DAWES CENTRE FOR FUTURE CRIME AT UCL

AI-enabled future crime

Artificial Intelligence (AI) technologies have applications for crime prevention and detection, but they could be exploited for criminal purposes in many different ways. This briefing identifies 20 different potential AI-enabled future crimes.

Summary

This study identified 20 applications of AI and related technologies which could be used for crime now or in the future. Future crimes were ranked as either low, medium or high concern in relation to the harm they could cause, the criminal profit (achieving a financial return, terror, harm or reputational goal), the achievability of the crime and its difficulty to defeat. Six crimes were identified as most concerning: audio and video impersonation, driverless vehicles as weapons, tailored phishing, disrupting AI-controlled systems, large-scale blackmail and AI-authored fake news.

Introduction

Applications of AI are a growing feature of, and are improving modern life: from 'personal assistants' (such as Amazon Alexa and Google Home) and satellite navigation, to 'behind the scenes' applications in language translation, biometric identification (such as fingerprint and face recognition) and industrial process management. Emerging AI applications include systems for crime prevention and detection, but the technology also has the potential to be misused.

This briefing sets out some of the possible ways in which AI technology could be exploited for criminal purposes. We also provide an assessment of the level of concern associated with each crime. Crime prevention and detection strategies must keep pace with an ever evolving technological landscape. An understanding of how new technologies could be exploited for crime is essential for policy actors, law enforcement agencies and technology developers alike.

Using AI for criminal purposes

AI can be exploited for criminal purposes in multiple ways, which are not mutually exclusive:

- As a **tool for crime**, where AI is used to undertake a traditional crime, such as theft, intimidation or terror.
- As a **target for criminal activity**, where AI systems are targeted by criminals – such as attempts to bypass protective AI systems or to make systems fail or behave erratically.
- As a **context for crime**, where fraudulent activities might depend on the victim believing that some AI functionality (such as predicting stock markets or manipulating voters) is possible even if it is not.

Unlike traditional crimes, crimes involving AI are highly scalable; once developed, techniques can be shared, repeated, or even sold. This raises the opportunity for **marketisation** of criminal techniques and 'Crime as a Service' (CaaS). We have already seen this type of (crime) business model for Denial of Service (DoS) attacks – where attacks can be hired to take websites or other online services offline for as little as $1 – but we may see CaaS emerge for a wider set of offences.

Future crimes involving AI

This scoping study identified 20 categories of AI-enabled future crime. These were then ranked by 31 experts (including representatives from academia, the police, the defence sector and government) at a two-day 'sandpit' event as being of either high, medium or low concern. In this briefing, crimes are not listed in a specific order within the high, medium and low concern categories.

The crimes were ranked according to four different dimensions:

- **Harm:** to individual victims or to society, including terror.
- **Criminal profit:** Realization of a criminal aim, e.g. financial return, terror, harm or reputational damage.
- **Achievability:** The readiness of the technology, its availability and the practicalities of achieving the crime.
- **Difficulty of defeat:** The difficulty of preventing, detecting or rendering the crime unprofitable. Consideration was given to whether measures would be obvious, simple or complex, and whether or not it required behaviour change.

High concern crimes **Crime dimension ranking key:** ● Low ●● Medium ●●● High	Harm	Criminal Profit	Achieveability	Difficulty of defeat
Audio/visual impersonation Impersonation of another person on video or audio. This could be impersonation of children to relatives over video calls to gain access to funds (there are examples of this in Mexico but with actors playing the role of relatives), phone conversations to request access to secure systems, or fake video calls of public figures speaking or acting in a different way to attempt to influence public opinion. Recent developments in deep learning (see glossary) have increased the scope for the generation of fake content, meaning achievability is high. Difficulty of defeat was considered high; although some success has been demonstrated in the use of algorithms to detect impersonation, there are many uncontrolled routes for fake material to spread.	●●●	●●	●●●	●●●
Driverless vehicles as weapons Motor vehicles have long been used both as a delivery mechanism for explosives and as kinetic weapons of terror in their own right. Fully autonomous AI-controlled driverless vehicles are not yet on the road, but numerous car manufacturers and technology companies are racing to deliver them. AI could expand vehicular terrorism by reducing the need for driver recruitment, enabling single perpetrators to perform multiple attacks, or even coordinating large numbers of vehicles at once. Difficulty of defeat was ranked as relatively low by delegates because driverless vehicles are expected to be susceptible to the same countermeasures (barriers and traffic restrictions) already in use for vehicles with drivers.	●●●	●●	●●●	●
Tailored phishing Phishing is a social engineering attack that aims to collect secure information or install malware via a digital message purporting to be from a trusted party, such as a bank. The attacker exploits the existing trust to persuade the user to perform actions they might otherwise be wary of, like revealing a password or clicking a link. Conventional phishing is already rife, but AI has potential to improve the success rates of phishing attacks by crafting messages that appear more genuine, and to discover 'what works' – by varying details of messages to "experiment" at scale and at almost no cost. This was rated as difficult to defeat; as phishing messages would be indistinguishable from genuine ones (aside from the link being fake.) The technology used to write the messages is improving, making messages look more like they have been written by a human.	●●	●●●	●●●	●●●
Disrupting AI-controlled systems As AI systems become ever more essential (in government, commerce and the home), the opportunities for attack will multiply, leading to many possible criminal and terror scenarios arising from targeted disruption of such systems, from causing widespread power failures to traffic gridlock and breakdown of food logistics. Systems with responsibility for public safety and security and systems overseeing financial transactions are likely to become key targets. Achievability was ranked as low because attacks typically require detailed knowledge of, or even access to, the systems involved, which may be difficult to obtain.	●●●	●●●	●	●●●
Large scale blackmail Traditional blackmail involves pressure under the threat of exposure of evidence of criminality or wrongdoing, or embarrassing personal information. Acquiring this evidence is a limiting factor: the crime is only worthwhile if the victim will pay more to suppress it than it costs to acquire. AI can be used to do this on a much larger scale, harvesting information from social media or large personal datasets such as email logs, browser history, hard drive or phone contents, then identifying specific vulnerabilities for a large number of potential targets and tailoring threat messages to each. Achievability was ranked as low because the crime requires large amounts of data and a combination of different AI techniques would be needed. However, difficulty of defeat was also rated as high because victims may be reluctant to come forward and face exposure.	●●	●●●	●	●●●
AI-authored fake news Fake news is propaganda that aims to appear credible by being, or seeming to be, issued from a trusted source. As well as delivering false information, fake news in sufficient quantity can displace attention from true information. AI could be used to generate many versions of a particular piece of content, apparently from multiple sources, to boost its visibility and credibility; and to choose content or its presentation, on a personalised basis, to boost impact. Criminal profit was ranked as low because it was perceived to be difficult to directly make financial profit from fake news, although there is potential to use fake news in market manipulation. It is highly achievable as the technology already exists and is hard to defeat both technically and because the boundary between fake and real news is vague.	●●●	●	●●●	●●●

Medium concern crimes

Misuse of military robots

The availability of any military hardware to criminal or terrorist organisations can be expected to pose a serious threat, including autonomous robots intended for battlefield or defensive deployment. The threat level is not known: military capabilities are shrouded in secrecy, and we have limited knowledge as to the current state of the art and rate of advancement.

Snake oil

Sales of fraudulent services, such as security screening, lie detection or targeted advertising, under the guise of AI or using a smokescreen of machine learning (ML) jargon. The products are fake but lent credibility by popular (mis) conceptions about AI. This type of fraud is highly achievable, with almost no technical barrier (since by definition the technology doesn't work). It should in theory be easy to defeat via education and due diligence by organisations who are purchasing products, who can often buy products with limited understanding of what AI is and isn't.

Data poisoning

The manipulation of machine learning training data to deliberately introduce specific biases, either as an end in itself or with the intention of subsequent exploitation. For example, making an automated X ray threat detector insensitive to weapons being smuggled on to a plane. Trusted data sources tend to be hard to change and are under frequent scrutiny, so this type of crime is likely to be difficult to achieve.

Learning-based cyber-attacks

Existing cyber-attacks tend to either be sophisticated and tailored to a particular target, or crude but heavily automated, relying on the sheer weight of numbers. AI raises the possibility of attacks that are both specific and massive, using, for example, approaches from reinforcement learning to probe the weaknesses of many systems in parallel before launching multiple attacks simultaneously.

Autonomous attack drones

Drones can be used for crimes such as smuggling drugs into prisons and have also been responsible for major transport disruptions. Autonomous drones under onboard AI control could enable greater coordination and complexity of attacks while the perpetrator is not required to be close by to the drone. Drones could cause harm, but in some contexts could be easily defeated; protection may be provided using physical barriers.

Online eviction

The prevalence of online activities within modern life, for finance, employment, social activity and accessing public services presents a novel target for attacks against the person: denial of access to what have become essential services is potentially debilitating.

This could be used as an extortion threat or to cause chaos.

Tricking face recognition

AI systems performing face recognition are increasingly used for proof of identity on devices like smartphones, and are also in testing by police e.g. on suspect tracking in public spaces and to speed up passenger checks at borders. These systems could be an attractive target for criminals.

Some successful attacks have already been demonstrated, for example "morphing" – the use of photo ID containing a graphical morph between two faces that can serve as ID for both.

Market bombing

The manipulation of financial or stock markets via targeted, probably high frequency, patterns of trades, in order to damage competitors, currencies or the economic system as a whole. The idea is an AI boosted version of the fictional Kholstomer cold war plot, which envisaged a Russian attempt to precipitate a financial crash by suddenly selling huge stockpiles of US currency via front companies. Achievability was rated low because of the extreme difficulty of accurately simulating market behavior (required to train the market manipulating AI) and the very high cost of entry to engage in large scale trading.

Low concern crimes

- **Bias exploitation:** taking advantage of existing biases in algorithms.
- **Burglar bots:** small autonomous robots used to commit burglaries.
- **Evading AI detection:** undermining AI systems that are used by police/security services.
- **AI-authored fake reviews:** automatic content generation to skew review scores.
- **AI-assisted stalking:** monitoring the location and activity of individuals.
- **Forgery:** generation of fake content such as art or music.

Further detail about the low ranked crimes can be found in the full report online.

Methods

- **A Review** of academic, news, current affairs, fiction and popular culture sources was conducted to identify and catalogue possible AI applicationsand related technologies for perpetration of crime.
- **A sandpit event** was held with 31 representatives from academia, the police, the defence sector, Government and the private sector to rank the threats identified during the review along four dimensions (harm, criminal profit, achievability, difficulty of defeat).
- **A ratings analysis exercise** took place after the sandpit to aggregate similar crimes and rank them. This resulted in 20 crimes with a rating value in these four dimensions.

Glossary

Artificial Intelligence (AI) describes efforts to enable computers to reproduce tasks that normally require human intelligence, including language use, vision and autonomous action. At present, most uses of AI are task specific rather than being capable of doing many different things.

Machine Learning (ML) is a subset of AI, where methods are based on discovering patterns in data. Because these algorithms "learn" how to perform tasks rather than being told how to do them, it can be difficult to understand how they work. ML 'training data' is the actual dataset used to train the algorithms.

Deep Learning (DL) is a class of ML methods using multi-layered 'artificial neural networks' to progressively extract complex features like faces or spoken words from raw data such as images or audio recordings.

Reinforcement learning (RL) is an exploratory ML approach in which independent software agents can observe and interact with some system (such as a game) and repeatedly try out different actions, with the goal of maximising a 'reward' (for example the score in the game). RL is especially relevant for dynamic problems such as how a robot should interact with its environment.

About the authors

Dr Matthew Caldwell
m.caldwell@ucl.ac.uk

Dr Jerone T. A. Andrews
jerone.andrews@cs.ucl.ac.uk

Dr Thomas Tanay

Professor Lewis Griffin
L.Griffin@cs.ucl.ac.uk

Contact: Mr Vaseem Khan, Business Director, UCL Security & Crime Science **vaseem.khan@ucl.ac.uk** or Professor Shane Johnson, Director, Dawes Centre **shane.johnson@ucl.ac.uk**.

Funders

This research was funded by the Dawes Centre for Future Crime at UCL. The Centre was established to identify how technological, social or environmental change might create new opportunities for crime and to conduct research to address them.

The Dawes Centre is funded by the Dawes Trust and UCL. These funds are limited and so we invite additional funding from the public and private sector. By funding the centre you will contribute to helping society better prepare for crimes of the future. We are also able to undertake research upon request, contributing to organisational goals and strategic thinking.

Find out more

A new PhD research project on adversarial perturbations (looking at how small changes to the data used to train a ML system can cause it to produce the wrong output) is underway following this scoping study.

The full research paper can be accessed online at this link: **http://doi.org/10.1186/s40163-020-00123-8**

This briefing was developed with the Policy Impact Unit. Find out more at **www.ucl.ac.uk/steapp/PIU** or email us on: **PolicyImpactUnit@ucl.ac.uk**