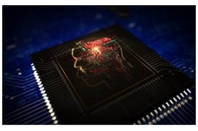
Infowars – the large language model threat

Technology vendors are keen to promote the benefits of artificial intelligence (AI) in fighting money laundering and terrorist financing, helping obliged entities detect and block illicit money flows. But every technology can also be wielded by bad actors; AI is no different. It can be abused to attack and evade know your customer (KYC) controls, ***Keith Nuthall*** *discovers.*



Knowledge exchange

RELATED

Level up – large language models raise the stakes

AML IT – legacy and change

Incoming tide – AI for AML

Fortified – the arms trade and AML

Effective sanctions list management: a seven-step guide

ChatGPT itself has told *MLB* how it and its AI cousins might undermine anti-money laundering/counter financing of terrorism (AML/CFT) defences. Asked how this could happen, it said attacks might be made via AI-assisted:

- Synthetic identity fraud generation through merging elements from multiple stolen identities to open fraudulent accounts, which can be abused for ML/TF;
- Document forgery using advanced machine learning algorithms to generate forged documents (such as IDs, utility bills, bank statements) that can pass know your customer (KYC) checks;
- The creation of synthetic or mimicked real biometric data to conduct biometric spoofing, fooling face, fingerprint and voice recognition systems;
- Requests by chatbots to engage with bank customer service representatives to extract personal information or manipulate them through social engineering, aiding impersonation techniques to subvert KYC;
- Analysis of transaction patterns of legitimate users, to mimic them, making fraudulent transactions seem normal and harder to detect;
- Flooding of KYC systems, with AI making multiple fake applications to overwhelm and slow IT and human controls, making them more likely to overlook genuine-looking fraudulent applications for accounts;
- Identification of vulnerabilities or loopholes in a verification process and exploiting them;
- Harvesting of vast amounts of data from various sources, including the dark web, to assemble comprehensive profiles of real people to bypass KYC checks;
- Undertaking of rapid testing of multiple strategies to breach KYC processes, to see what works: and
- Counteraction of anomaly detection systems through AI studying their red flags, telling launderers what behaviours are regarded as normal, helping them mask illicit activities.

ChatGPT advises: "To counter these threats, financial institutions must continuously update and evolve their KYC controls, leveraging AI and other technologies themselves to detect and deter fraudulent activities. Collaboration between institutions, sharing threat intelligence, and adopting a proactive security stance will be essential."

But is it right? How serious are these threats? And what can be done to counteract those that are a real danger?

Sharpen the spear faster

A report from WithSecure, a cyber-security advisory service owned by the Helsinki-based F-Secure Corporation, agrees the risk is real. In a January [2023] paper [1] it warned how large language model (LLM) chat-based AI can generate copy honed for social engineering and manipulation of subjects, such as bank staff, being "designed to trick a user into opening a malicious attachment or visiting a malicious link".

Such spear phishing of KYC professionals can be hastened by LLM services, said a paper from the UK Centre for the Governance of AI and the University of Oxford, Oxford Internet Institute: "Spear phishing is traditionally a time-consuming and labour-intensive process that can involve several steps, such as identifying high-value targets, conducting personalised research to gather relevant information on the target, and crafting a tailored message that appears to come from a trusted acquaintance. However, with the integration of AI, this process can be made more efficient." [2]

Leak check

Another paper, from SRM University, New Delhi, India, warned that banks needed to be careful if they integrate LLMs into their chatbots, lest they are persuaded by money launderers to give away information to bad actors that could be useful in evading KYC controls: "Compliance with financial regulations, data protection laws, and privacy regulations is essential. Banks need to ensure that the chatbot's functionalities align with these regulations, such as not disclosing sensitive customer information or violating anti-money laundering (AML) and know-your-customer (KYC) requirements." [3]

Maladjustment

An April [2023] paper from Spanish and Mexican researchers (Universidad de Navarra; Universidad Pontificia Comillas, Madrid; University of Guadalajara) warned how ChatGPT can automate malware production, with cybercriminals using it and other LLMs "for hacking, scamming, and other illegal activities". That included "creating malware to steal files or to phish for credentials". Also, "ChatGPT malware can spy on keyboard strokes; steal, compress, and distribute files; or install backdoors," all useful ML tradecraft for criminal looking to abuse financial networks and institution: "Because of ChatGPT's user friendliness, very little technical knowledge is required to automate malware creation," noted the paper. [4]

Speaking to *MLB*, Yong Li, Anti-Money Laundering Advisor, for Canada-based consultancy The AMLShop, said: "The technology itself is neutral, but the LLM abuse problem is real. LLMs can be used by bad actors to increase the effectiveness of crimes such as email scams, which can cause financial and psychological harm to public."

Already out there...?

He said financial institutions and other obliged entities need to be careful about information they have already put into the public domain, because that might have been read by an LLM as training data, helping a bad actor subvert AML controls such as KYC. "Typically, financial institutions wouldn't publish the top-secret internal information, but there's a chance they will publish policy and procedures. Based on the analysis of that text data maybe the bad actor can figure out that's something they don't do or they do [in KYC]... And then they can further question and refine their question [to an LLM chat box] to gain insight."

That risk will intensify as vertical LLMs are trained specifically to aid a particular industry or task, and that could include AML, or – if developed by a bad actor – ML. This verticalisation is already happening. For example, BloombergGPT was launched in March [2023], as a large language model with 50 billion parameters trained on financial data. [5]

A Bloomberg note said the LLM could generate financial queries, answer financial questions and suggest financial news headlines.

Dark designs

Yong Li said that should organised crime develop a vertical ML LLM, or abuse a vertical AML LLM, trained on KYC, CDD (customer due diligence), STR (suspicious transcation report), UBO (ultimate beneficial owner) data and procedures, this could have "very, very bad consequences". It could aid "very targeted actions that may cause more damage", he said. At present, while general LLMs can be useful for launderers, "they may not be able to figure out the technical detail of a particular organisation, of what they use for AML". But LLM protocols are open source, and existing LLMs help with programming a specialised LLM. "If a vertical LLM tool is built in future for AML, then that tool would be very, very powerful if it says: 'Yes, here for this situation, here's the possible control,' and the bad actors say: 'Okay, because this is giving me a list of controls, I will try to avoid all of them.'"

The risk gets worse if a vertical ML LLM, developed by organised crime, used leaked financial information culled from the dark web as training data. [6] For instance, in March [2023], a user of the hacker-focused BreachForums claimed to possess 60GB of stolen Deutsche Bank data and was offering to sell it to the highest bidder. [7]

A bad actor who has developed an ML vertical LLM could say, said Yong Li: "'How about feeding that data into my engine that generates some results which give me hints about how to commit crime?'. It's very possible everyone can download the source code as it's open source or at least learn the methodology – then they feed the data to the model to gain insight of something." He has yet to learn of an example of organised crime building a sophisticated LLM, but he said: "The risk is real. The probability is there. And the potential impact is there."

Risk management and testing

What should be done? Yong Li said all developers and commercial exploiters of LLM ecosystems should consider and mitigate the potential LLM risks of money laundering: "For example, regulators should provide ethical guidelines for LLMs; LLM creators/vendors should engage in responsible innovation to prevent insecure output and sensitive information disclosure." Also, organisations should continuously monitor and test common LLM tools' Q&A capabilities and adjust their internal KYC controls accordingly, so that bad actors won't be able to gain insights into how these operate.

The reason here is that the training data of LLMs and KYC-related data may have much in common. "LLMs are data-driven models [i.e., the models are the results of the training and tuning by using large data sets]", so obliged entities need to aware if their "secured information [such as internal KYC controls] has something in common with the training data of LLMs..."

For good actors to prevail in the use and abuse of AI in ML and AML, "everybody needs to take responsibility," said Yong Li – from vendors, to regulators to obliged entities. Governments need to provide guidelines and say: "'Here's an area we shouldn't touch before you train a model. Don't feed it data which could be exploited by bad actors for malicious use,'" he said. "It takes a village. Everybody in this ecosystem when you build a tool should be aware of this kind of issue. It will impact everybody."

Notes

1. https://labs.withsecure.com/publications/creatively-malicious-prompt-engineering
2. https://arxiv.org/ftp/arxiv/papers/2303/2303.09377.pdf
3. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4563696
4. https://arxiv.org/pdf/2304.11215.pdf
5. https://arxiv.org/pdf/2303.17564.pdf
6. https://www.sleyber.io/what-financial-data-is-sold-on-dark-web-marketplaces/
7. https://cybernews.com/news/deutsche-bank-data-offered-up-on-dark-web/

Sep 17 2023

Print this page | Send to a colleague | Email the Editor

Comments

TOPIC ALERTS

Customer Due Diligence
Training
Technology
Frequency: Daily
No. of articles: up to 10
Email Address: timon.molloy@informa.com
Manage topic alerts | Set up Alert

TRAINING

The money launderer's bookshelf: part I

CUSTOMER DUE DILIGENCE

FATF targets Recommendations at tax and transparency

Latin lessons

News EU proposes zero threshold for sender data on money transfers