CHATGPT

Impacts
on
Law Enforcement

August 2023

INTERPOL

BACKGROUND PAPER

INTERPOL Innovation Centre

BACKGROUND PAPER

ChatGPT:

IMPACTS ON LAW ENFORCEMENT

1 Introduction

ChatGPT is an artificial intelligence (AI) language model that is capable of generating human-like text. It was developed by OpenAI,[1] an AI research organization, and is one of the most advanced language models currently available. To put it simply, ChatGPT is a computer program that has been trained on a vast amount of text data, such as books, articles, and web pages. ChatGPT has been taught how to understand the patterns and structures of language, and to generate new text that is similar in style and tone. ChatGPT is an interactive AI interface able to engage in conversations by addressing follow-up questions, recognizing its own errors, questioning inconsistencies, and rejecting certain requests.[2] It is an AI chatbot system capable of 'understanding natural human language' and providing information and solutions to complex questions.[3]

'GPT' stands for 'Generative Pre-trained Transformer', aptly describing the ability of the AI chatbot system to produce written text in different forms (e.g., articles, poems, essays, speeches) by using huge quantities of text collected from the internet.[4] It is an AI language model that is based on a type of neural network called a "transformer", which is designed to process sequential data such as text. The GPT model is "pre-trained", which means that it is trained on a large amount of text-based data before it is adapted for a specific task.

Due to the accessibility and power of ChatGPT and its potential to permeate every-day information-gathering activities in the same way as search engines do, law enforcement agencies need to anticipate future uses of ChatGPT and similar applications by criminals committing crimes and law enforcement investigating and prosecuting them. This background paper aims to provide an overview of ChatGPT's functionalities, uses and limitations and make recommendations for law enforcement.

---

[1] OpenAI LP is governed by the board of the OpenAI non-profit, comprised of OpenAI LP employees. https://openai.com/about/

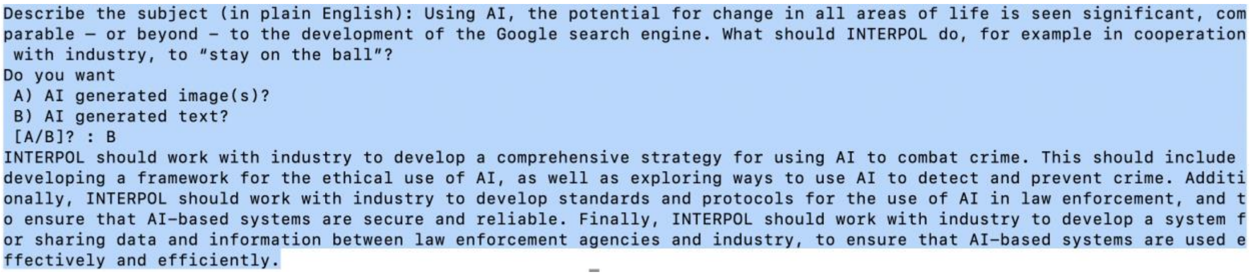
[2] https://openai.com/blog/chatgpt/

[3] https://www.theguardian.com/technology/2022/dec/05/what-is-ai-chatbot-phenomenon-chatgpt-and-could-it-replace-humans

[4] https://www.police1.com/chiefs-sheriffs/articles/will-ai-chatbots-power-the-future-of-police-language-translation-R8Ve7qQX26VReXGv/; https://www.cnet.com/tech/computing/why-the-chatgpt-ai-chatbot-is-blowing-everyones-mind/

2 ChatGPT Functionalities

ChatGPT offers an easy entry-level user access to AI platforms through a common interface using conversational prompts to generate results. Users need to register on OpenAI's website to obtain the necessary keys to configure its services. OpenAI offers a range of tutorials and documentation to help programmers interact with OpenAI platforms. The main models available generate texts, images, or computer programmes in response to plain language questions.



```
Describe the subject (in plain English): Using AI, the potential for change in all areas of life is seen significant, com
parable – or beyond – to the development of the Google search engine. What should INTERPOL do, for example in cooperation
 with industry, to "stay on the ball"?
Do you want
 A) AI generated image(s)?
 B) AI generated text?
 [A/B]? : B
INTERPOL should work with industry to develop a comprehensive strategy for using AI to combat crime. This should include
developing a framework for the ethical use of AI, as well as exploring ways to use AI to detect and prevent crime. Additi
onally, INTERPOL should work with industry to develop standards and protocols for the use of AI in law enforcement, and t
o ensure that AI-based systems are secure and reliable. Finally, INTERPOL should work with industry to develop a system f
or sharing data and information between law enforcement agencies and industry, to ensure that AI-based systems are used e
ffectively and efficiently.
```

Figure 1. A simple ChatGPT (3.5) integration using Python and OpenAI APIs.

To understand how AI works here, it is important to look at how the platform has been trained to operate. This is achieved in three steps:[5]

Step 1: Collect demonstration data and train a supervised policy.
Step 2: Collect comparison data and train a reward model.
Step 3: Optimize a policy against the reward model.

3 ChatGPT Use by Law Enforcement

The OpenAI platform is very versatile but common uses include searching for, generating and modifying images, writing content such as poems, songs, essays, and even blogs, and debugging code or getting Codex to write it. Overall, the GPT model represents a major breakthrough in the development of AI language models; its various versions have pushed the boundaries of what is possible with text-based AI and have opened up new opportunities for innovation and discovery. Some potential use cases of ChatGPT for law enforcement are given below. **However, they are purely hypothetical scenarios which require individual assessment and safeguards to be implemented by law enforcement and judicial authorities planning to use ChatGPT.**

ChatGPT should be used not only "in accordance with legal and ethical guidelines" but also in compliance with the relevant laws, regulations and internal policies and procedures of each law enforcement and judicial authority. In addition to national laws - and especially in the absence of specific AI laws and policies, all uses of ChatGPT must comply with international standards on responsible AI, adhering to principles such as lawfulness, minimization of harm, fairness, respect of human autonomy, and good governance.[6] ChatGPT should never replace human judgement and the final responsibility for the accuracy and quality of ChatGPT outputs and the decisions taken must rest with individual law enforcement officers who have the necessary training and expertise.

---

[5] https://openai.com/blog/chatgpt/

[6] INTERPOL and the United Nations Interregional Crime and Justice Research Institute (UNICRI) Toolkit for Responsible AI Innovation in Law Enforcement (see https://www.interpol.int/en/News-and-Events/News/2023/INTERPOL-and-UNICRI-release-blueprint-for-responsible-use-of-AI-by-law-enforcement#:~:text=INTERPOL%20and%20UNICRI%20launch%20the,and%20INTERPOL's%20Responsible%20AI%20Lab).

Translation: ChatGPT can be used to translate text from one language to another. This can be helpful in situations where language barriers may be a challenge. However, ChatGPT should not be relied upon without confirming and verifying the output, especially as regards translations that will be used to initiate a case, take action against an individual, process witness statements, or handle sensitive information.

Text data analysis: Law enforcement agencies often deal with large amounts of text data, such as emails, social media posts, and chat transcripts. ChatGPT can be used to analyse this data and extract insights that may be relevant to ongoing investigations.

Fraud detection: ChatGPT can be used to analyse text data and detect patterns of fraud or other criminal activity, such as phishing scams or fraudulent emails.

Training and education: ChatGPT can be used to provide training and education to law enforcement officers on topics such as de-escalation techniques, cultural awareness, and investigative methodologies.

Victim support: ChatGPT can be used to provide support and resources to victims of crime, such as information on legal rights and counselling services.

Investigative research: ChatGPT can be used to conduct investigative research, such as analysing online forums or social media groups to identify potential suspects or criminal activity.

Virtual assistants: ChatGPT can be used to provide virtual assistants which use natural language processing techniques to understand and respond to user queries.

It is important to note that particular care must be taken to ensure that the use of AI does not result in biases or violations of individual rights and privacy. When using ChatGPT for law enforcement purposes, agencies should exercise caution with regard to potential data confidentiality and disclosure issues. AI platforms like ChatGPT should not be used for sensitive police data as their providers may process this information on their servers for learning purposes. Additionally, it is likely that when ChatGPT is used to produce documents in a criminal case based on law enforcement input, court orders are issued to seek clarifications about the request made to ChatGPT by law enforcement.

OpenAI will soon switch to a fully commercial model for businesses that want to develop applications integrating OpenAI foundational layers into their products, while also having the ability to customize the models with proprietary data and additional AI features.[7] Microsoft and OpenAI recently announced their long-term partnership 'through a multiyear, multibillion dollar investment to accelerate AI breakthroughs to ensure these benefits are broadly shared with the world'.[8] It is expected that other global platform providers will start to create and roll out similar platforms in the near future as they see user engagement with such platforms increase.

---

[7] https://openai.com/api/pricing/

[8] https://blogs.microsoft.com/blog/2023/01/23/microsoftandopenaiextendpartnership/

Google and Alphabet have been developing similar AI layers (namely Bidirectional Encoder Representations from Transformers [BERT][9], Multitask Unified Model [MUM][10], and Language Model for Dialogue Applications [LaMDA][11]). Google recently unveiled a LaMDA-powered 'experimental conversational AI service' called 'Bard'.

4 Limitations of ChatGPT

An AI large language model like ChatGPT can have several limitations that can affect the quality and accuracy of its responses. These limitations are important to keep in mind when using ChatGPT as a tool for generating responses to text-based queries, especially in a law enforcement context.

Incomplete or outdated information: Although ChatGPT has been trained on a large text dataset, its performance is heavily dependent on the quality and relevance of its training data. Insufficient or low quality data can lead to poor performance and inaccurate responses. This means that it may have inadequate knowledge or understanding to provide accurate or complete answers to some questions, especially on very specific or rare topics. Moreover, ChatGPT may not be aware of recent events, updates, or advancements in various fields depending on the cut-off date of training data, which could also lead to outdated information.

Bias: Like any machine learning model, ChatGPT can also be biased towards certain groups, topics, or viewpoints depending on the training data it has been exposed to. ChatGPT's training data is derived from a wide range of sources, including social media, news articles, and books. This means that the model may contain biases from these sources in terms, for example, of gender, race, or culture which could potentially lead to responses that are inaccurate or discriminatory. It may sometimes generate responses that are unintentionally offensive or inappropriate.

Contextual understanding: ChatGPT may sometimes struggle to understand the nuances and context of a question or conversation, which can lead to responses that are irrelevant or inappropriate.

Vulnerability to adversarial attacks: ChatGPT can be vulnerable to attacks during which malicious users intentionally input incorrect or misleading information to manipulate the responses of the model.

Lack of legal expertise: While ChatGPT has been trained on a diverse range of content, it is not a legal expert. Its understanding of legal concepts, terminology, and procedures may be incomplete or incorrect so it should not be relied upon for professional legal advice.

Lack of professional judgment: Law enforcement officers are often required to write reports detailing their observations and actions which are generally submitted to the defence and used for cross-examination purposes. These reports – which often form the basis of a defence

---

[9] https://blog.google/products/search/search-language-understanding-bert/

[10] https://blog.google/products/search/introducing-mum/

[11] https://blog.google/technology/ai/lamda/

argument - are based on police officers' personal experiences or first-hand knowledge of events and are shaped by their judgment of the most appropriate information to include. This cannot be replicated by ChatGPT. Police reports are legal documents that require accuracy and impartiality (which ChatGPT does not necessarily possess due to its training dataset), and compliance with specific guidelines and procedures. The final responsibility for the accuracy and quality of police reports shall always remain with police officers who have the necessary training and expertise.

5 Awareness and Caution

Generative AI and large language models such as ChatGPT have many potential benefits. However, it is important to be aware of potential misuses, remain vigilant, identify potential vulnerabilities and take preventive action.

It is the duty of organizations using large language models such as ChatGPT to ensure ethical and responsible use of this technology. Since the area of generative AI is still unregulated with little or no moderation, it provides a new breeding ground for the expansion of existing and emergence of future criminal enterprises.

FRAUD, SCAMS AND IMPERSONATION

- **The public should stay alert and keep in mind that deepfake technology can generate realistic images and videos and be used to create fake content and identities.**
- **For example, AI-generated automated spam messages are used for phishing scams. Criminals could create convincing scams or phishing attempts by mimicking the voice of a trusted individual or organization to trick users into providing sensitive information.**
- **When using emails and social media platforms, users should be aware of the risk of fake profiles or fraudulent messages which can be used to steal personal information or money.**
- **Malicious chatbots attempting to impersonate real people or organizations can spread malware or gather sensitive information.**

MALWARE

- **Users should always be careful when clicking on links and downloading attachments as AI-generated links can be used to spread malware or files that install viruses or malicious software on users' devices.**

MISINFORMATION, PROPAGANDA AND MANIPULATION OF PUBLIC OPINION

- **Users should always be sceptical about content read online. AI-generated content can be used to spread false information or propaganda, either intentionally or unintentionally.**

- **Spreading misinformation and generating fake news using large language models could have catastrophic societal consequences, impacting public health for example.**
- **Users should be mindful that some messages and content are specifically created to promote extremist ideologies and incite violence leading to social unrest and fuelling terrorism.**
- **Fake reviews or testimonials for products or services can mislead consumers and harm genuine businesses.**

Criminals have only recently started using AI platforms for their illicit purposes and, like all other users, are getting better at using the correct expressions and questions to generate the desired results.

6 Role of Law Enforcement

AI is becoming increasingly accessible and will become more prevalent in cybercrime. Law enforcement needs to have tools and capabilities able to detect ChatGPT, or AI generated content and be able to share these tools and possibly the signature of identified AI-generated content. A number of such AI-generated text detection tools are being developed, for instance, GPTZero, Huggin Face GPT2, and Writer AI detector. These solutions need to be verified and benchmarked prior to being widely adopted by law enforcement. Law enforcement in INTERPOL member countries needs to be well prepared and ensure that ChatGPT and similar platforms are used for the common good as well as flagging criminal uses; it is recommended that the points described below be adopted and that a mechanism to share, receive, and exchange information from member countries be implemented.

1) **Technology alignment**
   For platforms that use AI and associated processes, it is important that INTERPOL member countries' law enforcement investigators, forensics specialists, and prosecutors all have the right level of know-how to ensure these technologies are applied in the appropriate manner in investigations.
2) **Standard investigative processes and procedures**
   It is also important for law enforcement agencies to work with industry in developing standard processes and procedures to address transnational crimes. A clear understanding of the shared needs of law enforcement and industry partners and reciprocal solutions is essential.
3) **Standardized training and education**
   There is huge interest in conversational AI platforms and associated technologies and processes because of their advanced functionalities, so it is vital to allocate resources and provide support to create training and education packages. INTERPOL member countries need to be aware of and familiar with these platforms and processes in order to successfully address these new challenges.
4) **Horizon scanning and foresight**
   INTERPOL aims to ensure that member countries are kept abreast of the latest advancements in this fast-evolving AI technology and the implementation of AI-enabled platforms and applications. Member countries are encouraged to exchange information with INTERPOL in order to detect early signs of major developments through a systematic examination of potential threats and opportunities, with an emphasis on the effects and implications for law enforcement.

5) **Clear rules and regulations**
With transnational crimes linked to the use of ChatGPT and similar platforms on the rise, it is crucial for INTERPOL member countries to start the discussion about identifying appropriate regulations that could be applied to different aspects of preventing and fighting crime.

6) **Responsible use of AI-enabled platforms by law enforcement**
AI is an incredibly promising technology and a tool which could offer tremendous benefits to law enforcement work. However, given the complex nature of the subject, and the importance of public trust in law enforcement, AI-enabled platforms need to be responsibly applied in the law enforcement environment. To this end, INTERPOL worked with the United Nations Interregional Crime and Justice Research Institute (UNICRI) to develop a *Toolkit for Responsible AI Innovation in Law Enforcement* published in June 2023. The guide aims to support law enforcement agencies around the world regarding the responsible design, development, procurement, and deployment of AI-enabled tools/platforms.

As AI-enabled chatbot systems and platforms such as ChatGPT are continuously evolving - in parallel with their implementation in law enforcement work - INTERPOL will continue monitoring developments by working closely with law enforcement agencies in member countries, industry and academic experts, and ensure a cohesive and collaborative approach is applied.

DISCLAIMER

The contents of this Background Paper are for information purposes only. INTERPOL assumes no liability or responsibility for any inaccurate, delayed or incomplete information, nor for any actions taken in reliance thereon. The published material is distributed without warranty of any kind, either express or implied, and the responsibility for the interpretation and use of the material lies with the reader.

In no event shall INTERPOL be liable for damages arising from its use. INTERPOL takes no responsibility for the content of any external website referenced in this publication or for any defamatory, offensive or misleading information which might be contained on these third party websites. Any links to external websites do not constitute an endorsement by INTERPOL, and are only provided as a convenience. It is the responsibility of the reader to evaluate the content and usefulness of information obtained from other sites.

The views, thoughts and opinions expressed in the content of this publication belong solely to the authors and do not necessarily reflect the views or policies of INTERPOL, its member countries, its governing bodies, or contributory organizations, nor does it imply any endorsement. Therefore, INTERPOL carries no responsibility for the opinions expressed in this publication.

INTERPOL does not endorse or recommend any product, process, or service. Therefore, mention of any products, processes, or services in this publication cannot be construed as an endorsement or recommendation by INTERPOL.

This publication must not be reproduced in whole or in part or in any form without special permission from the copyright holder. When the right to reproduce this publication is granted, INTERPOL would appreciate receiving a copy of any publication that uses it as a source.

INTERPOL

For questions or feedback regarding
this Background Paper, please contact
INTERPOL Innovation Centre at
innovation@interpol.int