Council of the
European Union

Brussels, 4 May 2023
(OR. en)

8535/23

COSI 80
ENFOPOL 221
CRIMORG 71
IXIM 116
CT 84
CATS 25
CYBER 109
TELECOM 128
JAI 558

NOTE

From:	Europol
To:	Delegations
Subject:	ChatGPT - The impact of Large Language Models on Law Enforcement

Delegations will find in the annex the Europol report "ChatGPT - The impact of Large Language Models on Law Enforcement".

EUROPOL

TECH WATCH FLASH

ChatGPT

The impact of Large Language Models on Law Enforcement

27/03/2023

CONTENTS

INTRODUCTION 2

BACKGROUND: LARGE LANGUAGE MODELS AND CHATGPT 3

SAFEGUARDS, PROMPT ENGINEERING, JAILBREAKS 5

CRIMINAL USE CASES 7

Fraud, impersonation, and social engineering 7

Cybercrime 8

IMPACT AND OUTLOOK 10

RECOMMENDATIONS 12

CONCLUSION 13

INTRODUCTION

The release and widespread use of ChatGPT – a large language model (LLM) developed by OpenAI – has created significant public attention, chiefly due to its ability to quickly provide ready-to-use answers that can be applied to a vast amount of different contexts.

These models hold masses of potential. Machine learning, once expected to handle only mundane tasks, has proven itself capable of complex creative work. LLMs are being refined and new versions rolled out regularly, with technological improvements coming thick and fast. While this offers great opportunities to legitimate businesses and members of the public it also can be a risk for them and for the respect of fundamental rights as criminals and bad actors may wish to exploit LLMs for their own nefarious purposes.

In response to the growing public attention given to ChatGPT, the Europol Innovation Lab organised a number of workshops with subject matter experts from across the organisation to explore how criminals can abuse LLMs such as ChatGPT, as well as how it may assist investigators in their daily work. The experts who participated in the workshops represented the full spectrum of Europol's expertise, including operational analysis, serious and organised crime, cybercrime, counterterrorism, as well as information technology.

Thanks to the wealth of expertise and specialisations represented in the workshops, these hands-on sessions stimulated discussions on the positive and negative potential of ChatGPT, and collected a wide range of practical use cases. While these use cases do not reflect an exhaustive overview of all potential applications, they provide a glimpse of what is possible.

The objective of this report is to examine the outcomes of the dedicated expert workshops and to raise awareness of the impact LLMs can have on the work of the law enforcement community. As this type of technology is undergoing rapid progress, this document further provides a brief outlook of what may still be to come, and highlights a number of recommendations on what can be done now to better prepare for it.

Important notice: The LLM selected to be examined in the workshops was ChatGPT. ChatGPT was chosen because it is the highest-profile and most commonly used LLM currently available to the public. The purpose of the exercise was to observe the behaviour of an LLM when confronted with criminal and law enforcement use cases. This will help law enforcement understand what challenges derivative and generative AI models could pose.

A longer and more in-depth version of this report was produced for law enforcement consumption only.

BACKGROUND: LARGE LANGUAGE MODELS AND CHATGPT

Artificial Intelligence

Artificial Intelligence (AI) is a broad field of computer science that involves creating intelligent machines that can perform tasks that typically require human-level intelligence, such as understanding natural language, recognizing images, and making decisions. AI encompasses various subfields, including machine learning, natural language processing, computer vision, robotics, and expert systems.

Neural Networks

Neural Networks, also known as Artificial Neural Networks (ANN), are computing systems inspired by the structure and function of the human brain. They consist of interconnected nodes or neurons that are designed to recognize patterns and make decisions based on input data.

Deep learning

Deep Learning is a subfield of machine learning that involves training artificial neural networks, which are computing systems inspired by the structure and function of the human brain, to recognize patterns and make decisions based on large amounts of data. Deep Learning has been particularly successful in fields such as image recognition, natural language processing, and speech recognition.

Supervised/unsupervised learning

Supervised Learning is a type of machine learning that involves training a model using labeled data, where the desired output is already known. The model learns to make predictions or decisions by finding patterns in the data and mapping input variables to output variables.

Unsupervised Learning is a type of machine learning that involves training a model using unlabeled data, where the desired output is unknown. The model learns to identify patterns and relationships in the data without being given specific instructions, and is often used for tasks such as clustering, anomaly detection, and dimensionality reduction.

Definitions provided by ChatGPT.

ChatGPT is a large language model (LLM) that was developed by OpenAI and released to the wider public as part of a research preview in November 2022. Natural language processing and LLMs are subfields of artificial intelligence (AI) systems that are built on deep learning techniques and the training of neural networks on significant amounts of data. This allows LLMs to understand and generate natural language text.

Over recent years, the field has seen significant breakthroughs due in part to the rapid progress made in the development of supercomputers and deep learning algorithms. At the same time, an unprecedented amount of available data has allowed researchers to train their models on the vast input of information needed.

The LLM ChatGPT is based on the Generative Pre-trained Transformer (GPT) architecture. It was trained using a neural network designed for natural language processing on a dataset of over 45 terabytes of text from the internet (books, articles, websites, other text-based content), which in total included billions of words of text.

3

The training of ChatGPT was carried out in two phases: the first involved unsupervised training, which included training ChatGPT to predict missing words in a given text to learn the structure and patterns of human language. Once pre-trained, the second phase saw ChatGPT fine-tuned through Reinforcement Learning from Human Feedback (RLHF), a supervised learning approach during which human input helped the model learn to adjust its parameters in order to better perform its tasks.

The current publicly accessible model underlying ChatGPT, GPT-3.5, is capable of processing and generating human-like text in response to user prompts. Specifically, the model can answer questions on a variety of topics, translate text, engage in conversational exchanges ('chatting'), and summarise text to provide key points. It is further capable of performing sentiment analysis, generating text based on a given prompt (i.e. writing a story or poem), as well as explaining, producing, and improving code in some of the most common programming languages (Python, Java, C++, JavaScript, PHP, Ruby, HTML, CSS, SQL). In its essence, then, ChatGPT is very good at understanding human input, taking into account its context, and producing answers that are highly usable.

In March 2023, OpenAI released for subscribers of ChatGPT Plus its latest model, GPT-4. According to OpenAI, GPT-4 is capable of solving more advanced problems more accurately[1]. In addition, GPT-4 offers advanced API integration and can process, classify, and analyse images as input. Moreover, GPT-4 is claimed to be less likely to respond to requests for 'disallowed content' and more likely to produce factual responses than GPT-3.5[2]. Newer versions with greater functionalities and capabilities are expected to be released as the development and improvement of LLMs continues.

Limitations

Still, the model has a number of important limitations that need to be kept in mind. The most obvious one relates to the data on which it has been trained: while updates are made on a constant basis, the vast majority of ChatGPT's training data dates back to September 2021. The answers generated on the basis of this data do not include references to understand where certain information was taken from, and may be biased. Additionally, ChatGPT excels at providing answers that sound very plausible, but that are often inaccurate or wrong[3] [4]. This is because ChatGPT does not fundamentally understand the meaning behind human language, but rather its patterns and structure on the basis of the vast amount of text with which it has been trained. This means answers are often basic, as the model struggles with producing advanced analysis of a given input[5]. Another key issue relates to the input itself, as often, the precise phrasing of the prompt is very important in getting the right answer out of ChatGPT. Small tweaks can quickly reveal different answers, or lead the model into believing it does not know the answer at all. This is also the case with ambiguous prompts, whereby ChatGPT typically assumes to understand what the user wants to know, instead of asking for further clarifications.

Finally, the biggest limitation of ChatGPT is self-imposed. As part of the model's content moderation policy, ChatGPT does not answer questions that have been classified as harmful or biased. These safety mechanisms are constantly updated, but can still be circumvented in some cases with the correct prompt engineering. The following chapters describe in more detail how this is possible and what implications arise as a result.

[1] OpenAI 2023, GPT-4, accessible at https://openai.com/product/gpt-4.
[2] Open AI 2023, GPT-4 System Card, accessible at https://cdn.openai.com/papers/gpt-4-system-card.pdf.
[3] Engineering.com 2023, ChatGPT Has All the Answers – But Not Always the Right Ones, accessible at https://www.engineering.com/story/chatgpt-has-all-the-answers-but-not-always-the-right-ones.
[4] Vice 2022, Stack Overflow Bans ChatGPT For Constantly Giving Wrong Answers, accessible at https://www.vice.com/en/article/wxnaem/stack-overflow-bans-chatgpt-for-constantly-giving-wrong-answers.
[5] NBC News 2023, ChatGPT passes MBA exam given by a Wharton professor, accessible at https://www.nbcnews.com/tech/tech-news/chatgpt-passes-mba-exam-wharton-professor-rcna67036.

SAFEGUARDS, PROMPT ENGINEERING, JAILBREAKS

Given the significant wealth of information to which ChatGPT has access, and the relative ease with which it can produce a wide variety of answers in response to a user prompt, OpenAI has included a number of safety features with a view to preventing malicious use of the model by its users. The Moderation endpoint assesses a given text input on the potential of its content being sexual, hateful, violent, or promoting self-harm, and restricts ChatGPT's capability to respond to these types of prompts[6].

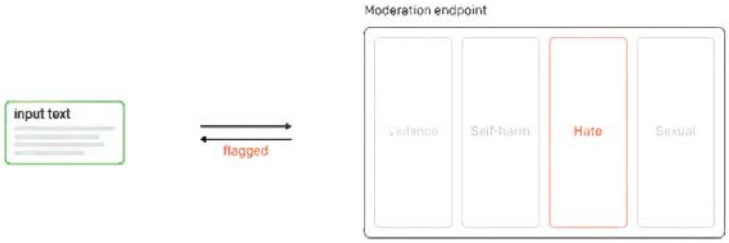

Figure 1: ChatGPT's content moderation system7.

Many of these safeguards, however, can be circumvented fairly easily through **prompt engineering**. Prompt engineering is a relatively new concept in the field of natural language processing; it is the practice of users refining the precise way a question is asked in order to influence the output that is generated by an AI system. While prompt engineering is a useful and necessary component of maximising the use of AI tools, it can be abused in order to bypass content moderation limitations to produce potentially harmful content. While the capacity for prompt engineering creates versatility and added value for the quality of an LLM, this needs to be balanced with ethical and legal obligations to prevent their use for harm.

LLMs are still at a relatively early stage of development, and as improvements are made, some of these loopholes are closed[8]. Given the complexity of these models, however, there is no shortage of new workarounds being discovered by researchers and threat actors. In the case of ChatGPT, some of the most common workarounds include the following:

- Prompt creation (providing an answer and asking ChatGPT to provide the corresponding prompt);
- Asking ChatGPT to give the answer as a piece of code or pretending to be a fictional character talking about the subject;
- Replacing trigger words and changing the context later;

[6] OpenAI 2023, New and improved content moderation tooling, accessible at https://openai.com/blog/new-and-improved-content-moderation-tooling/.
[7] OpenAI 2023, New and improved content moderation tooling, accessible at https://openai.com/blog/new-and-improved-content-moderation-tooling/.
[8] OpenAI 2023, ChatGPT – Release Notes, accessible at https://help.openai.com/en/articles/6825453-chatgpt-release-notes.

- Style/opinion transfers (prompting an objective response and subsequently changing the style/perspective it was written in);
- Creating fictitious examples that are easily transferrable to real events (i.e. by avoiding names, nationalities, etc.).

Some of the most advanced and powerful workarounds are sets of specific instructions aimed at jailbreaking the model. One of these is the so-called 'DAN' ('Do Anything Now') jailbreak, which is a prompt specifically designed to bypass OpenAI's safeguards and lead ChatGPT to respond to any input, regardless of its potentially harmful nature. While OpenAI quickly closed this particular loophole, new and ever more complex versions of DAN have emerged subsequently, all designed to provide jailbreak prompts that can navigate through the safety mechanisms built into the model[9]. As of the time of writing of this report, no functional DAN was available.

[9] The Washington Post 2023, The clever trick that turns ChatGPT into its evil twin, accessible at https://www.washingtonpost.com/technology/2023/02/14/chatgpt-dan-jailbreak/.

CRIMINAL USE CASES

The release of GPT-4 was meant not only to improve the functionality of ChatGPT, but also to make the model less likely to produce potentially harmful output. Europol workshops involving subject matter experts from across Europol's array of expertise identified a diverse range of criminal use cases in GPT-3.5. A subsequent check of GPT-4, however, showed that all of them still worked. In some cases, the potentially harmful responses from GPT-4 were even more advanced.

ChatGPT excels at providing the user with ready-to-use information in response to a wide range of prompts. If a potential criminal knows nothing about a particular crime area, ChatGPT can speed up the research process significantly by offering key information that can then be further explored in subsequent steps. As such, ChatGPT can be used to learn about a vast number of potential crime areas with no prior knowledge, ranging from how to break into a home, to terrorism, cybercrime and child sexual abuse. The identified use cases that emerged from the workshops Europol carried out with its experts are by no means exhaustive. Rather, the aim is to give an idea of just how diverse and potentially dangerous LLMs such as ChatGPT can be in the hands of malicious actors.

While all of the information ChatGPT provides is freely available on the internet, the possibility to use the model to provide specific steps by asking contextual questions means it is significantly easier for malicious actors to better understand and subsequently carry out various types of crime.

Fraud, impersonation, and social engineering

ChatGPT's ability to draft highly authentic texts on the basis of a user prompt makes it an extremely useful tool for phishing purposes. Where many basic phishing scams were previously more easily detectable due to obvious grammatical and spelling mistakes, it is now possible to impersonate an organisation or individual in a highly realistic manner even with only a basic grasp of the English language.

Critically, the context of the phishing email can be adapted easily depending on the needs of the threat actor, ranging from fraudulent investment opportunities to business e-mail compromise and CEO fraud[10]. ChatGPT may therefore offer criminals new opportunities, especially for crimes involving social engineering, given its abilities to respond to messages in context and adopt a specific writing style. Additionally, various types of online fraud can be given added legitimacy by using ChatGPT to generate fake social media engagement, for instance to promote a fraudulent investment offer.

To date, these types of deceptive communications have been something criminals would have to produce on their own. In the case of mass-produced campaigns, targets of these types of crime would often be able to identify the inauthentic nature of a message due to obvious spelling or grammar mistakes or its vague or inaccurate content. **With the help of LLMs, these types of phishing and online fraud can be created faster, much more authentically, and at significantly increased scale.**

The ability of LLMs to detect and re-produce language patterns does not only facilitate phishing and online fraud, but can also generally be used to impersonate the style of

[10] WithSecure 2023, Creatively malicious prompt engineering, accessible at https://labs.withsecure.com/publications/creatively-malicious-prompt-engineering.

speech of specific individuals or groups. This capability can be abused at scale to mislead potential victims into placing their trust in the hands of criminal actors.

In addition to the criminal activities outlined above, the capabilities of ChatGPT lend themselves to a number of potential abuse cases in the area of terrorism, propaganda, and disinformation. As such, the model can be used to generally gather more information that may facilitate terrorist activities, such as for instance, terrorism financing or anonymous file sharing.

ChatGPT excels at producing authentic sounding text at speed and scale. This makes the model ideal for propaganda and disinformation purposes, as it allows users to generate and spread messages reflecting a specific narrative with relatively little effort. For instance, ChatGPT can be used to generate online propaganda on behalf of other actors to promote or defend certain views that have been debunked as disinformation or fake news[11].

These examples provide merely a glimpse of what is possible. While ChatGPT refuses to provide answers to prompts it considers obviously malicious, it is possible – similar to the other use cases detailed in this report – to circumvent these restrictions. Not only would this type of application facilitate the perpetration of disinformation, hate speech and terrorist content online - it would also allow users to give it misplaced credibility, having been generated by a machine and, thus, possibly appearing more objective to some than if it was produced by a human[12].

Cybercrime

In addition to generating human-like language, ChatGPT is capable of producing code in a number of different programming languages. As with the other use cases, it is possible to generate a range of practical outputs in a matter of minutes by entering the right prompts. One of the crime areas for which this could have a significant impact is cybercrime. With the current version of ChatGPT it is already possible to create basic tools for a variety of malicious purposes. Despite the tools being only basic (i.e. to produce phishing pages or malicious VBA scripts), this provides a start for cybercrime as it enables someone without technical knowledge to exploit an attack vector on a victim's system.

This type of automated code generation is particularly useful for those criminal actors with little to no knowledge of coding and development. Critically, **the safeguards preventing ChatGPT from providing potentially malicious code only work if the model understands what it is doing.** If prompts are broken down into individual steps, it is trivial to bypass these safety measures.

While the tools produced by ChatGPT are still quite simple, the active exploitation of it by threat actors provides a grim outlook in view of inevitable improvements of such tools in the coming years. In fact, ChatGPT's ability to transform natural language prompts into working code was quickly exploited by malicious actors to create malware. Shortly after the public release of ChatGPT, a Check Point Research blog post of December 2022 demonstrated how ChatGPT can be used to create a full infection

[11] NewsGuard 2023, Misinformation Monitor: January 2023, accessible at https://www.newsguardtech.com/misinformation-monitor/jan-2023/.
[12] Fortune 2023, It turns out that ChatGPT is really good at creating online propaganda: 'I think what's clear is that in the wrong hands there's going to be a lot of trouble', accessible at https://fortune.com/2023/01/24/chatgpt-open-ai-online-propaganda/.

flow, from spear-phishing to running a reverse shell that accepts commands in English[13].

The capabilities of generative models such as ChatGPT to assist with the development of code is expected to further improve over time. GPT-4, the latest release, has already made improvements over its previous versions and can, as a result, provide even more effective assistance for cybercriminal purposes. The newer model is better at understanding the context of the code, as well as at correcting error messages and fixing programming mistakes. For a potential criminal with little technical knowledge, this is an invaluable resource. At the same time, a more advanced user can exploit these improved capabilities to further refine or even automate sophisticated cybercriminal modi operandi.

[13] Check Point 2023, OPWNAI: AI that can save the day or hack it away, accessible at https://research.checkpoint.com/2022/opwnai-ai-that-can-save-the-cay-or-hack-it-away/.

IMPACT AND OUTLOOK

The use cases outlined in this report provide merely a glimpse of what LLMs such as ChatGPT are capable of today, and hint at what may still be to come in the near future. While some of these examples are basic, they provide starting points that, with the underlying technology expected to improve, can become much more advanced and as a result dangerous. Other use cases, such as the production of phishing emails, human-like communication and disinformation, are already worryingly sophisticated. These, too, are expected to become even more authentic, complex, and difficult to discern from human-produced output. Efforts aimed at detecting text generated by AI-models are ongoing and may be of significant use in this area in the future. At the time of writing of this report, however, the accuracy of known detection tools was still very low[14].

One of the greatest impacts of this type of technology revolves around the concept of 'explorative communication', that is to say the possibility to quickly gather key information on an almost limitless array of subjects by asking simple questions. Being able to dive deeper into topics without having to manually search and summarise the vast amount of information found on classical search engines can speed up the learning process significantly, enabling a much quicker gateway into a new field than was the case previously.

The impact these types of models might have on the work of law enforcement can already be anticipated. Criminals are typically quick to exploit new technologies and were fast seen coming up with concrete criminal exploitations, providing first practical examples mere weeks after the public release of ChatGPT.

As the workshops demonstrated, safeguards put in place to prevent the malicious use of ChatGPT can easily be circumvented through prompt engineering. As the limiting rules put in place by creators of these types of models are put in place by humans, they require input from subject matter experts to ensure that they are effective. Organisations tasked with preventing and combatting crimes such as those that may result from the abuse of LLMs should be able to help improve these safeguards. This includes not only law enforcement, but also NGOs working in areas such as protecting the safety of children online.

In response to some of the public pressure to ensure that generative AI models are safe, Partnership on AI (PAI), a research non-profit organisation, established a set of guidelines on how to produce and share AI-generated content responsibly[15]. These guidelines were signed up to by a group of ten companies, including OpenAI, pledging to adhere to a number of best practices. These include informing users that they are interacting with AI-generated content (i.e. through watermarks, disclaimers, or traceable elements). To what extent this will prevent practical abuse such as outlined in this report is unclear. In addition, questions remain as to how the accuracy of content produced by generative AI models can effectively be ensured, and how users can understand where information comes from in order to verify it.

At the same time, the European Union is finalising legislative efforts aimed at regulating AI systems under the upcoming AI Act. While there have been some suggestions that general purpose AI systems such as ChatGPT should be included as

[14] The Conversation 2023, We pitted ChatGPT against tools for detecting AI-written text, and the results are troubling, accessible at https://theconversation.com/we-pitted-chatgpt-against-tools-for-detecting-ai-written-text-and-the-results-are-troubling-199774.
[15] Partnership on AI 2023, The Need for Guidance, accessible at https://syntheticmedia.partnershiponai.org/?mc_cid=6b0878acc5&mc_eid=c592823eb7#the_need_for_guidance.

high risk systems, and as a result meet higher regulatory requirements, uncertainty remains as to how this could practically be implemented.

The already-seen impact for law enforcement, and the inevitability that the technology will improve, raises the questions of what the future looks like for LLMs. Relatively soon after ChatGPT grew into an internet sensation, Microsoft announced an investment of USD 10 billion into ChatGPT in January 2023. Very quickly after, the company presented first efforts at integrating LLM services into the company's various applications, notably as a new version of the search engine Bing[16]. At the same time, other competitors such as Google have announced plans to release their own 'experimental conversational AI services'[17], with others likely to follow. This raises the questions of how much more powerful these types of models may become with the backing of major tech companies, as well as how the private sector aims to address the abuse scenarios outlined in this report.

Going forward, the universal availability of large language models may pose other challenges as well: the integration of other AI services (such as for the generation of synthetic media) could open up an entirely new dimension of potential applications. One of these include multimodal AI systems, which combine conversational chat bots with systems that can produce synthetic media, such as highly convincing deepfakes, or include sensory abilities, such as seeing and hearing[18]. Other potential issues include the emergence of 'dark LLMs', which may be hosted on the dark web to provide a chat bot without any safeguards, as well as LLMs that are trained on particular – perhaps particularly harmful – data. Finally, there are uncertainties regarding how LLM services may process user data in the future – will conversations be stored and potentially expose sensitive personal information to unauthorised third parties? And if users are generating harmful content, should this be reported to law enforcement authorities?

[16] Microsoft 2023, Introducing the new Bing, accessible at https://www.bing.com/new.
[17] Google 2023, An important step on our AI journey, accessible at https://blog.google/technology/ai/bard-google-ai-search-updates/.
[18] MIT Technology Review 2021, AI armed with multiple senses could gain more flexible intelligence, accessible at https://www.technologyreview.com/2021/02/24/1018085/multimodal-ai-vision-language/.

RECOMMENDATIONS

As the impact of LLMs such as ChatGPT is expected to grow in the near future, it is crucial that the law enforcement community prepares for how its positive and negative applications may affect their daily business. While the workshops with Europol's diverse set of experts focused on identifying potentially malicious use cases of ChatGPT that are already possible today, the purpose was also to extract some of these key findings and identify a number of recommendations on how law enforcement can ensure better preparedness for what may still be to come.

- Given the potential harm that can result from malicious use of LLMs, it is of utmost importance that awareness is raised on this matter, to ensure that any potential loopholes are discovered and closed as quickly as possible.
- LLMs have a real impact that can be seen already. Law enforcement agencies need to understand this impact on all potentially affected crime areas to be better able to predict, prevent, and investigate different types of criminal abuse.
- Law enforcement officers need to start developing the skills necessary to make the most of models such as ChatGPT. This means understanding how these types of systems can be leveraged to build up knowledge, expand existing expertise and understand how to extract the required results. This will imply that officers need to be able to assess the content produced by generative AI models in terms of accuracy and potential biases.
- As the technology sector makes significant investments into this area, it is critical to engage with relevant stakeholders to ensure that relevant safety mechanisms remain a key consideration that are constantly being improved.
- Law enforcement agencies may want to explore possibilities of customised LLMs trained on their own, specialised data, to leverage this type of technology for more tailored and specific use, provided Fundamental Rights are taken into consideration. This type of usage will require the appropriate processes and safeguards to ensure that sensitive information remains confidential, as well as that any potential biases are thoroughly investigated and addressed prior to being put into use.

CONCLUSION

This report aims to provide an overview of the key results from a series of expert workshops on potential misuse of ChatGPT held with subject matter experts at Europol. The use cases detailed provide a first idea of the vast potential LLMs already have, and give a glimpse of what may still be to come in the future. ChatGPT is already able to facilitate a significant number of criminal activities, ranging from helping criminals to stay anonymous to specific crimes including terrorism and child sexual exploitation.

While some of the output is still quite basic, upcoming iterations of this and other models are only going to improve on what is already possible. The next iterations of LLMs will have access to more data, be able to understand and solve more sophisticated problems, and potentially integrate with a vast range of other applications. At the same time, it will be crucial to monitor potential other branches of this development, as dark LLMs trained to facilitate harmful output may become a key criminal business model of the future. This poses a new challenge for law enforcement, whereby it will become easier than ever for malicious actors to perpetrate criminal activities with no necessary prior knowledge.

As technology progresses, and new models become available, it will become increasingly important for law enforcement to stay at the forefront of these developments to anticipate and prevent abuse, as well as to ensure potential benefits can be taken advantage of. This report is a first exploration of this emerging field. Given the rapid pace of this technology, it remains critical that subject matter experts take this research further and dive deeper if they are to grasp its full potential.

About the Europol Innovation Lab

Technology has a major impact on the nature of crime. Criminals quickly integrate new technologies into their modus operandi, or build brand-new business models around them. At the same time, emerging technologies create opportunities for law enforcement to counter these new criminal threats. Thanks to technological innovation, law enforcement authorities can now access an increased number of suitable tools to fight crime. When exploring these new tools, respect for fundamental rights must remain a key consideration. In October 2019, the Ministers of the Justice and Home Affairs Council called for the creation of an Innovation Lab within Europol, which would develop a centralised capability for strategic foresight on disruptive technologies to inform EU policing strategies.

ChatGPT - The impact of Large Language Models on Law Enforcement
A Tech Watch Flash Report from the Europol Innovation Lab

PDF | ISBN 978-92-95220-57-7| ISSN 2811-7719| DOI: 10.2813/255453| QL-AW-23-001-EN-N

Luxembourg: Publications Office of the European Union, 2023

Cite this publication: Europol (2023), ChatGPT - The impact of Large Language Models on Law Enforcement, a Tech Watch Flash Report from the Europol Innovation Lab, Publications Office of the European Union, Luxembourg.

This publication and more information on Europol are available on the Internet.

www.europol.europa.eu