A Review of *'Artificial intelligence and crime: A primer for criminologists'*

Maximilian Matthews

Keith J. Hayward and Matthijs M. Maas' *Artificial Intelligence and Crime: A Primer for Criminologists* is a primer on *artificial intelligence* (AI) for the criminologist community. The paper begins by demystifying the term AI, introducing terms such as *Machine Learning* and *Deep Learning* and providing a high level overview of how it works. The authors continue by exploring the potential applications of AI by criminals, splitting this into three categories: crimes with AI, crimes against AI, and crimes by AI. Finally, the authors discuss how law enforcement is using the technology and how it is changing their underlying enforcement strategy. In this review, I aim to summarise the key points raised by the authors, as well as offer occasional personal remarks on the points and conclusions they draw.

The article begins by discussing the phenomenon of *techno-optimism*, whereby computer scientists have historically downplayed the potential negative consequences of digital innovation, both in terms of use by criminals as well as overly intrusive policing practices. The authors also hint at the tendency for criminologists to ignore advances in AI and technology, pitching this paper as a wake up call for the profession. In particular, the authors discuss how, despite the popularity of AI as a media talking point, it is strangely opaque, with most people having little understanding of what AI truly is, its capabilities, and its limits. They go on to discuss how this causes a *chasm of computational thinking*, whereby people cede decision-making power to automated AI, leading to an opaque concentration of power. They postulate that people outside the computer-science field, in this case criminologists, must engage in conversations around AI to shape and direct its development.

I agree with this line of thought, far too often, consumers and business people place blind trust in AI (and other high tech systems) without thinking critically about their implementation. When they are mentioned, people either catastrophize about a robot apocalypse or ignore them. I support the sentiment that people in all fields must engage with the reality of AI to ensure it is ethical.

Next, the authors offer an explanation of what AI is, how it works, and what its capabilities and limitations are. First, the "intelligence" in artificial intelligence is explored. They write about how, rather than mimicking human consciousness, like many people think, AI aims to replicate (or exceed the human performance of) specific tasks and that *its humanness is besides the point*. Modern forms of AI, brought about by advances in algorithm science, big data, and the processing power of chips, are explored. Specifically, the learning process of deep learning neural networks. Next, the different approaches to learning are examined: supervised learning, unsupervised learning, and reinforcement learning. General adversarial networks are also touched on.

The article goes on to discuss some of the capabilities and limitations of AI. Firstly, it's use in classification tasks, anomaly detection tasks, prediction tasks, systems optimisation tasks, and autonomous robot operation tasks. The authors comment that, despite the AI performing very well at these tasks, they're all quite limited in scope. This paper was written in 2020, since then there have been numerous innovations in AI which have expanded its use case. Mostly notably *ChatGPT*, a natural language chat-bot introduced in November 2022, has a vast scope of use cases producing natural content and answering questions across many domains. The amount of development in only two years demonstrates the fast pace of AI product development, and how information can be out-of-date very quickly.

Numerous limitations are also mentioned, these include physical issues such as access to data, human talent, and adequate hardware. As well as limitations with AI technology, these include the tendency *of artificial stupidity* and *catastrophic forgetting*, the limitation core to AI models whereby models trained on one situation cannot easily adapt and transfer the learning to another given situation. Additionally the level of unpredictability of AI when encountering a new situation is mentioned. Together these present core limitations to AI, and are why AI use is often avoided in control systems with high safety requirements. It is important that the author raised this limitation as it is often an oversight, especially when many people view AI as a natural solution to everything.

When discussing use of AI by criminals and other malicious actors, the authors of the article split this into three categories: crimes *with* AI, crimes *on* AI, crimes *by* AI.

When discussing crimes *with* AI, the authors first explore how it can expand the threat of crime in a physical context, some examples raised are the use of unmanned vehicles for smuggling, as well as autonomous drones with explosives posing a national security threat. However they conclude that compared to other use cases these pose a limited threat, I agree both due to the cost of entry as well as specialised technical knowledge required for this to function.

One of the biggest criminal threats raised by AI, in the article, is the possibility of automating cybercrime, especially phishing attacks. The authors raise how 91% of cybercrimes start with a user clicking on a phishing email. It is stressed that while currently they're fairly suspicious (e.g. *You've won $1 Million!*), natural language AI software, trained on previous phishing attempts, including success rates, can now be used to automatically generate highly personalised phishing messages. These can be up to four times more effective than regular messages.

Furthermore, the use of GANs to forge various types of media is a risk. The authors discuss how deepfakes have gained much attention in the press, particularly in the realm of political manipulation. They go on to state, however, that the primary threat with deepfakes is the production of plausible, non-consensual, intimate content used for extortion, blackmail, and harassment. They state that 96% of deepfakes online are pornographic material.

To me reading this was very shocking, I am very aware of the role deepfakes play in *Fake News* and electoral manipulation, but I was not aware of the extent to which they're used for

extortion. Perhaps the authors of the article should have gone into more detail on the gap between public knowledge of AI related issues, and the reality of how they are used.

When discussing crimes against AI, the paper explores attacks that exploit and reverse-engineer system vulnerabilities to fool AI systems. It discusses how neural networks rely disproportionally on counterintuitive details and patterns, which hackers can exploit by feeding them specific data, causing the AI to misclassify the subject or perform an unexpected action. The authors explore how this is used for both good and bad. On the one hand, adversarial images (e.g. printed on a t-shirt) have been used to render people invisible to AI-powered surveillance cameras in authoritarian countries, while on the other hand, criminals have been able to embed hidden voice commands in innocent-sounding audio, which can force smart speakers to dial certain numbers or open certain websites.

I find this to be particularly alarming, especially since AI systems have been embedded in so many everyday products. While it is important that this topic be covered in the paper, it could have been used to include information on how security professionals and individuals can safeguard against adversarial attacks, particularly with a focus on adversarial attacks that execute commands. This is especially important as Internet of Things (IoT) devices become commonplace.

Finally, the article discusses crimes committed by AI as an intermediary. The authors explore how, once a person creates an AI code, they often have little control over what that AI does in order to achieve the intended outcome. Thus, if an AI commits a crime (e.g., an AI financial trading agent committing actions that amount to the crime of market manipulation while deployed), there is uncertainty over whether the person who created the AI is to blame or whether it is its own legal actor.

This is an interesting topic, since there is no concrete precedent for crimes committed by AI. It would have been interesting for the article to have also explored to what extent (human) criminals are able to hide behind the defence of their AI having done a crime, and also to what extent people have control over the actions of their AI.

The third part of the paper deals with use of AI technology by the Police.

The first key police use of AI is for expanded digital photographic surveillance of citizens. AI tools can be layered on top existing 'dumb' CCTV infrastructure. This facial recognition makes it very easy for the police to track people around cities. Furthermore, AI tools that can recognise a person from their voice are also becoming more established, a further surveillance tool to police can use.

The authors note a number of ethical concerns with this widespread AI surveillance. They note the ACLU has suggested these technologies are untested, discrimination, and subject to abuse. Furthermore they also note examples where the technology has been used by governments for the purpose of persecuting minority groups, for example in China where such systems are configured to *flag members of the ethnic Uighur minority* by their facial features. I don't believe the article goes far enough in exploring the negative consequences of

facial recognition technology, particular in terms of inherent racial bias and privacy concerns.

The second key point surrounding police use of AI is the transformation of police strategy. According to the authors developments in AI and data collection now allow vast areas (> 20 km^2) to be monitored from a single camera. Furthermore, AI policing systems (pooling data from various sources) are able to pick up un subtle patterns to offer predictions of future criminal behaviour, these allow law enforcement to make subtle interventions shaping urban architecture. The use of (automated) *hypernudge* techniques allows police to make subtle changes to prevent crimes before they happen. This marks a shift from the traditional *detection and enforcement* strategy, to a strategy of *prediction and prevention*.

The authors argue that these changes in fact make policing less intrusive since instead of traditional tactics such as stop and search and obvious CCTV, policing will become more subtle and also more precise, meaning fewer people will need to deal with difficult police encounters.

However I believe that here the authors fail to take into account the erosions into personal freedom and privacy that come with the mass data surveillance needed for a system such as thing to function. Moreover, a move to a *prediction and prevention* model has difficult implications for the free will of citizens, if peoples' behaviour is consistently being manipulated by police (or by autonomous AI systems). A more holistic view of the pros and cons of each needs to be achieve before an adequate judgment can be made.

In conclusion, the article shows how AI is already having a substantial impact in the law enforcement sector, both through changes in how crime is committed as well as leading to changes in the methods of law enforcement. Overall, the authors presented a highly informative, balanced picture of what is happening. However they could have gone into more detail on the negative consequences of AI-based policing.