Crime Media Culture
1–25
© The Author(s) 2020
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/1741659020917434
journals.sagepub.com/home/cmc

Artificial intelligence and crime: A primer for criminologists

Keith J Hayward and Matthijs M Maas
University of Copenhagen, Denmark

Abstract

This article introduces the concept of Artificial Intelligence (AI) to a criminological audience. After a general review of the phenomenon (including brief explanations of important cognate fields such as 'machine learning', 'deep learning', and 'reinforcement learning'), the paper then turns to the potential application of AI by criminals, including what we term here 'crimes *with* AI', 'crimes *against* AI', and 'crimes *by* AI'. In these sections, our aim is to highlight AI's potential as a criminogenic phenomenon, both in terms of scaling up existing crimes and facilitating new digital transgressions. In the third part of the article, we turn our attention to the main ways the AI paradigm is transforming policing, surveillance, and criminal justice practices via diffuse monitoring modalities based on prediction and prevention. Throughout the paper, we deploy an array of programmatic examples which, collectively, we hope will serve as a useful AI primer for criminologists interested in the 'tech-crime nexus'.

Keywords

Artificial Intelligence, Big Data criminology, cybercrime, digital criminology, inescapable surveillance, machine learning, technology and crime

Introduction

Standing behind a desk on the type of gargantuan conference stage favoured by chino-wearing Microsoft executives and headset-sporting tech futurologists, computer scientist Zeyu Jin addresses delegates at a 2016 Adobe MAX event in San Diego. Jin is a designer behind 'Project Voco', a prototype audio editing-and-generating software that enables users to manufacture digital sound. Dubbed 'Photoshop-for-voice', Voco is at the vanguard of a suite of technologies that essentially allows anyone with a laptop to forge or manipulate voice-recordings. A hushed atmosphere descends as the software is cued up on a billboard-size video screen. The silence doesn't last long.

Corresponding author:
Keith J Hayward, Faculty of Law, University of Copenhagen, Njalsgade, 76, 2300 Copenhagen S, Denmark.
Email: keith.hayward@jur.ku.dk

After Jin demonstrates the ease by which he can fabricate the voice of the American comedian Keegan Michael-Key, the audience of bloggers and start-up entrepreneurs erupts with all the fervour of a Southern Baptist revival. Jin, thrilled with the response, holds his arms aloft in recognition of the applause. However, even before the cheering subsides, a cautionary note is sounded by Key's comedy partner, Jordan Peele, who jokes: 'You could get in big trouble for something like this'. Jin is unfazed. 'Don't worry', he later responds 'We have actually researched how to prevent a forgery . . . it's really not for bad stuff'.[1]

It's really not for bad stuff. Jin's sanguine position is unsurprising. For decades, computer scientists have been so captivated by the unlimited potential of new technologies that the negative effects of these systems have been downplayed or often ignored entirely.[2] Known as techno-optimism, this failure to effectively balance reward and risk was famously reflected in 'Don't be evil', the former motto of Google's corporate code of conduct. However, just as Google quietly dropped the phrase in the wake of concerns about the company's involvement in censorship, tax avoidance, and privacy-violation scandals, tech developers are now realizing that Artificial Intelligence (AI) systems will not only enable unpalatable policing practices, but will also provide a host of new avenues for serious criminal exploitation.[3]

Yet if many computer scientists are understandably guilty of concentrating their efforts on the technology side of what we might call here the 'tech-crime nexus', then the same is true in reverse of criminologists. While our discipline continues to advance knowledge about crime and punishment, it does so largely oblivious of the many social challenges posed by technological disruption (Brown, 2006; Hayward, 2012; Holt and Bossler, 2014).[4] Not only do most criminologists ignore matters relating to technology itself, but with few exceptions they have shown a studied disregard of theories from other disciplines that have sought to open a space for dialogue between the social sciences and information and communications technology. As a result, while recent years have seen scholars begin to survey and warn of the potential 'malicious uses' of AI (Brundage et al., 2018), and its offensive and defensive applications (Broadhurst et al., 2018), mainstream criminologists have yet to play any significant role in what King et al. (2019) have usefully described as the emerging interdisciplinary field of 'AI-Crime' (AIC).[5]

Thankfully, this situation is changing. Recently, criminology has been enlivened by a series of new research clusters concerned with how crime will be transformed by the impact of what Greenfield (2017) calls the 'radical new technologies of the networked era'. Here, we refer to new subfields such as digital criminology (Powell et al., 2018); computational (Williams and Burnap, 2016) and Big Data-era criminology (Chan and Bennett Moses, 2017; Smith et al., 2017); and a growing body of research into 'technocrimes' (Steinmetz and Nobles, 2017) involving encryption, cryptocurrencies, illicit trade and 'dropgangs' on the dark web, and 'stalkerware' (e.g. Aldridge, 2019; Kruithof et al., 2016; Munksgaard et al., 2016; Paoli et al., 2017; Parsons et al., 2019). The recent publication of McGuire and Holt's (2017) impressive and much-needed *Routledge Handbook of Technology, Crime and Justice* further evidences criminology's burgeoning interest in matters technological.[6] It is our hope, this article will lend further weight to this body of work by training attention on criminological issues related to one particular facet of contemporary technology: AI.

To lay observers, AI can be a difficult concept to grasp – a phenomenon that is seemingly everywhere, yet at the same time strangely opaque. In popular culture and news reporting, AI is often the stuff of fanciful narratives about 'killer robots' or dystopian surveillance systems. However, in terms of people's everyday lives, AI tends to function at a much more prosaic level, driving everything from

Smart TVs to language translation applications. It is perhaps this universality that confuses people, not least because each putative AI future evokes its own particular array of concerns about safety, ethics, legality and liability. If society is to overcome this confusion, what is required are clear answers to straightforward questions such as 'What exactly is AI?' 'What are its capabilities and limits?' And, most importantly for criminologists, 'What are the consequences of its proliferation and use in society, both as a tool for criminal or illegitimate ends, and as a means of security and social control?'

In a bid to answer such questions the paper will proceed in three parts. We start with a short, accessible, introduction to AI. As we are not writing for computer scientists, we will avoid diving into technical details and computational matters, and instead offer a more general overview aimed specifically at criminologists and other interested social scientists. In part 2, we turn to the potential application of AI by criminals, including 'crimes *with* AI', 'crimes *against* AI', and 'crimes *by* AI'. Part 3 addresses the use of AI applications by police forces, criminal justice agencies and governments and how the scaling up of granular, ubiquitous and predictive surveillance will redefine the culture and texture of future urban landscapes. Importantly, these three areas do not exhaust the space for future criminological enquiry in relation to AI. For instance, one additional salient area concerns potential methodological innovations in the use of AI systems to assist in the study of criminological phenomena. This will no doubt offer tremendous opportunity to criminologists, but will also require their careful critical scrutiny. However, because the goal of this paper is to lay the groundwork for criminology's future direct engagement with AI, our concern here does not extend to future digital methods.

Throughout, our critical position is one best described as dichotomous. On the one hand, our rationale for writing this paper is to reassure readers about AI, to demystify the concept, and to suggest that, as criminologists, we have much to offer this field. On the other hand, this paper also makes clear that, if deployed maliciously or without due diligence, AI applications could cause unfathomable damage. Thus, while it is not our intention to scare you about AI, we have in part written this to train attention on some of AI's more worrying tendencies. To achieve these aims, the paper relies on a series of programmatic examples which we hope will serve as a useful AI primer for interested criminologists.

What is AI?

'By far the greatest danger of Artificial Intelligence', AI theorist Eliezer Yudkowsky (2008) has quipped, 'is that people conclude too early that they understand it' (p. 308). Indeed, according to the journalist Kelsey Piper (2018), 'The conversation about AI is full of confusion, misinformation, and people talking past each other – in large part because we use the word 'AI' to refer to so many things'. In an effort to overcome the confusion that stems from semantic proliferation, this section clarifies key AI terminology, and some of the main functions and limitations of today's AI systems.

Competing notions of artificial 'intelligence'

'Intelligence' within the AI paradigm is a loaded and deeply contested philosophical and scientific concept. Legg and Hutter (2007a), for example, identified over 70 definitions. One way of overcoming this confusion is to adhere to Stuart Russell and Peter Norvig's (2009: 5) influential dyadic framework that turns around two interrelated philosophical questions:

1. *What is it we seek to create in AI?* The manifestation of certain internal thought processes, or more simply specific external behaviours/outcomes?
2. *How do we measure an AI system's performance?* Is the goal simply to reproduce/mimic human beings, or to surpass them by achieving 'optimal' performance towards certain outcomes?

This distinction is worth exploring, not least because it illustrates how thinking about AI has developed over time.

The first question essentially drives most of the popular commentary around whether or not robots could ever possess 'the right type' of thought processes ('sentience'/'consciousness').[7] However, today most AI research has moved away from attempts to reproduce *internal* notions of intelligence, to focus instead on more *external* (measurable) criteria. Indeed, this was even the actual point of the oft-misunderstood 'Turing Test', which after all was envisaged as an 'Imitation Game' (Turing, 1950: 460), rather than as a direct test of metaphysical computer sentience.[8]

This leads to the second question about measuring AI performance. Whereas earlier research was driven almost exclusively by the 'human-centric' definition – and this understanding still drives some adjacent social scientific analysis of AI today[9] – much contemporary computer science takes the 'humanness' of AI as beside the point, or of symbolic value at best. After all, the value of a high-frequency stock-trading algorithm is not that it can conduct a conversation with colleagues at the watercooler, but rather that it is effective at trading derivatives. And even there, the point is not that the algorithm performs *exactly* the same as a human trader, but that it can exceed human performance at that task. Indeed, in many recent landmark cases, the super-human performance of AI systems even results in explicitly non-human strategies that would never pass the Turing test. Such was the case with AlphaGo's famous initially-inscrutable but game-winning 'move 37' in its match against Lee Sedol, or with AlphaZero's 'alien' chess strategies (Knight, 2017). In modern AI, all too often it is not about 'sentience', but about capability–less about precisely imitating human performance, and more about surpassing it in various domains.

Modern AI terminology

Historically, there have been a range of distinct approaches or 'tribes' in AI (Domigos, 2015), many of which draw cultural and intellectual inspiration from fields such as logic, biology, statistics, or psychology. One type of AI system that has been used for decades is 'symbolic AI', which approaches tasks by following a set of explicit, logical 'if-then rules'. For example, an aircraft autopilot will keep an aeroplane within pre-set safe ranges in terms of altitude/airspeed. These rules have been pre-programmed by human experts on the basis of their domain knowledge. Symbolic AI underpins these so-called 'expert systems', which are so widely used that we often don't think of them as AI at all (Scharre, 2019).

However, in the last decade, three developments – advances in 'Big Data', processing power, and algorithmic innovations – have led to the rise of '*machine learning*' (ML), which is a more dynamic, less 'brittle' AI approach. ML involves the system gradually teaching itself the 'correct' (or 'useful') rules it needs to perform tasks effectively. Importantly, it does so on the basis of training data, rather than (as with 'expert systems') having these rules explicitly programmed.

One specific type of ML, responsible for the current AI boom, is *deep learning* (DL). DL involves deep neural networks – an AI technique inspired by how neurons communicate with each other in biological brains. Artificial neural networks consist of layers of digital interconnected 'neurons', some of which receive an 'input' (e.g. information about a certain pixel in an image), others provide an 'output' (e.g. a 'classification' of the image). Each neuron monitors others in the layer before it, and only if enough of those neurons send it a signal, will it then signal specific neurons in the subsequent layer. After each overall wrong/right response to training data sample, the system alters the strength of the connections between the involved neurons. In this way, it 'learns' to encode the rules to do its task.[10] For instance, an image-recognition algorithm will create clusters of neurons dedicated to the detection of increasingly more abstract concepts – from 'pixel colour', to 'edges and corners', to 'shapes' (eyes, noses), to concepts ('human', 'dog').

Importantly, how a given AI is trained depends on the specific type of ML algorithm, and of the sort of data used by the developers. There are a number of distinct approaches in use (Scharre and Horowitz, 2018). First, in *supervised learning*, the algorithm uses training data that has already been correctly pre-labelled by humans (e.g. photos of skin lesions, digitally labelled by doctors as either cancerous or benign). Second, in *unsupervised learning*, algorithms can independently identify patterns/correlations in 'raw', unlabelled data. This is useful not just because it saves on the cost of compiling large datasets of labelled images, but because it enables AI to identify patterns that humans cannot spot. Third, in *reinforcement learning*, AI systems 'learn' from feedback from their (real or simulated) environment, discovering by trial-and-error what different (combinations or sequences of) actions allow it to 'win' or maximize a 'score' metric (see Knight, 2017 on AlphaZero). Finally, a somewhat distinct application can be found in *generative adversarial networks* (GANs), whereby a neural network trains itself to generate fake data (images/sounds/videos), iteratively improving the quality of its creations until they are so indistinguishable from real data that they can fool another (regular and pre-trained) recognition-algorithm.

Modern AI: uses, pre-conditions, and limitations

But beyond terminology and technical perspectives, what exactly is AI used for? What are the preconditions for its deployment, and what, if any, are AI's limitations and weaknesses? Here it's useful to disaggregate the distinct functions that AI can serve (Scharre and Horowitz, 2018). AI systems can be used in any tasks involving data classification and generation, anomaly detection (e.g. detecting fraudulent financial transactions or new malware), prediction (e.g. re-offence rates for criminals), optimization of complex systems and tasks, and autonomous operation of robots/cyber-physical platforms. What's important to note about these tasks is they are all somewhat narrow, as we do not yet currently possess so-called artificial 'general' intelligence (Legg and Hutter, 2007b) capable of outperforming humans in any task.[11] Nonetheless, while narrow, many of these functions are useful in diverse sectors and contexts, including *inter alia* healthcare, law enforcement, advertising, and traffic management.

This does not mean AI is without limits. There are currently a range of preconditions for the effective application of AI to a given problem. These include primarily access to large (and sometimes labelled) datasets, but also pragmatic issues relating to hardware, human talent, and investment availability. Nonetheless, both the computational and data barriers are falling; moreover, they do not constrain the dissemination of already-trained systems – which is the concern in many criminal contexts. At the same time, even when all pre-conditions are in place, today's AI systems still suffer from a cluster of problems sometimes known as 'artificial stupidity' (Domigos, 2015). To start with, AI is often prone to 'catastrophic forgetting', the inability to easily transfer learning from one context to another. Second, AI is intrinsically susceptible to 'adversarial input' – data (e.g. visual or sound patterns) designed to alter the way the system processes stimuli, making it 'hallucinate'. Third, AI systems do not have 'common sense', and thus contrary to some of the connotations around the word 'intelligent', they inevitably suffer from the old 'Garbage-in, Garbage-out' (GIGO) problem, which as we already stated leads to bias problems (Barocas and Selbst, 2016). One final problem stems from the *unpredictability* of autonomous AI, which can often react unexpectedly when encountering unforeseen new situations. It is not uncommon, for example, for an AI programme to technically solve a problem, but not in a manner that was intended (Lehman et al., 2018). For example, one algorithm tasked with learning to walk in a simulated environment found that, rather than 'evolving' rudimentary legs, it could move forward by growing very tall and then repeatedly falling over. As noted by Scharre (2019), 'In the wrong situation, AI systems go from supersmart to superdumb in an instant'.

So what does all this mean for criminology? Since at least the turn of the century, scientists, science-fiction writers, and even occasionally criminologists (McGuire, 2007; Zedner, 2007) have all predicted that the type of developments outlined above would lead inevitably not just to new types of crime, but to new types of policing, punishment, and legal and pre-emptive decision making. At this point, an important question emerges: what, functionally, should or should not be considered 'AI' by criminologists? For example, does the term 'AI' include only narrowly (humanoid) robots and modern neural networks, or does it also extend to include older 'symbolic' expert systems, automated logistic regression (simple machine-learning), or indeed all algorithmic or computational processes such as those that drive the apps in your smart phone? This paper does not seek to definitively determine the contours of what is or is not 'AI'. Instead, our goal is to conceive of the 'AI-crime nexus' as a broad and fast-developing landscape that involves a diverse range of criminal, policing, and security practices that are now using technologies from across this spectrum. In what follows, we draw on the latest research – and a selection of recent incidents – to offer our own schematic vision of AI's imminent impact on crime and criminal justice.[12] Or in other words: now that we are in the future, what does it look like?

Criminal uses of AI

In their thoughtful recent survey article, King et al. (2019: 9–18) identify a range of threats posed by 'AI Crime' (AIC), including drug trafficking, sexual crimes, theft, fraud, and forgery. While theirs is a useful departure point, we structure our typology differently – not in terms of the area of the law affected, but rather in how criminals might use AI. Here, we identify three categories: (1) crimes *with* AI, (2) crimes *on* AI, and (3) crimes *by* AI.[13]

Crimes *with* AI (AI as tool)

Most fundamentally, AI can serve as a potent *tool* for 'malicious' criminal use by expanding and changing the inherent nature of existing threats, or by introducing new threats altogether (Brundage et al., 2018).

The expansion of existing threats can happen in a physical context. For example, drug traffickers could turn to unmanned vehicles (especially unmanned underwater vehicles) to improve smuggling success rates and the resilience of smuggling networks (Sharkey et al., 2010). More dramatically, some have cautioned that the combination of cheap quadcopter drones with facial recognition software and small explosive charges could, before long, create a new vector for mass terrorist attacks on civilians (Topol, 2016). They warn of a new 'weapon of mass destruction', one made more disturbing by the fact that it can (ethnically or politically) 'discriminate' in its targeting.[14]

While these are significant concerns, it will be in its native cyberspace where AI poses the greatest criminal threat. One use is in expanding existing hacking and malware threats. Researchers have already developed GANS to generate new malware that can slip by virus filters (Kolosnjaji et al., 2018). Another use is in scaling-up social engineering cyberattacks. Today, 91% of cybercrimes/ attacks start with a phishing email (Bahnsen et al., 2018) – a message which invites someone to click on a link which then takes them to a website that allows criminals to obtain sensitive personal information for the purposes of identity theft or fraud. However, to date, phishing emails are often generic (e.g. 'You've won $1million!') and therefore either easily caught by spam filters or unconvincing to all but a relatively small sub-population of particularly vulnerable users. More personalized phishing attacks ('spear phishing') are up to four times more effective than untargeted ones (Jagatic et al., 2007), but are labour-intensive as they need to be hand-tailored to target specific demographics or individuals. However, with 'DeepPhish' AI (Bahnsen et al., 2018), systems can automatically learn from and combine features (synthetic URLs, etc.) from other phishing attacks, avoiding spam filters and improving success rates. At the same time, AI may also play a role in improving defense: the recently developed 'Panacea' AI system uses natural language processing to respond to incoming fraudulent emails, engaging attackers in conversation to gain information about their true identity, while also wasting their time (Dalton, et al., 2020).

In another experiment, two researchers used AI to automatically generate large numbers of social media messages, all tailored to specific targets' profiles and past behaviour, convincing these users to click on phishing links (Seymour and Tully, 2016). Likewise, 'identity-cloning' bots which mimic people on social media, have shown high success rates in embedding themselves in social networks, since many users habitually accept all friend requests (Bilge et al., 2009). Indeed, in 2019, the Associated Press reported that AI-generated faces had been used to create phantom LinkedIn accounts for the purposes of getting embedded in the Washington D.C. policy establishment (Satter, 2019).[15]

Beyond scaling up existing threats, how might AI be used to develop *new* threats currently beyond the scope of human actors? A vivid and now increasingly common example here are so-called 'DeepFakes' (Chesney and Citron, 2019) – GAN applications capable of forging any type of media, including photographs of faces (Vincent, 2018),[16] video footage, voices 'cloned' from one-minute speech samples (Gholipour, 2017), or coherent text for targeted 'neural fake news' (Zellers et al., 2019), as illustrated by the GPT-2 and Grover systems. These DeepFakes have already proved convincing. In March 2019, thieves used voice-mimicking software to copy the voice of a CEO, calling the director of a subsidiary British energy company. This resulted in the latter executive transferring $243,000 to a fraudulent account (Harwell, 2019). Moreover, much of the commentary around DeepFakes has been concerned with their potential misuse for political manipulation. This is not unwarranted. For instance, in Belgium in 2018, the Flemish Socialist Party used these techniques to create a fake video of Donald Trump supposedly calling on Belgium to exit the Paris Climate Agreement. Meant as a climate-change stunt, many of the party's supporters shared the video (Von der Burchard, 2018). Since then, DeepFakes have gradually cropped up in a range of electoral contexts. However, while DeepFakes will undoubtedly contribute to our 'post-truth' political discourse, it remains the case that, currently, the primary function of this technology lies in the ability of criminals to create synthetic, yet plausible, intimate material for harassment, blackmail or 'sextortion' (cyber blackmail) (Spera et al., 2016). Indeed, a 2019 report shows that 96% of the 14,600 online deep fake videos involved the forging of non-consensual pornographic material (Ajder et al., 2019).[17]

AI can also forge other types of imaging-data. A recent study showed how malicious actors could use it to tamper with hospital data, adding or removing evidence of medical conditions from volumetric (3D) medical scans. This fake information could then be used to sabotage political candidates, corrupt research, compromise health infrastructure, or even commit murder (Mirsky et al., 2019).

Finally, AI is already illustrating the frailty of existing (cyber)security protocols. In 2017, New York University (NYU) researchers used GANs to generate ‘DeepMasterPrints’ – synthetic fake ‘fingerprints’ that can serve as a master key for biometric identification systems (Bontrager et al., 2017; Hern, 2018).[18] In the same year, ‘PassGAN’ was trained on datasets of leaked passwords, learning to generate likely candidates for human passwords in order to generate high-quality password guesses. In tests, this system outperformed existing state-of-the-art tools like HashCat, matching 51%–73% more passwords (Hitaj et al., 2017).

Crimes *on* AI (AI as attack surface)

Crimes ‘on’ AI involve attacks that exploit and reverse-engineer system vulnerabilities in a bid to fool or *‘hypnotise’* AI systems. It has been possible for some time to ‘poison’ a system’s training data. Infamously, Microsoft Twitter chatbot, ‘Tay’, was turned racist inside a day after users fed it a slurry of right-wing phrases (Gershgorn, 2016). Such incidents are just the tip of the iceberg. ML systems, when classifying data, often rely disproportionately on counterintuitive details and patterns. Hackers can use this feature to reverse-engineer input data to spoof systems into displaying particular behaviour (Nguyen et al., 2015). Worse, this can be done in ways that are not apparent to human inspection (Goodfellow et al., 2014, 2017). Moreover, attacks can often be carried out even in ‘black-box settings’, where an attacker does not have access to the internal weighting of the network. Researchers have shown it is even possible to generate a custom-made ‘adversarial patch’ sticker which causes an AI to misclassify objects as ‘toasters’ (Brown et al., 2017). Elsewhere, researchers managed to 3D-print a model of a turtle, altered to be perceived by AI as a ‘rifle’ from nearly every angle (Athalye et al., 2018).

Gu et al. (2017) have demonstrated that at times these problems are exacerbated by vulnerabilities in the ML model supply-chain. Given that many users outsource the (computationally intensive) training procedure or use pre-trained models, adversaries can create a ‘BadNet’, a maliciously trained network that performs very well on the user’s regular scenario, but which contains ‘environmental backdoors’ – specific inputs that fool the system into incorrect or dangerous behaviour. For example, in several cases, researchers showed that placing stickers on traffic signs and street surfaces can cause self-driving cars to ignore speed restrictions and swerve headlong into oncoming traffic (Evtimov et al., 2017; Tencent Keen Security Lab, 2019). Such problems are likely to become ever more commonplace. In 2018, Google researchers demonstrated that image-recognition neural networks can be tricked into performing free computations for attackers, potentially turning smartphones into botnets by exposing them to doctored images (Elsayed et al., 2018).

Such ‘AI hacks’ have serious real-world implications across diverse sectors. In healthcare, researchers have demonstrated that adversarial attacks can co-opt diagnostic algorithms to facilitate medical insurance fraud (Finlayson et al., 2019). Others have shown how even text-processing AI can be vulnerable to manipulation, as was shown recently by the system ’TextFooler’, which could analyze texts and suggest strategic synonyms to be changed in order to dramatically alter the decisions of AI systems in areas from job applications to fake news detection (Jin 2020; Knight 2020). Another application exploited vulnerabilities in voice-recognition systems such as Alexa, Siri and Google Assistant. Replicating audio waveforms (some accurate to within 99.9% of the original), researchers sent hidden voice commands to these smart speakers, making them dial phone numbers or open websites.[19] In theory, this could be used to attack 'smart homes' – unlocking doors, wiring money, or triggering buy-orders for incriminating or embarrassing products (Smith, 2018).[20]

These same adversarial techniques are also being used in the context of activism and resistance against pervasive surveillance culture(s). In Belgium, researchers designed an adversarial image which, if printed and carried around, rendered a person invisible to AI computer-vision systems (Thys et al., 2019).[21] In a similar vein, artists and fashion designers have started to collaborate with tech researchers to create wearable items like the 'anti-AI T-shirt' (Xu et al., 2019) and cosmetic 'dazzle camouflage' (Eckert et al., 2013), both of which (it is claimed) might protect protestors from identification by face-detection cameras.[22]

In summary, adversarial input shows how diverse actors can sprinkle an environment with what Scharre and Horowitz (2018, 15) call 'cognitive landmines', raising critical security vulnerabilities in 'Internet-of-Things' systems and posing problems for advocates of the 'smart city paradigm'. While work is underway on detecting adversarial examples (Xiao et al., 2018), many of these detection methods can themselves be easily bypassed (Carlini and Wagner, 2017), suggesting that those seeking to assure the security of AI infrastructure have their hands full.

Crimes *by* AI (AI as intermediary)

In 2015, a group of artists released a random shopping bot on the dark web – with the unsurprising outcome that it eventually bought drugs, and was arrested by Swiss police (Kasperkevic, 2015). This incident not only provides a lucid example of our third AIC category, '*Crimes by AI*', but also, more importantly, it raises the thorny issue of AI's legal status – and its potential misuse as 'criminal shield/intermediary'.

Some lawyers have suggested that it may already be possible to grant certain algorithms some semblance of legal personhood. Bayern (2016) has argued that loopholes in existing US company laws allow for the functional incorporation of 'artificially intelligent entities' with legal personhood. The legal intricacies behind such arrangements, or the broader merits or societal value of such algorithmic personhood (cf. Turner, 2018), are beyond the scope of this paper. However, what is of interest here is how such legal chicanery enables new modalities of white-collar crime (LoPucki, 2017). Specifically, using AI as an 'independent' criminal intermediary poses serious questions for cornerstone legal standards such as the voluntarily undertaken criminal act (*actus reus*), criminal intent (*mens rea*), and various issues surrounding the knowledge threshold, foreseeability and liability (Williams, 2017: 25; McAllister, 2018: 47; King et al., 2019: 6–7).

These concerns over criminal liability and intentionality are likely to play a role in the context of algorithmic market manipulation, price fixing, and collusion (King et al., 2019: 9–12). In a 2016 experiment, computer scientists showed that AI trading agents can discover and learn to execute profitable strategies that amount to market manipulation. Using reinforcement learning, an 'artificial agent' explored the space of actions it could take on the market, and found that placing deceitful false buying orders was a profitable strategy (Miranda et al., 2016). Likewise, near-instantaneous pricing information among algorithms ensures that different companies' algorithms can sometimes artificially, inadvertently and tacitly settle on higher prices – essentially resulting in actions equivalent to collusion (Ezrachi and Stucke, 2017). Such behaviour can emerge quickly, possibly as a result of unanticipated interactions with other algorithms. While in many cases, these failures are easily spotted – compare the two pricing algorithms which, in 2011, engaged in a robotic price war over a book on flies, pushing up the price to US $23.7 million[23] (Sutter, 2011) – in other systems they are much harder to detect.

AIC: estimating the threat

Let us conclude our discussion of AIC with a question: will the type of crimes outlined above become commonplace, or will these examples prove yet again that what is easy in the lab can be difficult to scale in the real-world? To estimate how widely available these AI tools may be, it is instructive to draw on the example of the 'Blackshades Remote Access Tool', which, although not technically an AI application, can provide an illustration of how digital technologies can be disseminated and made rapidly accessible between different actors. Described as 'a criminal franchise in a box' (Markoff, 2016), and sold via PayPal for as little as US $40, Blackshades allowed users without any technical skills to effectively deploy ransomware and conduct eavesdropping operations. Prevalent in 2014, and only stopped after a major international cyber crackdown (Sullivan, 2014), Blackshades was, as the cybersecurity expert Brian Krebs (2014) succinctly observed, 'a tool created and marketed principally for buyers who wouldn't know how to hack their way out of a paper bag'. While Blackshades did not involve AI, it is not hard to see how criminal incentives are the same when it comes to emerging criminal AI tools. Many if not most of the AI capabilities described above are – or derive from – dual-use capabilities that are innocuous or beneficial in other applications. Moreover, the culture of AI is characterized by a high degree of openness, and even in cases where the source code is not already openly shared, many new AI algorithms can be independently reproduced by other researchers in a matter of months, making for a low barrier to proliferation (Brundage et al., 2018: 17; Shevlane and Dafoe, 2020). On the supply-side, AI tools, especially pre-trained versions, are as accessible as any software; on the demand-side, many of these tools offer extensions on, or improvements over, the precise sort of criminal capabilities or technologies which (cyber)criminals have long sought to acquire, whether in terms of pursuing 'zero-day exploits', or through tools such as Blackshades. This being the case, we estimate that AIC will be a major phenomenon within five years.[24]

Policing uses of AI

Having discussed the short-term future of AIC, we now address the reverse side of the criminological discussion. In the face of crimes both old and new, how will police departments harness these technologies to even – or overturn – the playing field? Furthermore, how might this further accelerate the existing militarization of police cultures (Wall and Linnemann, 2014)?

Recently there has been extensive attention on the potential uses of AI and robotics for law enforcement (INTERPOL and UNICRI, 2019; Zardiashvili et al., 2019), including critical examinations of how to ensure democratic accountability for ML-based predictive policing technologies (Vestby and Vestby, 2019). One question that frequently emerges, as in other areas of human activity, is whether AI and robots will ultimately replace human actors? Here, as Danaher (2018) suggests, a few distinctions must be made.

The first is between 'tasks' and 'jobs'. Police work involves a wide range of discrete tasks (patrolling, form-filling, ascribing 'a crime number', etc.). Here, AI may well be used to considerable effect.

A police officer's 'job', however, extends beyond these tasks, to broader roles of 'community policing', investigation, reassurance, arrest, and so on and these activities may be harder to replace. The second distinction is between AI's use as 'tool', 'partner' and 'usurper'. When technology is a simple *tool* of policing, it is used to assist in certain specific tasks that are part of a police officer's job. One could compare this to past advances in fingerprint analysis or DNA evidence.[25] For instance, recent research has demonstrated how ML can help with evidential links, such as recognizing a gun's caliber and model from audio recordings of shots (Raponi et al., 2020), matching crime-scene gunshot residue with the chemical characteristics of unspent ammunition (Gallidabino et al., 2018), or determining what type of shoe left a given imprint at a crime scene (Kong, 2017). To some extent, AI tools can even play a role in identifying and flagging signs of crimes mediated by other AI systems, as with the errant trading agents discussed above (King et al., 2019: 23–25).[26] When AI operates as *partner*, certain aspects of technology function autonomously, but still require human input and analysis – such as crime-prediction algorithms (see below). Finally, where technology serves as *usurper*, no human input is required. Keeping this framework in mind, we now outline three themes that we believe will be constitutive of the coming 'police-tech nexus'.

The seeing state: scalable, comprehensive, inescapable surveillance

First, AI promises (or threatens) the expansion of highly granular digital photography, described succinctly by the ACLU as the 'dawn of robot surveillance' (Stanley, 2019). Developments in data storage, along with advances in AI-enabled automatic video analytics, can turn passive, scattershot monitoring into an ever-more granular, comprehensive, and searchable surveillance record. Modern AI can already identify and distinguish emotions (Schwartz, 2019), forms of 'suspicious' behaviour (Schneier, 2019), and in one recent case was allegedly able to flag potential shoplifters by their body language, alerting grocery-store staff via a smartphone app (Du and Maki, 2019). Moreover, systems such as the 'iBorderCtrl' AI, funded since 2018 under the European Union's Horizon 2020 programme and since deployed at airports, purport to provide automated deception detection on the basis of facial micro-expressions, though many concerns have been raised over the scientific basis of this approach (Jupe and Keatley, 2019).

Beyond interpreting *what* people are doing in videos, AI can also recognize *who* is doing it (Phillips, 2018),[27] even when faces are disguised with masks (Singh et al., 2017). Moreover, AI surveillance is now very easy to graft onto extant surveillance infrastructures. Chinoy (2019), for example, used 'Rekognition' (Amazon's commercially-available facial recognition tool) to illustrate how easy it is to match employee photos from public sources against footage collected from three regular cameras around Bryant Park in New York City. In one day, the system detected 2,750 faces, including a local State University of New York (SUNY) professor. The total cost of the set-up was US $60.

Because extensive citizen name and face databases already exist, and because AI face-recognition can easily be layered on top of 'dumb' CCTV architectures, the ease and speed by which AI surveillance can be rolled out is truly staggering. Indeed, because faces (unlike fingerprints) are hard to conceal and can be scanned and recorded unknowingly from distance, some have argued that facial recognition is categorically different from other forms of surveillance and should be subject to an outright ban (Hartzog and Selinger, 2018). The ACLU, likewise, has expressed concerns that such algorithmic systems are untested, discriminatory, and subject to abuse (Stanley, 2019: 34–41). It is because of these and other related concerns that Stark (2019), a media scholar who works for Microsoft Research Montreal, referred to facial recognition as 'the plutonium of AI'.

The ability of AI systems to detect and infer identities does not, however, stop at cameras. Other technologies are now being trialled, including echolocation to identify human activity (Chen, 2019); 'Speech2Face', which allows users to reconstruct loosely identifying facial images – including age, gender and ethnicity – from voice audio alone (Oht et al., 2019); and the Pentagon's new 'Jetson' laser which identifies unique heartbeat signatures (through clothes) at a range of up to 200 metres. (Hambling, 2019). Increasingly, it seems one can scan anyone with almost anything. Rolled out society-wide, these developments make for a powerful tool – not just for fighting crime, but for social control writ large. China has an estimated 200 million surveillance cameras countrywide, and has reportedly begun incorporating AI in these systems (Mozur, 2018). In some cases, this results in explicitly racialized surveillance systems, with algorithms configured to flag members of the ethnic Uighur minority by their facial features (Mozur, 2019).

One further dimension is the increasing intermingling of state police capabilities with private companies. For instance, Axon Enterprise (formerly Taser International) supplies 47 of the 69 largest US police agencies with body cameras, and has been involved in marketing an AI system trained on 30 petabytes of video (over ten times larger than the Netflix database) collected from 200,000 officers. This system processes bodycam footage to assist police by anticipating problems and generating situation reports (Perry, 2018).[28] Among a host of concerns (Patterson and Greene, 2018), Joh (2017) has argued that these developments also highlight how private surveillance vendors exercise undue influence over the investigative and arrest practices of police departments.[29] Furthermore, such unease is exacerbated by meaningful concerns over proprietary datasets and software, which typically are not publicly available. More generally, others have raised concerns about how the military heritage underlying many digital technologies such as the internet may inflect their usage for surveillance (Levine, 2018). These developments and others have also been read in the context of the creeping construction of larger architectures of 'surveillance capitalism' (Zuboff 2019).

The hidden state: ubiquitous yet tacit surveillance

Second, the integration of AI with drones and 'smart-city' sensors creates new forms of 'wide-area surveillance' that are ubiquitous, yet subtle, tacit, and deniable. In terms of ubiquity, the falling cost of sensors and drone platforms, coupled with the increasing 'stand-off' distance of camera functionality, is greatly extending the reach of AI surveillance. Today, accurate gigapixel cameras can recognize faces and licence plates in photos taken kilometres away (Schneier, 2019) such that a single drone overflight of a protest could in principle enable authorities to compile a list of all attendees.[30]

In fact, such capabilities are not even that new: a decade ago, DARPA introduced ARGUS-IS, a 1.8 gigapixel unmanned video drone platform capable of continuously recording an area of 25-square kilometres with a resolution of 15 cm (Hambling, 2009).[31] In 2014, the US Air Force integrated this programme into the 'Gorgon Stare' system, which deploys Wide-Angle-Motion-Imagery (WAMI) to enable drones to track multiple targets over large areas. These early applications of Gorgon Stare were beset with technical problems (Cockburn, 2016). However, subsequent iterations have proven more efficient, and have seen expanded use, most infamously as an investigative law enforcement tool in Baltimore (Michel, 2019). This merging of WAMI with other digital and biometric sensors ushers in a new era of the 'fully fused' or 'captured' (Sadowski, 2019; Sadowski and Bendor, 2019) city.

In another application of stand-off 'stealth surveillance', China has developed drones in the shape of small robotic 'doves' allowing them to blend in with bird flocks across several provinces (Chen, 2018). As a citizen, it is easy to object to clearly visible public surveillance cameras; it is much harder to notice, let alone resist, distant, unseen technologies, especially those that do not simply watch, but also infer sensitive facts from statistical projections based on one's demographic profile.

Indeed, in some cases, the governmental role in explicitly monitoring behaviour is rendered opaque because it is further sublimated into a technologically mediated decentralized system of social control. Such is the case in the much-discussed Chinese Social Credit system – in reality a patchwork of different systems which collect data on a range of online and offline activities. This information is then aggregated into a score out of 800, which is then tied to benefits, discounts and other incentives (Mozur, 2018). Others have cautioned that the sophistication and Orwellian reach of the Chinese Social Credit System is often overstated (Ahmed, 2019). Yet even in its nascent stage, it already demonstrates one 'evolving practice of control' (Creemers, 2018) that will only be strengthened as AI allows the further leveraging of citizen data.

More abstractly, these trends also demonstrate how technology accelerates and exacerbates the underlying transition in the means by which governments seek to regulate citizens' behaviour. This marks a shift away from traditional overt and explicitly normative enforcement of law, to the non-intrusive shaping of (urban) architecture and space. For instance, Joh (2019) has described how policing in the smart city follows the Disneyland model, likening it to the way that some high-tech amusement parks anticipate and prevent disorder by shaping visitor's behaviour through physical barriers, as well as through the omnipresence of employees who notice and intercept errors. Such architectures do not feel intrusive or coercive, but can still be effective tools of governance. We see this in smart cities in the development of (algorithmic) tools to 'hypernudge' (Yeung, 2017) citizens into adopting certain prosocial behaviours.

Finally, Brownsword (2015) has extended this argument, suggesting that the emergence of regulatory technologies of control (including but not limited to AI), can lead to a shift among the 'regulatory modality'. He illustrates this argument with reference to a golf club that is experiencing problems with visitors driving golf carts over flowerbeds. According to Brownsword, the possible (successive) preventive options open to the club mirror or anticipate a larger trend in policing. Originally, the golf club relied on social norm-enforcement among members (shaming; censure for flowerbed-violators); when this proved insufficient, they switched to formal normative 'law' (setting up rules with specified fines for members caught damaging the flowers); the enforcement of this normative rule was eventually supported through the use of technology (the use of CCTV cameras to monitor violations). Finally, however, technology enabled the wholesale substitution of the normative rule: GPS chips were embedded in the golf carts, and the flowerbed areas geo-fenced to ensure that golf carts would shut down when approaching the flowerbeds. This example of a fundamental shift to a non-normative regulatory modality for controlling human behaviour, nicely illustrates how, in the future, AI technologies will come to facilitate the sublimation of policing architectures.

The oracle state: from detection and enforcement, to prediction and prevention

Third, as in other fields, AI systems can pick up on subtle patterns to offer (ostensibly) accurate predictions of future behaviour, including criminal conduct. Increasingly, this facilitates a shift in policing practices, from those aimed at detecting violations in order to enforce the law, to those that seek to predict criminal acts in order to prevent them entirely (Danaher, 2018). This is seen in high-profile debates over the use of algorithms to predict re-offence rates for pre-trial bail decisions. In one study, Kleinberg et al. (2017) found that their algorithm outperformed human judges at predicting a defendant's risk of re-offending. The authors argued that adopting this mode of prediction could yield 'potentially large welfare gains [. . .] crime can be reduced by up to 24.8% with no change in jailing rates, or jail populations can be reduced by 42.0% with no increase in crime rates'. While such results are alluring, one must be mindful of both the hype surrounding these systems, and their underlying 'political patterning' (see Kaufmann et al., 2018).

To start with, many of the (in)famous examples of 'predictive policing' do not in fact involve all that much AI. The predictive programme run by Palantir Technologies in New Orleans from 2012 onwards (in cooperation with police, but without the city council's knowledge), was based on human-curated social network-mapping and fairly simple scoring algorithms (Winston, 2018). Likewise, 'PredPol' primarily looks at just three variables of crime, in order to create 'crime hotspots' to guide police resource allocation and patrol routes. This is a far cry from the complex patterns distilled by deep neural networks.

Other problems abound – including questions about accuracy. For example, there is considerable contestation about whether the recidivism predictions of the much-vaunted COMPAS programme are any more accurate than those made by random people (Dressel and Farid, 2018; Lin et al, 2020). Various studies have suggested that these programmes are plagued by baked-in racial bias (Kirchner et al., 2016; Lum and Isaac, 2016, but see, however, Kamyshev, 2019), whether they use AI or conventional statistics. In other words, if predictive AI systems are trained on unrepresentative or biased datasets, the inevitable result is a 'runaway feedback loop' of self-confirming predictions (Ensign et al., 2017). As the algorithm designates certain areas as at high risk of crime, police forces dispatch more patrols, ensuring they arrest proportionally more people committing crime, which the algorithm then processes as further evidence of a high-risk crime area. In effect, the system corrupts its own future training data. As such, as Kamyshev (2019) has argued, the policing potential of predictive AI systems is undermined by their actual sub-par accuracy, lack of transparency, susceptibility to self-corrupting feedback loops, and failure to cohere to the basic goals of a judicial system.

AI and policing: predictions and meditations

Discussion of AI policing tools can rapidly turn dystopian. But how widely will such tools proliferate? On the one hand, the suppliers and customers are there. For instance, in 2018, Lookout and the Electronic Frontier Foundation (2018) revealed an extensive spying campaign by an elusive group called 'Dark Caracal', which used advanced hacking and surveillance tools apparently disseminated by an unknown third party, which had supplied at least half a dozen other surveillance campaigns.[32] While it should be stressed that Dark Caracal did not involve AI tools, it is not hard to see how, in the future, more sophisticated 'AI-surveillance-by-subscription' tools would be very appealing to some governments as indicated by reports that show how China is selling plug-and-play surveillance tech to states such as Ecuador (Mozur et al., 2019). Conversely, it is also easy to overstate the ease and speed of adoption of new surveillance technologies. For instance, in his study of the Danish police, Sausdal (2018) showed that, contrary to the bold claims of tech companies, detectives actually saw high-tech surveillance tools as frustrating and frequently a hindrance to their work.

The first law of technology, as coined by Melvin Kranzberg (1986), reads, 'technology is neither good nor bad; nor is it neutral'. The above discussion has certainly demonstrated room for concern over policing uses of AI. At the same time, just as we reject naïve techno-optimism, we must also not fall prey to doom-and-gloom techno-pessimism. In truth, much of AI's net impact will depend on critical political and cultural choices which societies will make more generally. At least as significantly, these technologies may come to shift the very terms of the societal trade-offs which we have long taken as axiomatic. To give one provocative example: the term 'privacy-preserving surveillance' may strike some as an oxymoron. Yet while criminologists are right to approach such new concepts with suspicion, they will nevertheless have to engage with new developments in AI – in the areas of 'homomorphic encryption' or 'federated learning' – that could potentially develop monitoring systems that are (at least on the surface) less intrusive and more accountable than old surveillance approaches. The dissolving or softening of such trade-offs is not unprecedented: the introduction of sniffer dogs at airports offered a new way to detect drugs or bombs, which was both more effective and less intrusive than previous security measures (Trask, 2017). Likewise, if configured and utilized appropriately, AI could even serve as a 'privacy-enhancing technology' (Els, 2017; see, Birnstill, et al., 2015), or at least potentially reduce the intrusiveness of policing, transforming digital surveillance from a blunt to a sharp instrument. At a larger scale, the ways we choose to resolve the safety/privacy balance may also come to be reassessed, if or when increasingly powerful technologies usher us into a 'vulnerable world' (Bostrom, 2019). We offer this example not in the hope of convincing the reader. Instead, we propose it is an illustration of the debates that new AI systems will – and perhaps should – re-open. At its best, we hope that AI, and the choices societies make around it, can help criminologists re-examine – and if necessary, reconsider – certain foundational or treasured assumptions that, like it or not, are going to be tested by the coming new technological age.

Conclusion

This article introduced the concept of AI to a criminological audience by offering a cautionary but measured overview of the technology's workings, applications, strengths and limits. One could summarize our argument simply, as follows: for all its utility, *AI is not magic.* Just like any data-driven programme, its objectivity and efficacy are still determined by that old computational axiom, 'GIGO'. Indeed, given the essentially brittle nature of neural network decision-making, it's clear that fine-grained human expertise is even more important in computing today than it has ever been; both in relation to parameter/hypothesis framing and overall system governance.

This article's ambitions, however, extended beyond simply striking a balance between overwrought narratives of tech dystopia and Silicon Valley utopia (see e.g. Barbrook and Cameron 2001, on the 'Californian Ideology'). Most of all, we have a more specific, disciplinary aim to encourage criminologists of all stripes to extend their research interests towards the 'tech-crime nexus'. Many criminologists will naturally balk at such an idea, believing that without a mathematics or computer science degree they are incapable of making an informed contribution to debates about AI and ML. We disagree. Indeed, as AI's impact on policing, punishment, sentencing, and inevitably criminality, continues to grow, the need for criminologists to engage fully with networked technology and its complexity is not just desirable but essential if we are to stand any chance of limiting its potential excesses.

Our thinking here is shaped by what James Bridle (2018) calls *the chasm of computational thinking*: the disconcerting tendency evident across all spheres of contemporary life, from education to warfare, to cede power and problem-prioritization to reified technologies and networked systems connected to vast repositories of data in the belief that any social challenge can be solved solely by the application of computation and technological acceleration. For Bridle (2018), the unthinking faith in a combination of information and automation represents a 'cognitive hack' wherein decision-making and consciousness are offloaded onto the machine, resulting in 'an ever-increasing opacity allied to a concentration of power, and the retreat of that power into ever more narrow domains of experience' (p. 34). To counter this direction of travel, criminologists must do more than simply criticize AI's problematic tendencies. Instead, we must look to proactively 'shape and direct' the conversation about technology within our field. Only then, when we have expanded the criminological imagination sufficiently to fully embrace the tech-crime nexus will we be in a position to ensure digital systems and practices that are both ethical and non-discriminatory.

Finally, a word about nomenclature. Over the years, criminologists have proven extremely creative when it comes to adding a prefix to their discipline. In recent times we have seen the emergence of a host of interesting subfields such as 'border criminology', 'visual criminology', 'queer criminology', 'Southern criminology', and now even a 'ghost criminology'. As criminologists turn their attention to the study of technology, further prefixes are likely to emerge. However, if criminology is to fully engage with the complex algorithms, networks and digital infrastructures now mediating human beings and their environment, further fracturing the discipline into sub-specialisms is not enough. Much better to make changes at a more universal level; to forge a rounded, tech-literate criminology that will be able to deal with the next wave of scientific disruptions and realignments as they arrive. It is for this reason that we are not calling here for anything as specific as 'a criminology of AI'. Such a development is too narrow and would, in due time, need to be buttressed (or replaced) by the likes of, say, a 'quantum-computing criminology', or a 'biohacking criminology', and so on. If, as now seems clear, radical new technologies are redrawing the contours of the existing liberal order with profound implications for crime and punishment, criminology must adapt if it is to remain relevant. This will, of course, involve reimaging criminology's existing theoretical and methodological horizons, including an overdue abandonment of some of the classical models of human behaviour and theories of crime devised in the twentieth century. To do anything less would mean to risk criminology's obsolescence.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iDs

Keith J Hayward https://orcid.org/0000-0001-7135-9131
Matthijs M Maas https://orcid.org/0000-0002-6170-9393

Notes

1. https://www.youtube.com/watch?v=QUK6rEUZAcA
2. This is not, of course, a universal position. Over the same period, one group of computer scientists – cybersecurity experts – have done much to promote a more cautious 'security mind-set' (Schneier, 2008).
3. Even when caution is exercised, this does not necessarily garner unanimous support from a tech community that adheres to a long-established open-source culture. In 2019, the Elon Musk-backed non-profit OpenAI developed 'GPT-2', a language model capable of composing coherent prose (including news releases) given just two sentences of context (demo available at https://transformer.huggingface.co/). Initially, OpenAI only publicly released smaller versions of its system, expressing concern that the full tool might be used by malicious actors 'to generate deceptive, biased, or abusive language at scale' (Radford et al., 2019). This decision sparked an ongoing – and at times divisive – debate in the AI community over when, if ever, it is appropriate to withhold AI research (Leibowicz et al., 2019), with another lab later releasing their own 'neural fake news' AI, to help researchers identify fake news (Zellers et al., 2019). Eventually, by November 2019, OpenAI released the full model, along with some reflections on future responsible release strategies for AI applications (Solaiman et al., 2019).
4. We define 'technology' here as any combination of tools, skills, processes, and techniques by which human capability is extended (Bennett Moses, 2007: 592).
5. King et al. (2019: 2) describe AIC as 'a relatively young and inherently interdisciplinary area – spanning socio-legal studies to formal science'.
6. Naturally, we acknowledge the pioneering work on cybercrime by Shelia Brown, Yvonne Jewkes, David Wall, Majid Yar and others. However, at this point we would stress the key distinction between early research into 'online crimes' and the potentially more expansive world of criminality associated with radical new technologies.
7. For example, in 2011 a little robot named Qbo passed a 'mirror test', chirping, '*Oh. This is me. Nice*', as it was prompted with its reflection (Ackerman, 2011). These experiments are often heralded in the media as evidence of 'self-aware' robots. But to most researchers such tests reveal more about the shortcomings of psychological testing than offering any profound evidence of 'sentient' AI.
8. The Turing test itself was 'passed' in 2014 by 'Eugene Goostman', a chatbot which convinced 33% of the judges it was human by impersonating a Ukrainian schoolboy.
9. The human-centric definition is utilized in King et al.'s (2019) AIC study. However, we consider that definition too narrow; while 'human-like' autonomous systems may pose peculiar challenges, many of the most potent criminal – or policing – uses of AI involve broader non-human and non-anthropomorphic conceptions of intelligence.
10. When using words such as 'learn' or 'discover' in the context of AI-systems, we should stress that we are deploying these terms in an explicitly non-anthropomorphic sense.
11. Surveys show AI-experts expect such capabilities within the next three to five decades (Grace et al., 2018).
12. It is important to acknowledge that some of the selected examples that follow derive from private research labs, and as such may have vested corporate interests in exaggerating the sophistication or performance of their system(s). As such it may be wise to treat some of the claims made with a degree of caution.
13. Hypothetically, one could entertain a fourth category of crimes against AI (as rights-carrying 'person'), but this is conditional on their being granted legal status. There is also another indirect category – the role of AI or robots in promoting general criminality, or *provoking* certain crimes – such as the concern over how interaction with social bots and 'sexbots' might desensitize perpetrators towards sexual harm (Danaher, 2017; King et al., 2019: 15–16).
14. For a dramatized depiction of this scenario, see the video 'Slaughterbots', by the Future of Life Institute, at https://www.youtube.com/watch?v=HipTO_7mUOw&t=203s .
15. One such fake account managed to connect with Paul Winfree, former Deputy Head of President Trump's domestic policy council. When contacted for comment, Winfree admitted that 'I literally accept every friend request that I get'.

16. For a vivid example, see https://thispersondoesnotexist.com/, which generates faces of non-existing people. At http://www.whichfaceisreal.com/, you can attempt to distinguish real from fake people.
17. One egregious example of DeepFake's potential for gendered violence was 'DeepNude', a June 2019 commercially available, $50 app which removed clothing from images of women, making them look realistically nude (Cole, 2019). The app was soon taken down after widespread outcry.
18. See Shumailov et al. (2019) on smartphone 'acoustic side channel attacks', which can detect the sound of fingers on touch-screen keyboards, recovering 61% of 4-digit PIN-codes within 20 attempts.
19. In another case, cybersecurity researchers used lasers to silently manipulate the microphones of computer voice-command systems. Capable of penetrating window glass, these so-called 'light commands' further expose digital locks and other smart household appliances to criminal exploitation (Sugawara et al., 2019).
20. See: https://nicholas.carlini.com/code/audio_adversarial_examples/
21. Interestingly, the researchers indicated that, while this defence only worked on one specific system, they would aim to generate images that work on multiple detectors simultaneously (Knight, 2019).
22. At the unfortunate cost of making one highly conspicuous to good old-fashioned human surveillance. See the project page, https://cvdazzle.com/
23. Plus $3.99 shipping.
24. To evidence this point, it was recently estimated that DeepFakes alone will, by the end of 2020, have accounted for in excess of $250 million in personal and corporate damages (Forrester, 2019).
25. However, in court cases where fingerprint and DNA evidence were first introduced, judges showed themselves too easily wowed by the technology, crediting these techniques with a degree of evidential authority and infallibility which they did not deserve. See also Alldredge (2015) on the so-called 'CSI effect'.
26. This does not mean that using AI as a *tool* will be uncontroversial – see, for instance, the problems that ensued when AI was used as a 'lie detector' in immigration procedures (Kendric, 2019; Molnar, 2019; Beduschi, 2020), or when facial recognition scans were implemented for prison visitors in England and Wales (Jee, 2019).
27. This does not mean that such technology is flawless. In 2018, the ACLU found that Amazon's 'Rekognition' system incorrectly matched 28 members of Congress with criminal mugshots (Snow, 2018). Moreover, facial or emotion-recognition systems have been plagued by racial bias (Rhue, 2018).
28. Notably, in June 2019, Axon announced a moratorium on the use of facial recognition in its bodycam devices on the advice of its independent ethics board, who deemed that such systems were not yet accurate enough (Warzel, 2019). However, an Axon spokesperson confirmed that police officials could in principle download the body cam footage and run it through third-party facial recognition services.
29. This development is salient in light of recent criminological research which reviews police body camera programmes, and found they 'have not had statistically significant or consistent effects on most measures of officer and citizen behavior or citizens' views of police' (Lum et al., 2019: 93).
30. Earlier this year in China, researchers used algorithms to process footage from a lidar-based camera, mounted on a Shanghai skyscraper, that was capable of resolving human-sized features through smog at a distance of 45 km (Li et al., 2019; MIT Technology Review, 2019).
31. Readers of this journal might note the irony that the original idea behind WAMI was conceived by an unnamed military scientist after he saw the 1998 Hollywood thriller *Enemy of the State*, a movie in which Will Smith is tracked by a rogue state agency deploying advanced satellite surveillance.
32. It is suspected that Dark Caracal is itself a hacking group sponsored by an unknown nation state. This illustrates the degree to which states may, publicly or through proxies, find ways to sell new surveillance technology to unscrupulous parties.

References

Ackerman E (2011) Qbo robot passes mirror test. *IEEE Spectrum: Technology, Engineering, and Science News*, 6 December. Available at: https://spectrum.ieee.org/automaton/robotics/artificial-intelligence/qbo-passes-mirror-test-is-therefore-selfaware

Ahmed S (2019) The messy truth about social credit. *Logic Magazine*, 1 May. Available at: https://logicmag.io/china/the-messy-truth-about-social-credit/

Ajder H, Patrini G, Cavalli F et al. (2019) The state of DeepFakes: landscape, threats, and impact. *Deeptrace Labs*, September. Available at: https://regmedia.co.uk/2019/10/08/deepfake_report.pdf

Aldridge J (2019) Does online anonymity boost illegal market trading? *Media, Culture & Society* 41: 578–583.

Alldredge J (2015) The 'CSI Effect' and its potential impact on juror decisions. *Themis: Research Journal of Justice Studies and Forensic Science* 3: 15.

Athalye A, Engstrom L, Ilyas A et al. (2018) Synthesizing robust adversarial examples. Available at: https://arxiv.org/abs/1707.07397

Bahnsen AC, Torroledo I, Camacho LD et al. (2018) DeepPhish: simulating malicious AI June 9. https://pdfs.semanticscholar.org/ae99/765d48ab80fe3e221f2eedec719af80b93f9.pdf?_ga=2.137195056.1064399283.1590653531-1585409390.1590653531

Barbrook R and Cameron A (2001) The Californian ideology. In: Ludlow P (ed.) *Crypto Anarchy, Cyberstates, and Pirate Utopias*. Cambridge: MIT Press, pp. 363–387.

Barocas S and Selbst AD (2016) Big Data's disparate impact. *California Law Review* 104: 671–732.

Bayern S (2016) The implications of modern business–entity law for the regulation of autonomous systems. *European Journal of Risk Regulation* 7: 297–309.

Beduschi A (2020) International migration management in the age of artificial intelligence. *Migration Studies*. DOI: 10.1093/migration/mnaa003.

Bennett Moses L (2007) Why have a theory of law and technological change? *Minnesota Journal of Law, Science & Technology* 8: 19.

Bilge L, Strufe T, Balzarotti D et al. (2009) All your contacts belong to us: automated identity theft attacks on social networks. In: *Proceedings of the 18th international conference on world wide web*, Madrid, 20–24 April, pp. 551–560. New York: ACM.

Birnstill P, Bretthauer S, Greiner S et al. (2015) Privacy-preserving surveillance: an interdisciplinary approach. *International Data Privacy Law* 5(4): 298–308.

Bontrager P, Roy A, Togelius J et al. (2017) DeepMasterPrints: generating MasterPrints for dictionary attacks via latent variable evolution. Available at: https://arxiv.org/abs/1705.07386

Bostrom N (2019) The vulnerable World hypothesis. *Global Policy* 10: 455–476.

Bridle J (2018) *New Dark Age*. London: Verso.

Broadhurst R, Maxim D, Brown P et al. (2018) *Artificial Intelligence and Crime: A Report for the Korean Institute of Criminology*. Canberra, ACT, Australia: ANU Cybercrime Observatory.

Brown S (2006) The criminology of hybrids: rethinking crime and law in technosocial networks. *Theoretical Criminology* 10: 223–244.

Brown TB, Mané D, Roy A et al. (2017) Adversarial patch. Available at: https://arxiv.org/abs/1712.09665

Brownsword R (2015) In the year 2061: from law to technological management. *Law, Innovation and Technology* 7: 1–51.

Brundage M, Avin S, Clark J et al. (2018) The malicious use of artificial intelligence. Available at: https://arxiv.org/ftp/arxiv/papers/1802/1802.07228.pdf

Carlini N and Wagner D (2017) Adversarial examples are not easily detected: bypassing ten detection methods. In: *Proceedings of the 10th ACM workshop on artificial intelligence and security (AISec '17)*, Dallas, TX, 3 November, pp. 3–14. New York: ACM.

Chan J and Bennett Moses L (2017) Making sense of Big Data for security. *British Journal of Criminology* 57: 299–319.

Chen S (2018) China's robotic spy birds take surveillance to new heights. *South China Morning Post*, 24 June. Available at: https://www.scmp.com/news/china/society/article/2152027/china-takes-surveillance-new-heights-flock-robotic-doves-do-they

Chen S (2019) This AI uses echolocation to identify what you're doing. *Wired*, 28 May. Available at: https://www.wired.com/story/this-ai-uses-echolocation-to-identify-what-youre-doing/

Chesney R and Citron DK (2019) Deep fakes: a looming challenge for privacy, democracy, and national security. *California Law Review* 107: 1753.

Chinoy S (2019) We built an ‘unbelievable’ (but legal) facial recognition machine. *The New York Times*, 16 April. Available at: https://www.nytimes.com/interactive/2019/04/16/opinion/facial-recognition-new-york-city.html

Cockburn A (2016) *Kill Chain: The Rise of the High-Tech Assassins*. London: Verso.

Cole S (2019) This horrifying app undresses a photo of any women with a single click. *Motherboard*, 26 June. Available at: https://me.me/i/motherboard-this-horrifying-app-undresses-a-photo-of-any-woman-b07945025d024a4aa830a505dc09cc24

Creemers R (2018) *China’s Social Credit System: An Evolving Practice of Control*. Rochester, NY: Social Science Research Network.

Dalton A, Aghaei E, Al-Shaer E et al. (2020) The Panacea Threat Intelligence and Active Defense Platform. Available at: http://arxiv.org/abs/2004.09662

Danaher J (2017) Robotic rape and robotic child sexual abuse: should they be criminalised? *Criminal Law, Philosophy* 11: 71–95.

Danaher J (2018) The automation of policing: challenges and opportunities. Available at: https://philosophicaldisquisitions.blogspot.com/2018/10/the-automation-of-policing-challenges.html (accessed 15 October 2018).

Domigos P (2015) *The Master Algorithm*. New York: Basic Books.

Dressel J and Farid H (2018) The accuracy, fairness, and limits of predicting recidivism. *Science Advances* 4: eaao5580.

Du L and Maki A (2019) These cameras can spot shoplifters even before they steal. *Bloomberg*, 4 March. Available at: https://www.bloomberg.com/news/articles/2019-03-04/the-ai-cameras-that-can-spot-shoplifters-even-before-they-steal

Eckert M-L, Kose N and Dugelay J-L (2013) Facial cosmetics database and impact analysis on automatic face recognition. In: *2013 IEEE 15th international workshop on multimedia signal processing (MMSP)*, Pula, 30 September–2 October 2013, pp. 434–443. New York: IEEE.

Els AS (2017) Artificial intelligence as a digital privacy protector. *Harvard Journal of Law & Technology* 31: 217–235.

Elsayed G, Goodfellow I and Sohl-Dickstein J (2018) Adversarial reprogramming of neural networks. Available at: https://arxiv.org/abs/1806.11146

Ensign D, Friedler SA, Neville S et al. (2017) Runaway feedback loops in predictive policing. Available at: https://arxiv.org/abs/1706.09847

Evtimov I, Eykholt K, Fernandes E et al. (2017) Robust physical-world attacks on deep learning models. Available at: https://arxiv.org/abs/1707.08945

Ezrachi A and Stucke ME (2017) *Two Artificial Neural Networks Meet in an Online Hub and Change the Future (Of Competition, Market Dynamics and Society)*. Rochester, NY: Social Science Research Network.

Finlayson SG, Bowers JD, Ito J et al. (2019) Adversarial attacks on medical machine learning. *Science* 363: 1287–1289.

Forrester (2019) *Predictions 2020: On the Precipice of Far-Reaching Change*. Forrester Research, 30 October. Available at: https://go.forrester.com/predictions-2020/

Gallidabino MD, Barron LP, Weyermann C et al. (2018) Quantitative Profile–Profile Relationship (QPPR) modelling: a novel machine learning approach to predict and associate chemical characteristics of unspent ammunition from Gunshot Residue (GSR). *Analyst*. DOI: 10.1039/C8AN01841C.

Gershgorn D (2016) Here’s how we prevent the next racist chatbot. *Popular Science*, 24 March. Available at: https://www.popsci.com/heres-how-we-prevent-next-racist-chatbot (accessed 21 February 2019).

Gholipour B (2017) New AI tech can mimic any voice. *Scientific American*, 2 May. Available at: https://www.scientificamerican.com/article/new-ai-tech-can-mimic-any-voice/

Goodfellow IJ, Papernot N, Huang S et al. (2017) Attacking machine learning with adversarial examples. *OpenAI Blog*, 24 February. Available at: https://blog.openai.com/adversarial-example-research/

Goodfellow IJ, Shlens J and Szegedy C (2014) Explaining and harnessing adversarial examples. Available at: https://arxiv.org/abs/1412.6572

Grace K, Salvatier J, Dafoe A et al. (2018) When will AI exceed human performance? Evidence from AI experts. *Journal of Artificial Intelligence Research* 62: 729–754. DOI: 10.1613/jair.1.11222.

Greenfield A (2017) *Radical Technologies*. London: Verso.

Gu T, Dola-Gavitt B and Gar S (2017) BadNets: identifying vulnerabilities in the machine learning model supply chain. Available at: https://arxiv.org/abs/1708.06733

Hambling D (2009) Special forces' Gigapixel flying spy sees all. *Wired*, 12 February. Available at: https://www.wired.com/2009/02/gigapixel-flyin/

Hambling D (2019) The Pentagon has a laser that can identify people from a distance—by their heartbeat. *MIT Technology Review*, 27 June. Available at: http://www.technologyreview.com/s/613891/the-pentagon-has-a-laser-that-can-identify-people-from-a-distanceby-their-heartbeat

Hartzog W and Selinger E (2019) Why you can no longer get lost in the crowd. *The New York Times*, 17 April. Available at: https://www.nytimes.com/2019/04/17/opinion/data-privacy.html

Harwell D (2019) An artificial-intelligence first: voice-mimicking software reportedly used in a major theft. *Washington Post*, 4 September. Available at: https://www.washingtonpost.com/technology/2019/09/04/an-artificial-intelligence-first-voice-mimicking-software-reportedly-used-major-theft/

Hayward KJ (2012) Five spaces of cultural criminology. *The British Journal of Criminology* 52(3): 441–462.

Hern A (2018) Fake fingerprints can imitate real ones in biometric systems – research. *The Guardian*, 15 November. Available at: https://www.theguardian.com/technology/2018/nov/15/fake-fingerprints-can-imitate-real-fingerprints-in-biometric-systems-research

Hitaj B, Gasti P, Ateniese G et al. (2017) PassGAN: a deep learning approach for password guessing. Available at: https://arxiv.org/pdf/1709.00440.pdf

Holt TJ and Bossler AM (2014) An assessment of the current state of cybercrime scholarship. *Deviant Behavior* 35: 20–40.

INTERPOL and UNICRI (2019) *Artificial Intelligence and Robotics for Law Enforcement*. Available at: http://www.unicri.it/news/files/ARTIFICIAL_INTELLIGENCE_ROBOTICS_LAW%20ENFORCEMENT_WEB.pdf

Jagatic TN, Johnson NA, Jakobsson M et al. (2007) Social phishing. *Communications of the ACM* 50: 94–100.

Jee C (2019) Prisons are using face recognition on visitors to prevent drug smuggling. *MIT Technology Review*, 6 March. Available at: https://www.technologyreview.com/the-download/613080/prisons-are-using-face-recognition-on-visitors-to-prevent-drug-smuggling/ (accessed 9 March 2019).

Jin D (2020) *Jind11/TextFooler*. Python. Available at: https://github.com/jind11/TextFooler

Joh EE (2017) The undue influence of surveillance technology companies on policing. *NYU Legal Review Online*. Available at: https://www.nyulawreview.org/online-features/the-undue-influence-of-surveillance-technology-companies-on-policing/

Joh EE (2019) Policing the smart city. *International Journal of Law in Context* 15(2): 177–182.

Jupe LM, and Keatley DA (2019) Airport artificial intelligence can detect deception: or am i lying? *Security Journal*. DOI: 10.1057/s41284-019-00204-7.

Kamyshev P (2019) Machine Learning In The Judicial System Is Mostly Just Hype. *Palladium Magazine*. Available at: https://palladiummag.com/2019/03/29/machine-learning-in-the-judicial-system-is-mostly-just-hype/

Kasperkevic J (2015) Swiss police release robot that bought ecstasy online. *The Guardian*, 22 April. Available at: https://www.theguardian.com/world/2015/apr/22/swiss-police-release-robot-random-darknet-shopper-ecstasy-deep-web

Kaufmann M, Egbert S and Leese M (2018) Predictive policing and the politics of patterns. *British Journal of Criminology* 59: 674–692.

Kendric M (2019) The border guards you can't win over with a smile. *BBC Future*, 17 April. Available at: https://www.bbc.com/future/article/20190416-the-ai-border-guards-you-cant-reason-with

King TC, Aggarwal N, Taddeo M et al. (2019) Artificial intelligence crime: an interdisciplinary analysis of foreseeable threats and solutions. *Science and Engineering Ethics* 26: 89–120.

Kirchner L, Angwin J, Larson J et al. (2016) Machine bias: there's software used across the country to predict future criminals. And it's biased against blacks. *ProPublica*, 23 May. Available at: https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

Kleinberg J, Lakkaraju H, Leskovec J et al. (2017) *Human Decisions and Machine Predictions*. Cambridge, MA: National Bureau of Economic Research.

Knight W (2017) Alpha zero's 'Alien' chess shows the power, and the peculiarity, of AI. *MIT Technology Review*, 8 December. Available at: https://www.technologyreview.com/s/609736/alpha-zeros-alien-chess-shows-the-power-and-the-peculiarity-of-ai/

Knight W (2019) How to hide from the AI Surveillance State with a color printout. *MIT Technology Review*, 23 April. Available at: https://www.technologyreview.com/f/613409/how-to-hide-from-the-ai-surveillance-state-with-a-color-printout/

Knight W (2020) This technique uses AI to fool other AIs. *Wired*, 23 February. Available at: https://www.wired.com/story/technique-uses-ai-fool-other-ais/

Kolosnjaji B, Demontis A, Biggio B et al. (2018) Adversarial malware binaries: evading deep learning for malware detection in executables. Available at: https://arxiv.org/abs/1803.04173

Kong B (2017) Cross-domain forensic shoeprint matching. Available at: https://www.ics.uci.edu/~fowlkes/papers/KongSRF_BMVC_2017.pdf.

Kranzberg M (1986) Technology and history: 'Kranzberg's Laws'. *Technology and Culture* 27: 544–560.

Krebs B (2014) Blackshades Trojan users had it coming. *Krebs on Security*, 14 May. Available at: https://krebsonsecurity.com/2014/05/blackshades-trojan-users-had-it-coming/ (accessed 18 February 2019).

Kruithof K, Aldridge J, Hétu DD et al. (2016) *The Role of the 'Dark Web' in the Trade of Illicit Drugs*. Santa Monica, CA: RAND.

Legg S and Hutter M (2007a) A collection of definitions of intelligence. Available at: https://arxiv.org/abs/0706.3639

Legg S and Hutter M (2007b) Universal intelligence: a definition of machine intelligence. Available at: https://arxiv.org/abs/0712.3329

Lehman J, Clune J, Misevic D et al. (2018) The surprising creativity of digital evolution: a collection of anecdotes from the evolutionary computation and artificial life research communities. Available at: https://arxiv.org/abs/1803.03453

Leibowicz C, Adler S and Eckersley P (2019) When is it appropriate to publish high-stakes AI research? *The Partnership on AI*, 2 April. Available at: https://www.partnershiponai.org/when-is-it-appropriate-to-publish-high-stakes-ai-research/

Levine Y (2018) *Surveillance Valley: The Secret Military History of the Internet*. New York: Public Affairs.

Li Z-P, Huang X, Cao Y et al. (2019) Single-photon computational 3D imaging at 45 km. Available at: https://arxiv.org/abs/1904.10341

Lin Z, Jung J, Goel S et al. (2020) The limits of human predictions of recidivism. *Science Advances* 6(7): eaaz0652. Available at: https://doi.org/10.1126/sciadv.aaz0652

Lookout and Electronic Frontier Foundation (2018) *Dark Caracal: Cyber-espionage at a Global Scale*. San Francisco, CA: Lookout.

LoPucki LM (2017) Algorithmic entities. Law-Econ research paper, UCLA School of Law, Los Angeles, CA.

Lum C, Stoltz M, Koper CS et al. (2019) Research on body-worn cameras. *Criminology & Public Policy* 18: 93–118.

Lum K and Isaac W (2016) To predict and serve? *Significance* 13: 14–19.

McAllister A (2018) Stranger than science fiction: the rise of A.I. interrogation in the dawn of autonomous robots and the need for an additional protocol to the U.N. convention against torture. *Minnesota Law Review*. Available at: http://www.minnesotalawreview.org/wp-content/uploads/2017/06/McAllister.pdf

McGuire MR (2007) *Hypercrime*. London: Cavendish.

McGuire MR, and Holt TJ, (eds) (2017) *The Routledge Handbook of Technology, Crime and Justice*. Abingdon: Routledge.

Markoff J (2016) As artificial intelligence evolves, so does its criminal potential. *The New York Times*, 23 October. Available at: https://www.nytimes.com/2016/10/24/technology/artificial-intelligence-evolves-with-its-criminal-potential.html

Michel AH (2019) *Eyes in the Sky*. Boston, MA: HMH Books.

Miranda EM, McBurney P and Howard MJW (2016) Learning unfair trading: a market manipulation analysis from the reinforcement learning perspective. In: *Proceedings of the 2016 IEEE conference on evolving and adaptive intelligent systems, EAIS 2016*, pp. 103–109. Institute of Electrical and Electronics Engineers Inc. DOI: 10.1109/EAIS.2016.7502499.

Mirsky Y, Mahler T, Shelef I et al. (2019) CT-GAN: malicious tampering of 3D medical imagery using deep learning. Available at: https://arxiv.org/abs/1901.03597

MIT Technology Review (2019) A new camera can photograph you from 45 kilometres away. *MIT Technology Review*, 3 May. Available at: https://www.technologyreview.com/s/613457/a-new-camera-can-photograph-you-from-45-kilometers-away/

Molnar P (2019) Technology on the margins: AI and global migration management from a human rights perspective. *Cambridge International Law Journal* 8(2): 305–330. DOI: 10.4337/cilj.2019.02.07.

Mozur P (2018) Inside China's dystopian dreams: A.I., shame and lots of cameras. *The New York Times*, 8 July. Available at: https://www.nytimes.com/2018/07/08/business/china-surveillance-technology.html

Mozur P (2019) One month, 500,000 face scans: how China is using A.I.to profile a minority. *The New York Times*, 14 April. Available at: https://www.nytimes.com/2019/04/14/technology/china-surveillance-artificial-intelligence-racial-profiling.html

Mozur P, Kessel J and Chan M (2019) Made in China, exported to the world: the surveillance state. *The New York Times*, 24April. Available at: https://www.nytimes.com/2019/04/24/technology/ecuador-surveillance-cameras-police-government.html

Munksgaard R, Demant J and Branwen G (2016) A replication and methodological critique of the study 'Evaluating drug trafficking on the Tor Network'. *International Journal of Drug Policy* 35: 92–96.

Nguyen A, Yosinski J and Clune J (2015) Deep neural networks are easily fooled: high confidence predictions for unrecognizable images. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, Boston, MA, 7–12 June, pp. 427–436. New York: IEEE.

Oht H, Dekel T, Kim C et al. (2019) Speech2Face: learning the face behind a voice. Available at: https://arxiv.org/abs/1905.09773

Paoli GP, Aldridge J, Ryan N et al. (2017) *Behind the Curtain: The Illicit Trade of Firearms, Explosives and Ammunition on the Dark Web*. Santa Monica, CA: RAND.

Parsons C, Molnar A, Dalek J et al. (2019) The predator in your pocket: a multidisciplinary assessment of the Stalkerware application industry. *The Citizen Lab*, June. Available at: https://citizenlab.ca/docs/stalkerware-holistic.pdf

Patterson G and Greene D (2018) The trouble with trusting AI to interpret police body-cam video. *IEEE Spectrum: Technology, Engineering, and Science News*, 21 November. Available at: https://spectrum.ieee.org/computing/software/the-trouble-with-trusting-ai-to-interpret-police-bodycam-video

Perry N (2018) How Axon is accelerating tech advances in policing. *Policeone*, 22 June. Available at: https://www.policeone.com/police-products/body-cameras/articles/476840006-How-Axon-is-accelerating-tech-advances-in-policing/

Phillips PJ (2018) Face recognition accuracy of forensic examiners, superrecognizers, and face recognition algorithms. *Proceedings of the National Academy of Sciences of the United States of America* 115: 6171–6176.

Piper K (2018) The case for taking AI seriously as a threat to humanity. *Vox*. Available at: https://www.vox.com/future-perfect/2018/12/21/18126576/ai-artificial-intelligence-machine-learning-safety-alignment

Powell A, Stratton G and Cameron R (2018) *Digital Criminology*. London: Routledge.

Radford A, Wu J, Amodei D et al. (2019) Better language models and their implications. In: *Openai Blog*. Available at: https://blog.openai.com/better-language-models/

Raponi S, Ali I and Oligeri G (2020) Sound of guns: digital forensics of gun audio samples meets artificial intelligence. Available at: http://arxiv.org/abs/2004.07948

Rhue L (2018) *Racial Influence on Automated Perceptions of Emotions*. Rochester, NY: Social Science Research Network.

Russell SJ and Norvig P (2009) *Artificial Intelligence*. Harlow: Pearson

Sadowski J and Bendor R (2019) Selling smartness: corporate narratives and the smart city as a sociotechnical imaginary. *Science, Technology, & Human Values* 44(3): 540–563.

Sadowski J (2019) The captured city. *Real Life*, 12 November. Available at: https://reallifemag.com/the-captured-city/

Satter R (2019) Experts: spy used AI-generated face to connect with targets. *AP NEWS*, 13 June. Available at: https://apnews.com/bc2f19097a4c4fffaa00de6770b8a60d (accessed 13 June 2019).

Sausdal D (2018) Everyday deficiencies of police surveillance: a quotidian approach to surveillance studies. *Policing and Society*. Epub ahead of print 13 December. DOI: 10.1080/10439463.2018.1557659.

Scharre P (2019) Killer apps: the real danger of an AI arms race. *Foreign Affairs*. Available at: https://www.foreignaffairs.com/articles/2019-04-16/killer-apps

Scharre P and Horowitz MC (2018) *Artificial Intelligence*. Washington, DC: Center for a New American Security.

Schneier B (2008) Inside the twisted mind of the security professional. *Wired*, 20 March. Available at: https://www.wired.com/2008/03/securitymatters-0320/

Schneier B (2019) AI has made video surveillance automated and terrifying. *Vice*, 14 June. Available at: https://www.vice.com/en_in/article/bj93z5/ai-has-made-video-surveillance-automated-and-terrifying

Schwartz O (2019) Don’t look now: why you should be worried about machines reading your emotions. *The Guardian*, 6 March. Available at: https://www.theguardian.com/technology/2019/mar/06/facial-recognition-software-emotional-science

Seymour J and Tully P (2016) Weaponizing data science for social engineering: automated E2E spear phishing on Twitter. Available at: https://www.blackhat.com/docs/us-16/materials/us-16-Seymour-Tully-Weaponizing-Data-Science-For-Social-Engineering-Automated-E2E-Spear-Phishing-On-Twitter-wp.pdf

Sharkey N, Goodman M and Ros N (2010) The coming robot crime wave. *IEEE Computer Magazine* 43: 116–115.

Shevlane T and Dafoe A (2020) The offense-defense balance of scientific knowledge: does publishing AI research reduce misuse? In: *Proceedings of the AAAI/ACM conference on AI, ethics, and society*, pp. 173–179. AIES ’20. New York: ACM.

Shumailov I, Simon L, Yan J et al. (2019) Hearing your touch: a new acoustic side channel on smartphones. Available at: https://arxiv.org/abs/1903.11137

Singh A, Patil D, Reddy GM et al. (2017) Disguised Face Identification (DFI) with facial keypoints using spatial fusion convolutional network. Available at: https://arxiv.org/abs/1708.09317

Smith CS (2018) Alexa and Siri can hear this hidden command. You can’t. *The New York Times*, 10 May. Available at: https://www.nytimes.com/2018/05/10/technology/alexa-siri-hidden-command-audio-attacks.html

Smith GJD, Bennett Moses L and Chan J (2017) The challenges of doing criminology in the Big Data era: towards a digital and data-driven approach. *British Journal Criminology* 57: 259–274.

Snow J (2018) Amazon’s face recognition falsely matched 28 members of congress with mugshots. *American Civil Liberties Union*, 26 July. Available at: https://www.aclu.org/blog/privacy-technology/surveillance-technologies/amazons-face-recognition-falsely-matched-28

Solaiman I, Brundage I, Clark J et al. (2019) Release strategies and the social impacts of language models. Available at: https://arxiv.org/abs/1908.09203

Spera C, Wittes B, Poplin C et al. (2016) Sextortion: cybersecurity, teenagers, and remote sexual assault. *Brookings*, 11 May. Available at: https://www.brookings.edu/research/sextortion-cybersecurity-teenagers-and-remote-sexual-assault/

Stanley J (2019) *The Dawn of Robot Surveillance*. New York: American Civil Liberties Union.

Stark L (2019) Facial recognition is the plutonium of AI. *XRDS* 25(3): 50–55.

Steinmetz K and Nobles MR (2017) *Technocrime and Criminological Theory*. New York: Routledge.

Sugawara T, Genkin D, Cyr B et al. (2019) Light commands: laser-based audio injection attacks on voice-controllable systems. Available at: https://lightcommands.com/20191104-Light-Commands.pdf

Sullivan G (2014) 5 scary things about the ‘Blackshades’ RAT. *Washington Post*, 20 May. Available at: https://www.washingtonpost.com/news/morning-mix/wp/2014/05/20/5-scary-things-about-blackshades-malware/

Sutter JD (2011) Amazon seller lists book at $23,698,655.93–plus shipping. *CNN*, 25 April. Available at: http://edition.cnn.com/2011/TECH/web/04/25/amazon.price.algorithm/index.html

Tencent Keen Security Lab (2019) *Experimental Security Research of Tesla Autopilot*. Tencent. Available at: https://keenlab.tencent.com/en/whitepapers/Experimental_Security_Research_of_Tesla_Autopilot.pdf

Thys SVAN, Ranst W and Goedemè T (2019) Fooling automated surveillance cameras: adversarial patches to attack person detection. Available at: https://arxiv.org/abs/1904.08653

Topol SA (2016) Killer Robots are coming and these people are trying to stop them. *Buzzfeed*, 26 August. Available at: https://www.buzzfeed.com/sarahatopol/how-to-save-mankind-from-the-new-breed-of-killer-robots

Trask A (2017) Safe crime prediction: homomorphic encryption and deep learning for more effective, less intrusive digital surveillance. Available at: https://iamtrask.github.io/2017/06/05/homomorphic-surveillance/ (accessed 8 June 2017).

Turing AM (1950) Computing Machinery and Intelligence. *Mind: A Quarterly Review* 59: 433–460.

Turner J (2018) *Robot Rules: Regulating Artificial Intelligence*. New York: Springer Berlin Heidelberg.

Vestby A and Vestby J (2019) Machine learning and the police: asking the right questions. *Policing: A Journal of Policy and Practice*. Epub ahead of print 14 June. DOI: 10.1093/police/paz035.

Vincent J (2018) These faces show how far AI image generation has advanced in just four years. *The Verge*, 17 December. Available at: https://www.theverge.com/2018/12/17/18144356/ai-image-generation-fake-faces-people-nvidia-generative-adversarial-networks-gans

Von der Burchard H (2018) Belgian socialist party circulates 'deep fake' Donald Trump video. *POLITICO*, 21 May. Available at: https://www.politico.eu/article/spa-donald-trump-belgium-paris-climate-agreement-belgian-socialist-party-circulates-deep-fake-trump-video/

Wall T and Linnemann T (2014) Staring down the state: police power, visual economies, and the 'war on cameras'. *Crime, Media, Culture* 10(2): 133–149.

Warzel C (2019) A major police body cam company just banned facial recognition. *The New York Times*, 27 June. Available at: https://www.nytimes.com/2019/06/27/opinion/police-cam-facial-recognition.html

Williams ML and Burnap P (2016) Cyberhate on social media in the aftermath of woolwich: a case study in computational criminology and Big Data. *British Journal of Criminology* 56: 211–238.

Williams R (2017) Lords select committee, artificial intelligence committee, written evidence (AIC0206). Available at: http://data.parliament.uk/writtenevidence/committeeevidence.svc/evidencedocument/artificial-intelligence-committee/artificial-intelligence/written/70496.html#_ftn13

Winston A (2018) Palantir has secretly been using New Orleans to test its predictive policing technology. *The Verge*, 27 February. Available at: https://www.theverge.com/2018/2/27/17054740/palantir-predictive-policing-tool-new-orleans-nopd

Xiao C, Deng R, Li B et al. (2018) Characterizing adversarial examples based on spatial consistency information for semantic segmentation. Available at: http://arxiv.org/abs/1810.05162

Xu K, Zhang G, Liu S et al. (2019) Evading real-time person detectors by adversarial T-shirt. Available at: https://arxiv.org/abs/1910.11099

Yeung K (2017) 'Hypernudge': Big Data as a mode of regulation by design. *Information, Communication & Society* 20(1): 118–136.

Yudkowsky E (2008) Artificial intelligence as a positive and negative factor in global risk. In: Bostrom N and Cirkovic MM (eds) *Global Catastrophic Risks*. Oxford: Oxford University Press, pp. 308–345.

Zardiashvili L, Bieger J, Dechesne F et al. (2019) AI ethics for law enforcement. *Delphi* 4(7). Available at: https://delphi.lexxion.eu/article/DELPHI/2019/4/7

Zedner L (2007) Pre-crime and post-criminology? *Theoretical Criminology* 11: 261–281.

Zellers R, Holtzman A, Rashkin H et al. (2019) Defending against neural fake news. Available at: http://arxiv.org/abs/1905.12616

Zuboff S (2019) *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. New York: Public Affairs.

Author biographies

Keith Hayward is Professor of Criminology at the Faculty of Law, University of Copenhagen, Denmark and a Visiting Professor at the School of Justice, Queensland University of Technology, Australia. His research interests include criminological theory, space and crime, digital criminology, and terrorism and extremism.

Matthijs M. Maas is a PhD Fellow at the Centre for International Law and Governance (Faculty of Law, University of Copenhagen), and a Research Affiliate with the Center for the Governance of AI at the Future of Humanity Institute (University of Oxford). His research focuses on global governance strategies for artificial intelligence, and the disruptive effects of emerging technologies for law and power.