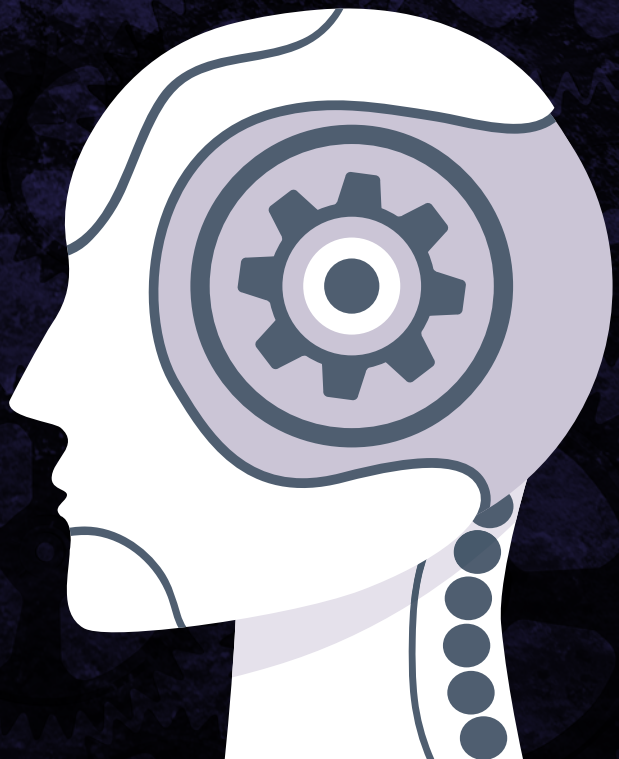
Government
Counter Fraud
Profession

Introduction to AI Guide

with a focus on Counter Fraud



Glossary of Terms

Key terms and meanings[1]

Artificial Intelligence (AI)	Machine driven capability to achieve a goal by performing cognitive tasks.
Generative AI	AI systems that create new content e.g. ChatGPT, generate text and images from text prompts - some use images to create audio and video content.
Large Language Models	Models trained on large volumes of text based data - typically from the internet.
Voice Cloning	Use of AI technology to create a simulation of a person's voice.
Deep fake	Videos or images that use a form of AI to digitally manipulate existing content e.g. replacing images of faces with someone's likeness. Deep fake can also be known as synthetic media
Ethical AI	Used to indicate the development, deployment and use of AI that ensures compliance with ethical norms, including fundamental rights as special moral entitlements, ethical principles and related core values. It is the second of the three core elements necessary for achieving Trustworthy AI.
Machine Learning	The use of algorithms that find patterns in data without explicit instruction. A system might learn how to associate features of inputs such as images with outputs such as labels.

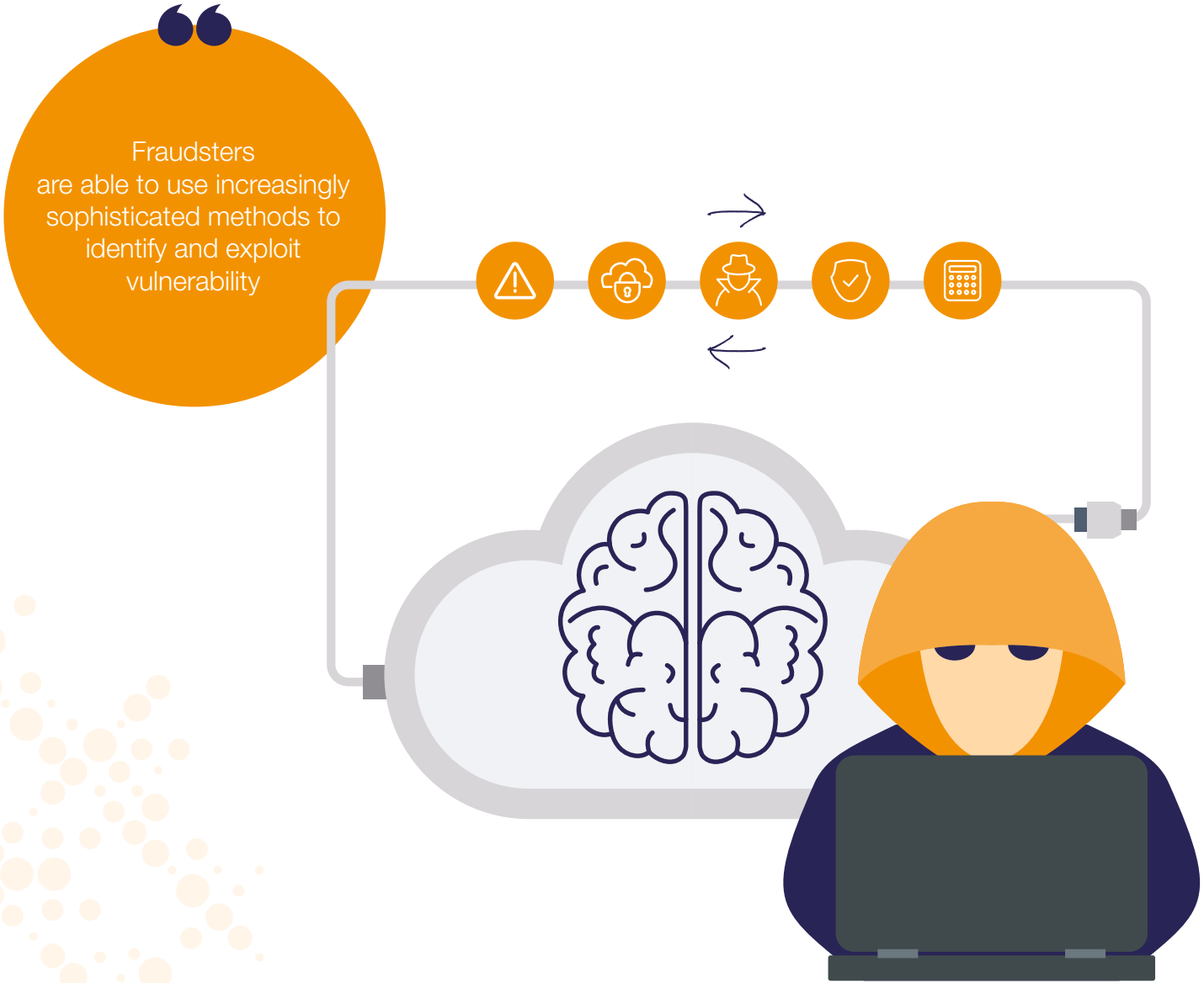
1 Partly sourced NCA, GenAI Threat Assessment 2023

Introduction

Technology does not stand still, whether we consider this from a counter fraud perspective or the view of the fraudsters we face. When technology evolves it can be harnessed by fraud practitioners to great benefit, but equally criminals and fraudsters will work at pace to embed these advances into their toolkits to attack systems and processes.

Fraudsters are able to use increasingly sophisticated methods, relying on the systematic analysis of large amounts of data in an effort to identify and exploit vulnerabilities that might exist in our organisations for their own gain.

Letting fraudsters lead the way in the use of Artificial Intelligence (AI) technology is not an option- so it is our collective role and responsibility in counter fraud to keep pace with developments and understand the impact and potential fraud threats they may bring and understand the opportunities that may arise. This short guide introduces, and we hope demystifies, AI, and signposts you onwards to build your knowledge and awareness.

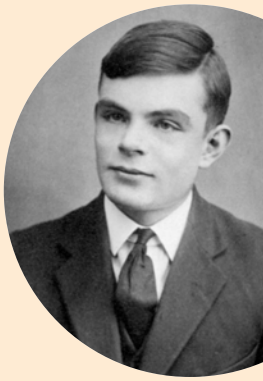


The rise of AI

Artificial intelligence is not new but we have seen accelerated coverage in the media and as a hot topic at public and private sector events in recent years. This is because access to AI tools has become more commonplace. GenAI platforms like ChatGPT are now widely available and used by the public in a variety of ways.

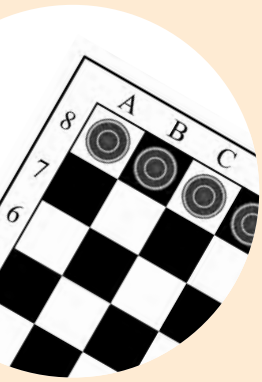
The rise of Artificial Intelligence does present a huge opportunity for those working in the public sector to detect and prevent fraud, at pace, using large quantities of information data. This aligns and supports the modern fraud approach which focuses on a deep understanding of risk and the use of data and intelligence to find fraud and irregular payments. When using data and AI it is important that users consider potential strategic, operational and reputational risks that may arise if key principles, ethical considerations and data management processes are not adhered to.

A brief timeline of AI[2]



1950

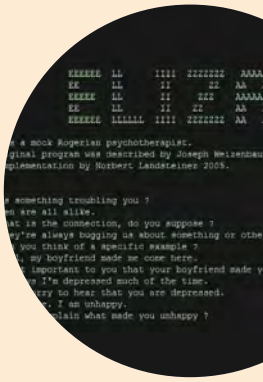
Alan Turing - Computer machinery and intelligence - ***'Turing Test'***



1956

Arthur Samuel IBM 701 invent checkers game ***'machine learning'***

John McCarthy 'the facts of AI' and Marvin Minsky logical theorist, coins ***'artificial intelligence'***



1966

'Eliza'- first chatbot by Joseph Wiezenbaum

1997

Deep Blue IBM chess computer wins v's world champion Garry Kasparov

Dragon software introduced e.g. ***'voice recognition'***

2011

'Siri' intelligence speech assistant introduced in i-phone 4

2014

'Alexa' virtual assistant by Amazon, learning from queries

2020

GPT-3 is released - enabling automated conversations

2022

2022 Chat GPT released - AI explodes!

2 Source: No10DS, 'The Bluffers Guide to AI' Dr Laura Gilbert 2023

What is Artificial Intelligence (AI)?

AI can range from predictive algorithms and machine learning all the way through to complex robotics[3]. It can be defined as the use of digital technology to create systems capable of performing tasks commonly thought to require intelligence.

In terms of its relationship to us as humans, it can be regarded as 'a collection of interrelated technologies used to solve problems autonomously and perform tasks to achieve defined objectives without explicit guidance from a human being'.

It will involve some element of learning by that system, but that can be supervised or unsupervised machines using statistics to find patterns in large amounts of data; and the ability to perform repetitive tasks with data without the need for constant human guidance.

> At its simplest form, artificial intelligence is a field, which combines computer science and robust datasets, to enable problem-solving

Supervised vs Unsupervised learning

Supervised learning, also known as supervised machine learning, is a subcategory of machine learning and artificial intelligence. It is defined by its use of labelled datasets to train algorithms to classify data or predict outcomes accurately.

Labelled data contains meaningful tags and unlabelled data does not contain any additional information. It is essentially raw data before any labels are applied.

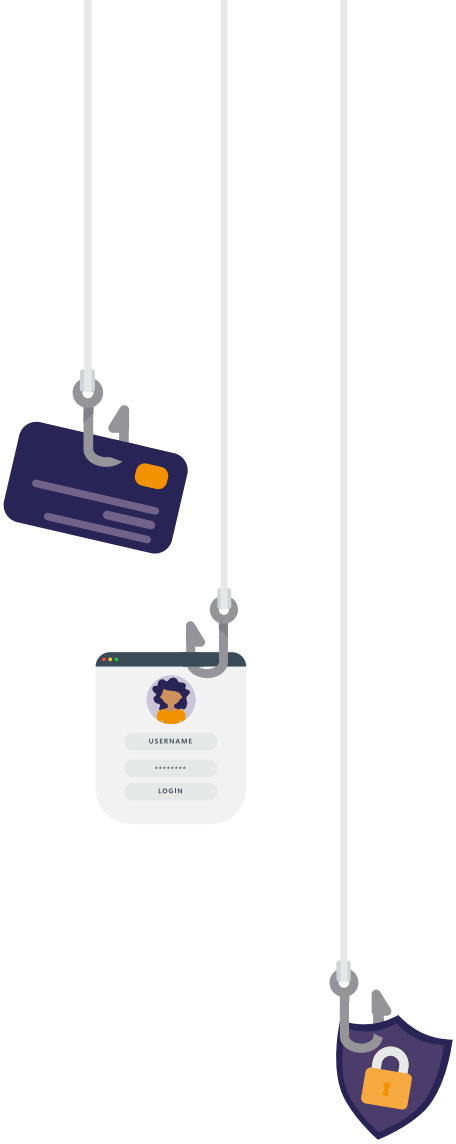
Supervised machine learning relies on labelled input and output training data, whereas unsupervised learning processes unlabelled or raw data.

Unsupervised learning in artificial intelligence is a type of machine learning that learns from data without human supervision. Unlike supervised learning, unsupervised machine learning models are given unlabelled data and allowed to discover patterns and insights without any explicit guidance or instruction.

3 Source: IPSFF guidance on AI, 2020

What is Generative AI (GenAI)?

Generative AI (GenAI) uses data and files online to create results that appear authentic to the audience, and these can include those created in the form of “language models”.[4] Language models are based on vast amounts of data, and they learn from these to generate outputs.



These can be used for good and may help to explore vast amounts of information and help to produce outcomes at a much more rapid pace than if attempted manually. For example language models can be used to summarise vast amounts of complex information. UK agencies like the Serious Fraud Office are already using this type of technology to support investigations and evidence review.

Fraudsters however, can use GenAI in an adverse way such as producing vast amounts of information that can be used to convince victims into handing over financial data and information, for example in the form of text or SMS (phishing or smishing) attacks. Gen AI models can learn from the patterns of information we input, and this can be used to generate data with similar characteristics.

> Generative AI models learn the patterns and structure of their input training data and then generate new data that has similar characteristics

> Language models can help to analyse and summarise vast amounts of data at pace

4 Source: ACFE introduction to AI, 2023

What is a deep fake?

This includes where data is used to mimic real on line interactions (such as a persons voice or image) that can have the illusion of being real.

Deep fakes are also known as synthetic media and can take the form of voice takeovers - these have been known to be used by fraudsters to take on identities and convince people into parting with money or information, or be used to open up credit or money transfer facilities. These are adopted to try and manipulate the controls used by organisations such as voice software recognition to verify authenticity of a user.

There have been examples of deep fakes being adopted by criminals to take on the persona of banking and government organisations, to convince victims to again handover personal identifiers, passwords and transfer over money.

1 Director cloned in elaborate fraud

The threat to individuals may feel dwarfed by the potential risks to business and corporations. One Japanese company lost $35 million after the voice of a company director was cloned–and used to pull off an elaborate fraud in 2020. The risks of this happening are increased now as AI tools for writing, voice impersonation and video manipulation are swiftly becoming more competent, more accessible and cheaper for even run-of-the-mill fraudsters.

In early 2020, a branch manager of a Japanese company in Hong Kong received a call from a man whose voice he recognized—the director of his parent business. The director had good news: the company was about to make an acquisition, so he needed to authorize some transfers to the tune of $35 million. A lawyer named Martin Zelner had been hired to coordinate the procedures and the branch manager could see in his inbox emails from the director and Zelner, confirming what money needed to move and to where. The manager, believing everything appeared legitimate, began making the transfers.

What he didn't know was that he'd been duped as part of an elaborate swindle, one in which fraudsters had used "deep voice" technology to clone the director's speech. The elaborate scheme was believed to involve at least 17 individuals, which sent the stolen money to bank accounts across the globe.

"Fraudsters had used "deep voice" technology to clone the director's speech

2 Romance fraud

Deep fakes are increasingly being used in romance scams[5] to trick victims into believing they are talking to a real person in order to steal large sums of money, a charity has warned.

Lisa Mills, relationship fraud expert at the UK charity Victim Support, said fraudsters have taken advantage of the latest deep fake technology to create video clips of themselves "manipulating victims into believing that they're real people". A fraudster, with whom the victim believed she was in a legitimate two-year relationship, used deep fake technology during video calls to steal £350,000 from her. The scammer, who met the victim on a dating website, had even proposed using a photo which had been digitally altered showing a man holding a sign that read: "Will you marry me?".

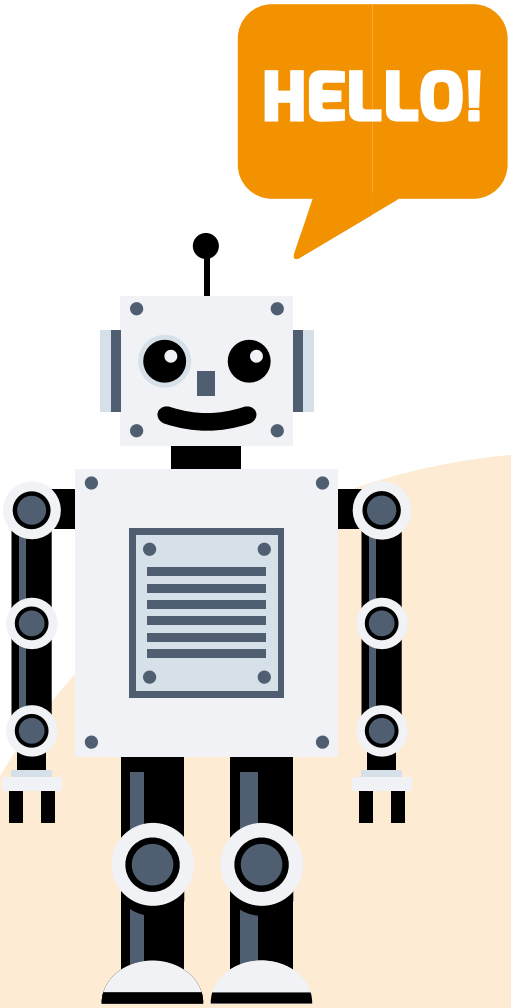
The victim, in their fifties, withdrew their pension pot early and even resorted to selling personal possessions after the fraudster convinced them they were being held hostage and tortured by loan sharks. "Aside from the financial aspect, the victims go through a lot of emotional stress because they feel like their boyfriend or girlfriend is in danger," said Ms Mills. She warned people that technology is getting "more sophisticated", with deep fakes set to become a "dangerous tactic in the fraudsters' toolkit".

Deep fakes – also known as "synthetic media" – are videos, images or audio files that use a form of artificial intelligence (AI) to digitally manipulate existing content, for example by replacing images of faces with someone else's likeness, to create fake events. As AI algorithms become increasingly sophisticated, it has become more difficult to distinguish fake content from reality.

3 AI Chatbots

AI Chatbots can be used to create fake online profiles that look like real people and talk to the victims. These chatbots can be so advanced to emotionally lure the victims into a relationship. They pretend to be real people and talk to the victims. They use special computer programs that can act like real people so they seem believable. These chatbots can also make victims feel a certain way so they are more likely to give out their money or personal information.

AI chatbots have the ability to be deployed at scale to engage millions of people and then detect the ones who can potentially fall into the romance scams.



5 Source: I-news, 2023, featuring Victim Support Charity

4 Voice Cloning

Voice cloning[6] is a technology that creates counterfeit conversations where an artificial voice imitates somebody's own voice.

As controversial as it may seem, cybercriminals have started taking advantage of this technology to deploy romance scams in dating websites and apps. The perpetrators have the tools to pretend to be someone else through digital impersonations – with nothing more than a pre-recorded conversation and some knowledge of the victim's life.

This advanced technology has been weaponised by internet scammers to deceive victims into believing they're speaking with someone they love or trust. The scammer then uses the reproduced voice to emotionally manipulate their victim into sending them money and sensitive information.

Knowing how to recognize a scammer is the best way to prevent becoming a victim of these types of fraud. Common indicators of a scam include fraudulent "emergency" requests for money or personal details, requests that you pay in a non-traditional manner such as gift cards, promises of large sums of money in return for minimal effort, and unexpected business offers.

LoveGPT, which combines OpenAI's ChatGPT with existing technology, is just one example of how generative artificial intelligence is used in scams. Such scams involve the use of LoveGPT to generate content to facilitate romantic connections, for use on dating sites or to target victims ultimately for financial gain (romance fraud). The content generated helps to convince the victim they are conversing with a genuine love interest.

The main goal is to create fake profiles on several dating platforms, while scraping data from interactions with the platforms' users, including their profile pictures, profile texts and dates of communication.

0 0 1 1 0 0 1 1 1 0 0 1 0 0 1 1 1 0 1 0
1 1 0 0 1 0 0 0 0 1 1 0 0 1 1 1 0 0 1 1
0 1 0 1 V O I C E 0 1 1 1 0 1 0 0 1 0 1
0 0 1 1 0 1 0 1 1 0 0 1 0 0 1 1 1 0 1 0
0 0 1 1 1 0 1 0 1 C L O N I N G 0 1 1 0
1 1 0 0 0 0 1 1 1 0 0 0 0 1 1 0 0 1 1 1
0 1 0 0 1 1 0 0 1 1 1 0 0 1 0 0 1 1 0 0
1 0 1 1 1 0 0 1 0 0 0 1 1 1 0 1 1 0 0 1

 Source, Avast, 2023

What are some of the challenges to consider?

We have already highlighted examples of where fraudsters are using AI to further their financial gains and other criminal pursuits. What are the challenges to be aware of and how are we responding[7] across the government?

The legislation and enforcement response is still catching up with the fast paced and emerging technology

- The Home Office are leading work with partners across industry to fight back and introduce preventative measures, legislative reform and controls to mitigate these risks, with the aim first and foremost to protect the public.

Fraudsters are evolving their modus operandi to attack and with the existence of AI has led to new methods in the application of these

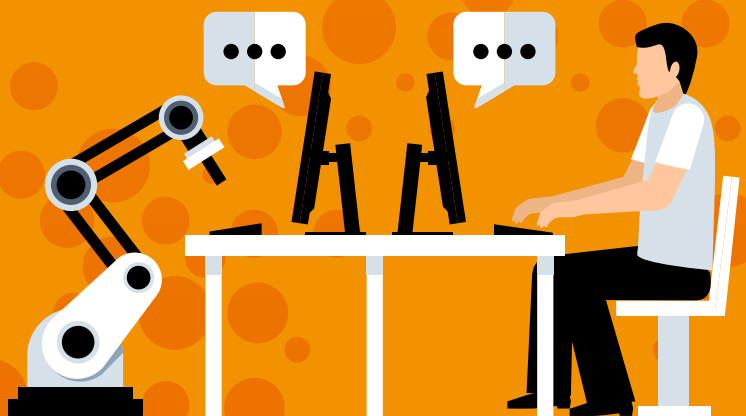
- Law enforcement partners are sharing across sectors emerging threats and intelligence to build understanding of the methods being deployed. This is supported in particular by products from the National Crime Agency (NCA), National Economic Crime Centre (NECC) and National Cyber Security Centre (NCSC).

Taking enforcement action when fraud occurs using AI techniques is difficult and not mature

- Work is ongoing across the public sector, with partnership working from the Public Sector Fraud Authority (PSFA), NCA, HM Revenue and Customs (HMRC), Home Office (HO) and others to determine what steps can be taken and then shared wider, to increase detection, prevention and recovery in this space.

AI allows perpetrators to act across jurisdictions, at pace and using realistic detail and information to defraud people

- Innovative work in the use of data analytics is live to be able to at pace identify criminal networks operating across departments so that action can be taken. This work is being led by the PSFA in collaboration with industry partners and law enforcement agencies.



7 Source NCA/PSFA 2023

Principles and considerations for use of AI to fight fraud[8]

AI can be used for beneficial purposes, such as machine learning and data analytics which can be used to sort data - to support and evaluate the quality.

It is expected that increasingly AI will be an aid for counter fraud including pre-investigation to:

- Identify lines of enquiry and during investigation to gather and organise evidence for disclosure.
- Summarise evidence to inform fraud measurement and risk assessment.
- Use patterns to detect and prevent fraud and irregularity upstream.
- To aid post event assurance.

In the future with the right data sets and business rules there could even be a role for AI in predicting fraud risks.

As we know from recent live cases that have played out in the UK media there is concern about sole reliance on AI, and therefore its important to consider how the wider application of technology can be utilised, while understanding and considering the principles to adopt.

Challenges and concerns

The use of AI and its increasing power and complexity, presents both opportunities and challenges. These include:

- the collection, transmission, processing, storage, and curation of potentially vast amounts of information that can be factored into decisions;
- the potential lack of transparency and understanding around machine classification algorithms and/or decision making processes; and
- the critical need for objectivity that must attach to the results.

> There are future opportunities to incorporate AI into the discipline of counter fraud

8 Source: IPSFF guide for AI, 2020

International guidance and insights

Back in 2020 the International Public Sector Fraud Forum (IPSFF), recognised the rise of AI presented an opportunity for the Public Sector to prevent and detect fraud.

The IPSFF guide centres on the following five central themes raised by the use of AI in combatting fraud:

- Accuracy
- Human control
- Transparency and explainability
- Fairness
- Privacy and civil liberties

The fraud experts from the Five Eyes countries came together to produce professional guidance. Our partners in the Serious Fraud Office New Zealand took the lead to develop The Use of Artificial Intelligence to Combat Public Sector Fraud Professional Guidance which covers the key considerations for utilising AI and technology advances to fight fraud. Further guidance is signposted at the end of the document but below are key highlights to introduce and broaden thinking in this space:

Key Considerations for the use of AI

Competencies

For public sector organisations. Recruitment and training in key disciplines, including natural language processing, data analytics, computer vision and machine learning alongside these competencies counter fraud knowledge will be essential.

Data governance

A lack of meaningful data sets and benchmarks to validate real-world performance as well as insufficient volume of labelled data for machine learning could slow the adoption of AI within the public sector. Recognising that the power of current AI technology is in the data, the more abundant and clean data available on fraud cases, the better the AI will perform. Data sharing between different platforms also raises debates on privacy, security, trust and accountability.

Fairness

It is important to ensure that when we are applying AI to business rules, processes and decisions we are considering up front any potential bias in the populations included, and particularly consider the application of any rules developed to vulnerable groups for example- “where an AI tool directs resources to a particular issue (for example fraud that is occurring in a certain sector or demographic) and then receives its ‘learning’ from that same issue, then the conclusions it reaches can be self-reinforcing” The risks outlined above can be mitigated somewhat where AI is used to support human decision making rather than replace it.

Checklist

- AI tools still require human input and if this information is flawed then outcomes will be affected.
- Agencies should not assume a level playing field and an AI system should have to prove it is correct.
- Inequities or biases (whether overt, latent, or historic) can be reinforced through the use of AI and the data fed into an AI system unless they are taken into account and normalised or corrected.
- A review process of any AI tool should consider the context for its use and those who may be most affected by it before drawing any definitive conclusions about its fairness or objectivity

> AI tools still require human input and oversight to safeguard their use

Key steps to consider before AI is used

Test	Before it is deployed, an AI tool should be tested against independent, and well understood data for accuracy.
Test again	Post-deployment, it should again be periodically tested and trained using quality, unbiased data (in certain cases, it may even need to be retrained). As part of this process, the items used to train the AI tool should be representative of the data on which the AI tool will be deployed.
Learn	There should be a process for regular gathering and curating of new training data so that the system does not become out of date or skew into unintended bias.
Scrutinise	The level of scrutiny applied before AI is deployed should reflect the fact that in the context of fraud detection or prevention, a punitive or intrusive intervention may follow.
Recruit	Agencies should ensure that they have appropriately qualified people to operate the AI tools and also, to the greatest extent possible, ensure that the data being analysed is meaningful. All developers should also understand the fraud risks and apply those in the deployment.

Human input

AI should be used to inform human decision making but should not entirely replace human oversight. The extent of oversight will depend on the significance of the decision and on other safeguards in place. Where a decision or selection being made about an individual is significant (its operation impacts benefits, freedom, or access to a service), careful consideration should be given to the level of human input required.

The use of AI should inform human decision making and should not entirely replace human oversight.

- Human oversight must be meaningful, or it will simply reinforce over-reliance on automated decision making. However, the oversight should not be so pronounced that it undermines the system's effectiveness or efficiency.
- Caution should be exercised when introducing AI, balancing the use of technology against maintaining the capability and skills of the operators, for example, language models may help to speedily summarise complex information. However, over reliance on such tools could lead to individuals losing their ability to analyse data and impact their subject matter expertise.
- Careful consideration must be given to the impact of AI on the delegation of decision making in both a public sector and criminal procedure context.
- Consideration should be given to ways in which agencies can develop formal policies regarding the balance between automated and human decision-making. Demonstrating accountability at an organisational level regarding decisions that affect the public directly is key to maintaining public confidence in the work of the public sector.

The Data Protection Act 2018 is a regulation in law on data protection and privacy for all UK individual citizens which sets out some key principles for the level of human input that is appropriate.

Transparency and ethics

The ability to explain the operation of an AI tool should be a key consideration in its selection and/or development.

- The legal right of the public to understand and potentially challenge government decisions through requests for information is important and must be preserved.
- Agencies should be prepared to explain their decision-making processes and how the AI works at a level that satisfies criminal procedure requirements.
- Where a technical explanation for an AI tool is not possible, practical or meaningful, an ability to explain the priorities or strategic basis for a decision may suffice and may even be more meaningful depending upon the context.

The aim for those in public sector agencies is use of AI that best ensures:

- the highest standards of legality, ethics, transparency and accountability are met;
- individuals remain accountable for decisions even where AI makes that decision;
- evidential/admissibility and data quality requirements are fulfilled;
- public trust and confidence in the use of AI by the public sector is maintained;
- the data collected, transmitted, processed, and stored is secure; and
- personal privacy and civil liberties are maintained.

> AI ethics is a system of moral principles and techniques intended to inform the development and responsible use of artificial intelligence technology

National Cyber Security Centre: Large Language Models - what's the risk?[9]

The National Cyber Security Centre (NCSC) is an organisation of the United Kingdom Government that provides advice and support for the public and private sector in how to avoid computer security threats. They recently published a blog on the potential risks of large language models, explored further below.

Large language models (LLM) are the fastest growing consumer applications. As with emerging technology, there are always going to be concerns about the security risks they may bring.

An LLM scrapes the internet for information, which can include books, research and social media posts. This therefore may include offensive or controversial data with it.

The algorithms analyse the relationships between different words and turn that into a probability model.

It is then possible to give the algorithm a 'prompt' (for example, by asking it a question), and it will provide an answer based on the relationships of the words in its model.

The NCSC blog goes on to explain that LLMs while helpful are not 'magic' or a silver bullet to solve all issues we are trying to navigate. Issues they identify include:

1. They can get things wrong and hallucinate facts.
2. They can be biased, and gullible (e.g. responding to leading questions).
3. They require huge compute resources and vast data to train from scratch.
4. They can be coaxed into creating toxic content and can be prone to injection attacks (untrusted input or unauthorized code injected).

Common concerns

A concern may be that the LLM learns from the queries input, which may or may not include sensitive data. The NCSC explain, 'Currently, LLMs are trained, and then the resulting model is queried. An LLM does not (as of writing) automatically add information from queries to its model for others to query. That is, including information in a query will not result in that data being incorporated into the LLM'.

However, its important to note the host or organisation holding the data of the LLM will retain that query/information- this means its very important to understand the privacy notice when engaging LLMs.

The NCSC recommends:

- Not to include sensitive information in queries to public LLMs.
- Not to submit queries to public LLMs that would lead to issues were they made public.

9 NCSC blog - Chat GPT and LLMs- what's the risk- accessed 2024

LLMs and Cyber Criminals

The NCSC highlight how criminals have used LLMs to develop malware, using the power of the technology to develop these tools.

The NCSC commented that for more complex tasks, it's currently easier for an expert to create the malware from scratch, rather than having to spend time correcting what the LLM has produced.

However, an expert capable of creating highly capable malware is likely to be able to coax an LLM into writing capable malware.

LLMs can also be queried to advise on technical problems. There is a risk that criminals might use LLMs to help with cyber attacks beyond their current capabilities, in particular once an attacker has accessed a network.

For example, if an attacker is struggling to escalate privileges or find data, they might ask an LLM, and receive an answer that's not unlike a search engine result, but with more context.

LLMs are able to replicate writing styles on demand, so there is a risk of criminals using LLMs to write convincing phishing emails, including emails in multiple languages. This may aid attackers with high technical capabilities but who lack linguistic skills, by helping them to create convincing phishing emails, in the native language of their targets.

To summarise, we might see:

- More convincing phishing emails as a result of LLMs.
- Attackers trying techniques they didn't have familiarity with previously.
- A low risk, of a lesser-skilled attacker writing highly capable malware.

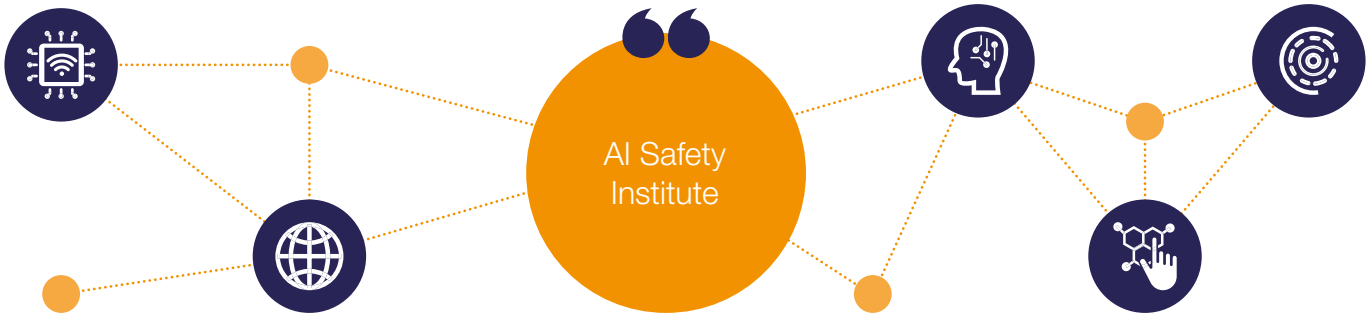
UK Government launches AI Safety Institute

In launching the AI Safety Institute, the UK is continuing to cement its position as a world leader in AI safety, working to develop the most advanced AI protections of any country in the world and giving the British people peace of mind that the countless benefits of AI can be safely captured for future generations to come.

The Frontier AI Taskforce will now evolve to become the AI Safety Institute, with Ian Hogarth continuing as its Chair. The External Advisory Board for the Taskforce, made up of industry heavyweights from national security to computer science, will now advise the new global hub.

The Institute will carefully test new types of frontier AI before and after they are released to address the potentially harmful capabilities of AI models, including exploring all the risks, from social harms like bias and misinformation, to the most unlikely but extreme risk, such as humanity losing control of AI completely. In undertaking this research, the AI Safety Institute will look to work closely with the Alan Turing Institute, as the national institute for data science and AI.

Already, the UK has agreed two partnerships: with the US AI Safety Institute, and with the Government of Singapore to collaborate on AI safety testing – two of the world's biggest AI powers.



Further reading

UK government	https://www.gov.uk/government/collections/a-guide-to-using-artificial-intelligence-in-the-public-sector
NCSC	https://www.ncsc.gov.uk/blog-post/chatgpt-and-large-language-models-whats-the-risk
IPSFF	https://assets.publishing.service.gov.uk/media/5e4545fe40f0b677be5fbd62/Artificial_intelligence_13_Feb.pdf
OECD principles and recommendations	https://www.oecd.org/going-digital/ai/principles/
Canada	https://www.tbs-sct.gc.ca/pol/doc-eng.aspx?id=32592
USA	https://www.gao.gov/products/GAO-18-142SP
New Zealand	https://www.lawfoundation.org.nz/wp-content/uploads/2019/05/2016_ILP_10_AILNZ-Report-released-27.5.2019.pdf
Australia	https://www.industry.gov.au/publications/australias-artificial-intelligence-ethics-framework/australias-ai-ethics-principles