IC3 received around 650 reports of grandparent scams, resulting in losses of approximately $13 million [11]. During that period, over 90 victims reported that money was picked up from their home, resulting in losses of approximately $3.6 million.

The United States Senate Special Committee on Aging [12] identifies various red flags in these types of scams. Typically, the grandchild or law enforcement officer asks you to keep the incident a secret. The demand for money is immediate and a suggestion is made to send money via a gift card or wire transfer. In these situations, it is recommended that you hang up and call back the number of the family member that you know to be their genuine number. The use of a family safe word can also be used as a simple test of the authenticity of the caller.

FRAUDGPT

As noted by Balona [13], large language models (LLMs) are trained on massive datasets of text, enabling them to learn complex patterns and relationships in language. Over the period from 2020 to 2023 various LLMs were released by OpenAI, with the most recent model GPT-4 (Generative Pre-trained Transformer).

GPT-4 was released on March 14, 2023. On July 22, 2023, FraudGPT emerged on the Dark Web Forum [14]. As noted by Falade [15], for an annual subscription fee of $ 1,700, inexperienced cybercriminals can unearth a range of capabilities including the creation of phishing emails, the development of exploits, hacking tools, the discovery of vulnerabilities in systems and the provision of guidance on cybercrime. FraudGPT marks the beginning of a new era of cybercriminal at scale. It is likely that we will soon see the end of badly punctuated, misspelled, misdirected and factually inaccurate phishing emails.

SOPHISTICATED TARGETING

Phishing is the term for generalized cyberattacks carried out by email or SMS. Among the common phishing scams is the 419-scam where victims are promised large sums of money in exchange for an investment in a business activity that does not exist.

Highly personalized cyberattacks are commonly referred to as "spearfishing." Whilst spearfishing attacks are not new, AI provides the opportunity to automate this process at scale. AI tools can review large volumes of data to identify potential victims and tailor messages that capture an individual's unique circumstances. This is likely an area of emerging risk for retirees.

An important consideration is to investigate how retirement changes the social life and social network of retirees. If they have a greater propensity to turn to social media to fill the void created by no longer interacting with colleagues in the workplace, there is the potential that more personalized information will become available to scammers to harvest. Social media literacy among retirees becomes key and workplace programs in preparation for retirement should include basic internet and social media training to avoid the many pitfalls.

Mitigation strategies for combatting identify fraud [16] and the healthy use of social media for retirees [17] is shown in Table 3.

Table 3
MITIGATION STRATEGIES FOR IDENTITY FRAUD AND THE HEALTHY USE OF SOCIAL MEDIA

Identity Fraud Mitigation Strategies	Healthy Use of Social Media
Don't share personal information on social media	Is it true, is it necessary, is it kind?
Regularly review and clean up social media profiles	Live in the moment
Don't answer unsolicited emails or calls	Link instead of comparing yourself
Always verify the identity of the caller	Follow people and things that bring you joy
Use strong, unique passwords for all accounts	Keep things in real life
Install and regularly update antivirus software	Start your day intentionally and use your time wisely
Store important documents in a secure place	Make events accessible
Regularly review bank and credit card statements	Take a break and support others in doing so
Monitor your credit profile periodically	Don't struggle alone and seek help for overuse
Don't click on suspicious links from unknown sources	
Logout of apps or websites after use	
Use biometrics or PINs to lock personal devices	

CONCLUSION

Historians date fraud back to around 300BC when a Greek trader named Hegestratos took out a large bottomry insurance policy (where typically a loan is received up front) with the deliberate intention of not wanting to repay the loan [18]. Ponzi schemes turned 100 years old in 2020 [19]. What is certain is that in 2024, AI technologies can improve the efficiency and scale of existing fraudulent activities, let alone develop entirely new schemes.

* * * * *

Gregory Whittaker, FSA, FASSA is a consulting actuary in Johannesburg, South Africa. He can be reached at gregory@algorithm-ca.com.

REFERENCES

[1] Tegmark, M. (2017). Life 3.0: *Being human in the age of Artificial Intelligence*. Penguin Books, United Kingdom.
[2] Sun, Z. & Huo, Y. (2019). The spectrum of big data analytics. *Journal of Computer Information Systems* 61(2), 154-62.
[3] Society of Actuaries (2011). Managing Post-Retirement Risks: A Guide to Retirement Planning. Online at: https://www.soa.org/globalassets/assets/files/research/projects/post-retirement-charts.pdf
[4] https://www.ic3.gov/Media/PDF/AnnualReport/2023_IC3Report.pdf
[5] European Parliament (2021). Tackling deepfakes in European policy. Online at: https://www.europarl.europa.eu/RegData/etudes/STUD/2021/690039/EPRS_STU(2021)690039_EN.pdf

[6] Singh, P. & Dhiman, B. (2024). Exploding AI-Generated Deep Fakes and Misinformation: A Threat to Global Concern in the 21st Century. *J Robot Auto Res, 5*(1), 01-07.
[7] Al-Khazraji, S., Saleh, H.H., Khalid, A.I., & Mishkhal, I.A. (2023). Impact of Deepfake Technology on Social Media: Detection, Misinformation and Societal Implications. *The Eurasia Proceedings of Science Technology Engineering and Mathematics.*
[8] https://artificialintelligenceact.eu/the-act/
[9] https://law.stanford.edu/stanford-lawyer/articles/artificial-intelligence-and-the-law/
[10] https://consumer.ftc.gov/articles/scammers-use-fake-emergencies-steal-your-money
[11] https://www.fbi.gov/contact-us/field-offices/miami/news/fbi-miami-warns-of-grandparent-fraud-scheme
[12] https://www.congress.gov/118/crpt/srpt111/CRPT-118srpt111.pdf
[13] Balona, C. (2023). ActuaryGPT: Applications of large language models to insurance and actuarial work. Convention, Actuarial Society of South Africa, 2023.
[14] https://netenrich.com/blog/fraudgpt-the-villain-avatar-of-chatgpt
[15] Falade, P.V. (2023). Decoding the threat landscape. *International Journal of Scientific Research in Computer Science, Engineering and Information Technology* 9(5), 185-198.
[16] Sumsub Identity Fraud Report (2023).
[17] https://news.mit.edu/2020/mindhandheart-nine-tips-healthy-social-media-use-0123
[18] Li, L., & McMurray, A. (2022). Corporate Fraud Trends. In *Corporate Fraud Across the Globe* (pp. 169-199). Singapore: Springer Nature Singapore.
[19] Artzrouni, M. (2009). The mathematics of Ponzi schemes. *Mathematical Social Sciences*, *58*(2), 190-201.