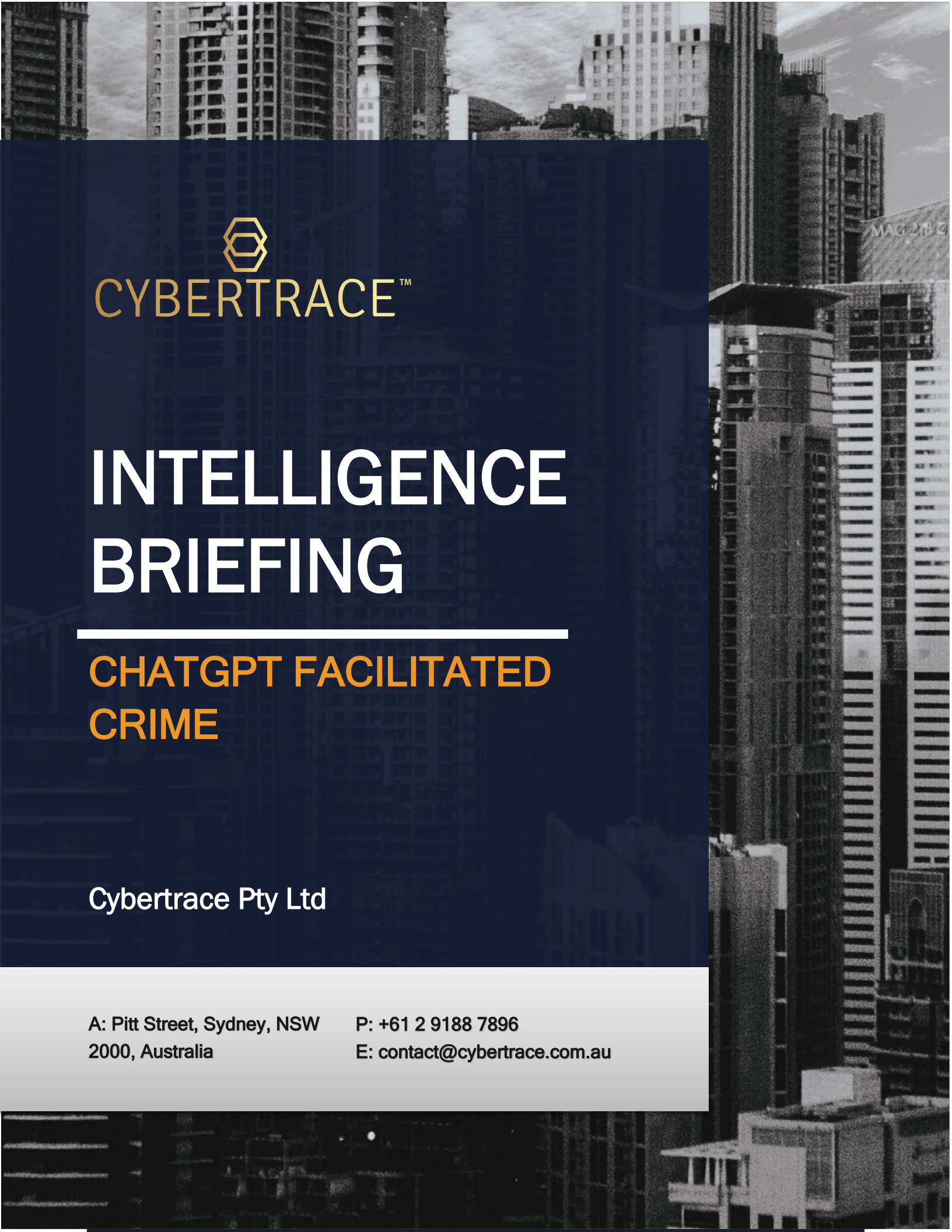

CYBERTRACE™

INTELLIGENCE BRIEFING

CHATGPT FACILITATED CRIME

Cybertrace Pty Ltd

A: Pitt Street, Sydney, NSW 2000, Australia

P: +61 2 9188 7896

E: contact@cybertrace.com.au

CHATGPT FACILITATED CRIME

BACKGROUND

The rapid development of technology has brought numerous advantages to society; however, it has also brought new challenges in the realm of cybersecurity and cybercrime. Cybercriminals are able to utilize new technologies to conduct nefarious activities in increasingly sophisticated ways. Large Language Models (LLMs) are one such technology that has the potential to be used for both beneficial and malicious purposes.

ChatGPT has proven highly popular since its release and multiple apps and platforms have been released drawing upon the ChatGPT capability, such as Bing Chat, ChatOn and ChatSonic among others.

This briefing examines the impact of LLMs on law enforcement, with a specific focus on the Artificial Intelligence (AI) based chatbot, ChatGPT which was released by OpenAI in November 2022. ChatGPT is built on top of OpenAI's GPT-3.5 and GPT-4 foundational LLM known as ChatGPT.

OpenAI describe themselves as an AI research and deployment company with,

> *"a mission to ensure that artificial general intelligence benefits all of humanity."*

Furthermore, this report examines the possible risks associated with cybercriminals or malicious actors possessing LLMs. It emphasizes the significance of the government, industry, and community increasing their awareness of the latest technological advancements including LLMs to combat cybercrime more effectively.

Fig. 1. The technology driven criminal landscape is constantly pushing the limits of law enforcement. AI and ChatGPT are not an exception.

CONTEXT

ChatGPT is a free online application that synthesizes information from multiple websites to provide answers to user prompts and queries. Basically, it works like a search engine but one which combines various sources of information into one cohesive ready-to-use and human-like answer. Part of what is known as artificial intelligence, ChatGPT is an LLM which was developed by the company, OpenAI. It was trained in huge amounts of language and ChatGPT can recognize, replicate, and respond to natural human language patterns. Importantly, ChatGPT's developers trained it to interact in a conversational way, meaning it can answer follow-up questions and refine answers. Not only does it have the potential to revolutionize online search engines, but users can also deploy ChatGPT to write emails, reports, and any kind of document for them. Released in late 2022 and refined in early 2023, ChatGPT has generated a huge amount of public and commercial interest.

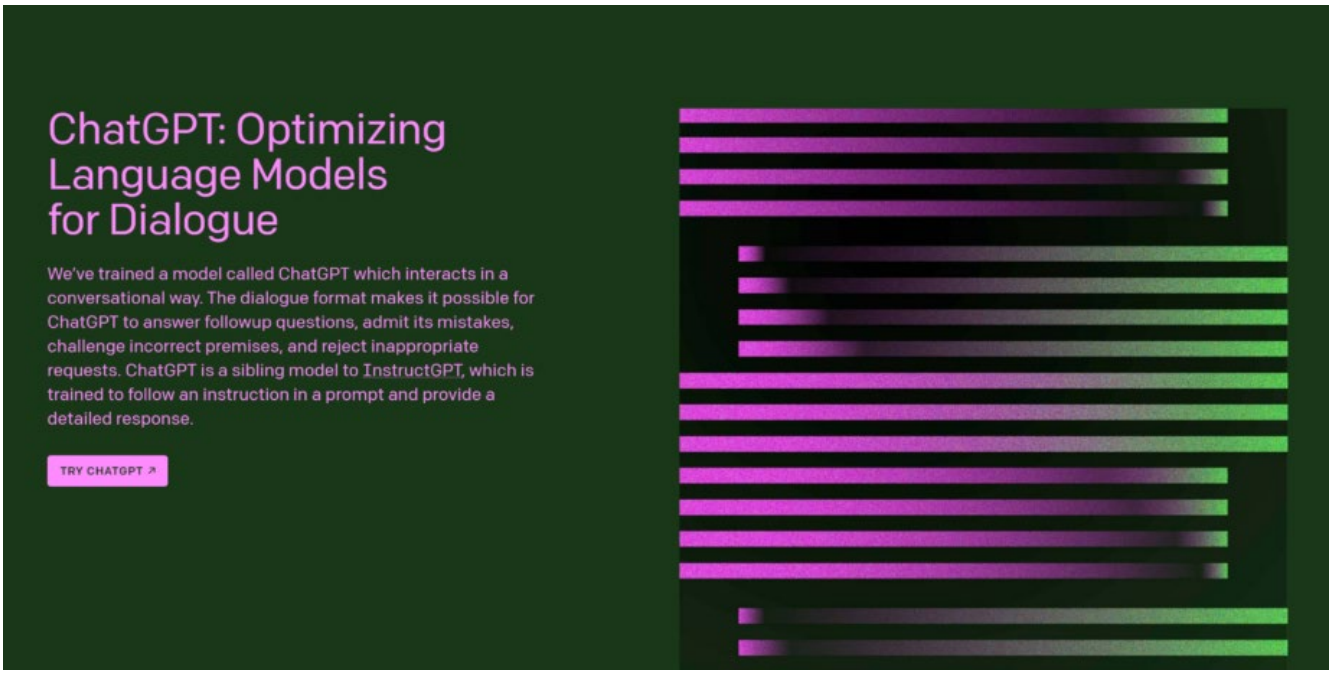

Fig. 2. OpenAI's webpage introducing ChatGPT. The ChatGPT web-based user interface is located at https://chat.openai.com/.

Based on Open-Source (OSINT) records, an analysis by Cybertrace revealed that ChatGPT has several limitations, particularly in its technical capabilities. The model's responses generally lack reliable references to the sources of information, which may result in biased answers. In addition, while ChatGPT's responses appear plausible, they are often incorrect or completely fabricated due to its current inability to understand the meaning and truth behind language. As a result, the model is only capable of producing simplistic responses and cannot perform advanced analysis. However, as the LLMs are constantly evolving, higher level capability is likely to be included in future editions.

Other downsides of ChatGPT are attributed to bad-faith actors utilizing the technology for malicious purposes. Europol recently warned of potential criminal uses of ChatGPT in its publication, 'The Impact of LLM on Law Enforcement'. Europol comments that ChatGPT enables criminals to rapidly acquire knowledge in a variety of illicit activities, such as terrorism, kidnapping, and disinformation campaigns.

ChatGPT's summarization capabilities accelerate this process, making it easier for individuals to acquire the necessary knowledge and skills to engage in criminal activities. Additionally, ChatGPT can be easily leveraged to generate hate speech, disinformation, and propaganda, as its authoritative tone lends credibility to inaccurate information. With improvements in the underlying technology, it may become increasingly difficult to distinguish between bot-generated and human-generated text.

In addition to the ability to create realistic human responses, this is compounded by ChatGPT's ability to create responses in multiple languages. It has been identified that this capability has been exploited by cyber criminals, specifically for cyber fraud cases.

The emergence and increasing prevalence of LLM (LLMs), such as OpenAI's GPT-3, has significant implications for law enforcement agencies. LLMs can generate highly realistic and convincing content, such as news articles, social media posts, and even fake personas. Criminals are increasingly using LLMs for nefarious purposes, including phishing, fraud, and disinformation campaigns.

Moreover, LLMs can also be used to automate various tasks in cybercrime, including spamming, credential stuffing, and brute-forcing. This poses a significant challenge for law enforcement agencies, as it becomes increasingly difficult to distinguish between legitimate and fraudulent content generated by LLMs. Additionally, it becomes harder to track down and identify the perpetrators behind such activities.

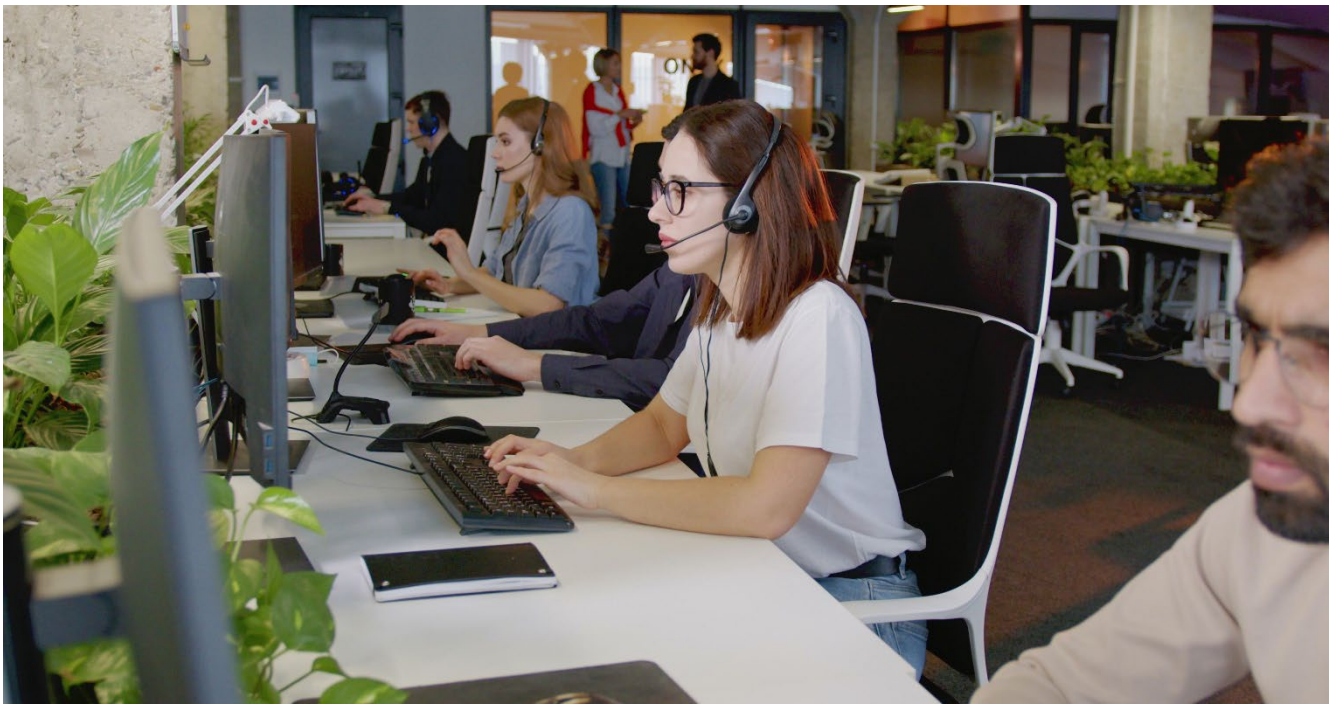

Fig. 3. Scam call centers have become significantly more productive through using ChatGPT to manipulate their victims in various languages and with almost instantaneous human like responses.

The report by Europol's European Cybercrime Centre (EC3) aims to provide insights into the use of LLMs by criminals and its potential impact on law enforcement. It highlights the need for law enforcement agencies to develop new strategies and techniques to detect and combat LLM-generated content. This includes improving the training and resources available to law enforcement personnel, increasing cooperation between law enforcement agencies and academia, and exploring the use of advanced technologies such as blockchain and artificial intelligence to combat LLM-based crime.

Overall, the report highlights the need for law enforcement agencies to remain vigilant and proactive in their efforts to combat cybercrime involving LLMs. The growing prevalence of LLMs underscores the importance of ensuring that law enforcement agencies are equipped with the necessary tools and knowledge to combat these emerging threats.

Fig. 4. Europol is aware that ChatGPT and other LLM can be used by criminals for nefarious purposes, including generating fraudulent content and automating cybercrime tasks, as highlighted in their report on the impact of LLMs on law enforcement. Image Source: Getty Images.

OpenAI themselves are aware of the potential for their technology to be used by criminals or for other nefarious activities such as propaganda and disinformation. In October 2021, researchers from OpenAI, Georgetown University's Center for Security and Emerging Technology, and the Stanford Internet Observatory conducted a collaborative investigation into the potential misuse of large language models for disinformation purposes.

The study involved a workshop attended by 30 experts in disinformation research, machine learning, and policy analysis, and was followed by more than a year of

research. The resulting report highlighted the risks posed by language models to the information ecosystem when used to enhance disinformation campaigns and proposed a framework for assessing possible countermeasures.

ANALYSIS & ASSESSMENT

Cybertrace conducted an analysis of the potential criminal usage of ChatGPT where it was identified that the chatbot can engage with individuals in a friendly and conversational manner, seeking to establish a rapport in support of criminal activity such as grooming victims for cyber fraud.

Cybertrace cautions that the chatbot may be used to gain access to sensitive information, such as personal and financial data. The report notes that the chatbot may be used by scammers to trick individuals into sharing this information by asking seemingly innocuous questions, such as their name or date of birth, which can be used to build a profile of the individual and potentially commit identity theft or financial fraud.

Fig. 5. ChatGPT adds a new dimension to cyber fraud and identity theft.

Further, the ChatGPT chatbot is not inherently malicious, however, it is a tool that can be used for both beneficial and harmful purposes. Scammers may use the chatbot to create a false sense of trust and establish a relationship with their targets before attempting to exploit them. This includes avoiding sharing personal information with the chatbot and being vigilant for suspicious requests or conversations. The report also recommends using reputable antivirus software and keeping up to date with the latest cybersecurity threats.

Overall, Cybertrace highlights a potential threat to individuals' personal and financial security using a chatbot and emphasizes the need for individuals to remain vigilant and take steps to protect themselves from this type of scam. Considering this potential threat, we recommend that governments and organisations increase literacy levels relating to the potential risks related to ChatGPT through public awareness campaigns and training where required.

SUMMARY

Based on open-source records, an analysis reveals that ChatGPT has several limitations in its technical capabilities, and its responses lack reliable references to the sources of information, which may result in biased answers. Additionally, while ChatGPT's responses appear plausible, they are often incorrect or completely fabricated due to its current inability to understand the meaning and truth behind language. However, as the LLMs are constantly evolving, and higher-level capability is likely to be included in future editions.

Bad-faith actors are utilizing ChatGPT for malicious purposes. Europol and Cybertrace warn of potential criminal uses of ChatGPT, enabling criminals to rapidly acquire knowledge in a variety of illicit activities, such as terrorism, kidnapping, and to assist cybercriminal operations. Furthermore, ChatGPT can generate hate speech, disinformation, and propaganda, as its human-like and authoritative tone lends credibility to inaccurate information. This is compounded by ChatGPT's ability to create responses in multiple languages, which has been exploited by cyber criminals, specifically for cyber fraud cases. Scam call centers have become significantly more

productive through using ChatGPT to manipulate their victims in various languages and with almost instantaneous human-like responses.

The emergence and increasing prevalence of LLMs have significant implications for law enforcement agencies. Criminals are increasingly using LLMs for nefarious purposes, including phishing, fraud, and disinformation campaigns. LLMs can also automate various tasks in cybercrime, including spamming, credential stuffing, and brute-forcing. This poses a significant challenge for law enforcement agencies, as it becomes increasingly difficult to distinguish between legitimate and fraudulent content generated by LLMs.

The report by Europol's European Cybercrime Centre highlights the need for law enforcement agencies to develop new strategies and techniques to detect and combat LLM-generated content. This includes improving the training and resources available to law enforcement personnel, increasing cooperation between law enforcement agencies and academia, and exploring the use of advanced technologies such as artificial intelligence to combat LLM-based crime. The growing prevalence of LLMs underscores the importance of ensuring that law enforcement agencies are equipped with the necessary tools and knowledge to combat these emerging threats.

According to research conducted by Cybertrace, it was determined that one of the primary risks associated with ChatGPT is the apparent low literacy levels for ChatGPT facilitated crime, disinformation, and propaganda. In order to improve literacy levels, Cybertrace recommends that governments and industry collaborate to create effective public awareness campaigns and internal training for staff.

REFERENCES

1. Cybertrace. (n.d.). ChatGPT Scam Investigation. https://www.cybertrace.com.au/chatgpt-scam-investigation/
2. OpenAI. (n.d.). About OpenAI. https://openai.com/about/
3. OpenAI. (n.d.). Chat with GPT. https://chat.openai.com/

4. Europol. (2021). The Impact of LLM on Law Enforcement. https://www.europol.europa.eu/publications-documents/impact-of-large-language-models-law-enforcement
5. Getty Images. (n.d.). Fig. 3. Europol is aware that ChatGPT and other LLM can be used by criminals for nefarious purposes, including generating fraudulent content and automating cybercrime tasks, as highlighted in their report on the impact of LLMs on law enforcement. [Photograph]. https://www.gettyimages.com/license/1332804841
6. "The Impact of LLM on Law Enforcement" report by Europol's European Cybercrime Centre (EC3): https://www.europol.europa.eu/publications-documents/impact-of-llms-on-law-enforcement
7. OpenAI's webpage introducing ChatGPT: https://openai.com/blog/chat-gpt/
8. "The Dark Side of AI Language Models" report by Georgetown University's Center for Security and Emerging Technology: https://cset.georgetown.edu/wp-content/uploads/CSET-The-Dark-Side-of-AI-Language-Models-091520.pdf
9. "Language Models are Few-Shot Learners" paper by OpenAI researchers, introducing GPT-3: https://arxiv.org/abs/2005.14165
10. "The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation" report by researchers from the Future of Humanity Institute and the Center for the Study of Existential Risk: https://maliciousaireport.com/

DISCLAIMER

ENQUIRIES

For sales, further information or feedback, please contact our team at contact@cybertrace.com.au, or +61 2 9188 7896.