www.kic.re.kr

Artificial Intelligence in the Context of Crime and Criminal Justice

A REPORT FOR THE KOREAN INSTITUTE OF CRIMINOLOGY

KIC | Korean Institute of Criminology

ICCC International Centre for Comparative Criminology

KIC Korean Institute of Criminology

ICCC International Centre for Comparative Criminology

Artificial Intelligence in the Context of Crime and Criminal Justice

Benoît Dupont, Yuan Stevens, Hannes Westermann, Michael Joyce

Canada Research Chair in Cybersecurity
International Centre for Comparative Criminology – Université de Montréal

A Report for the Korean Institute of Criminology

December, 2018

International Centre for Comparative Criminology

URL:https://www.cicc-iccc.org/en

Université de Montréal

Canada Research Chair in Cybersecurity
Université de Montréal, Montreal, Canada

URL:https://www.umontreal.ca/

Dr. Benoît Dupont
Email address: benoit.dupont@umontreal.ca

Acknowledgements

We thank the Cybercrime Laboratory at the University of Montreal, the International Centre for Comparative Criminology and Director Dr. Carlo Morselli for their support.
We especially thank Dr. Jea Hyen Soung and staff of the International Cooperation Center of the Korean Institute of Criminology for their assistance in producing this report.

CONTENTS PAGE

Foreword

It is my great pleasure to present Artificial Intelligence in the Context of Crime and Criminal Justice, the second joint-research project conducted by the Korean Institute of Criminology and the Université de Montréal. On behalf of the Korean Institute of Criminology, I would like to gratefully acknowledge the significant efforts put in by Professor Benoît Dupont and the researchers at the International Centre for Comparative Criminology (ICCC), Université de Montréal.

Unmanned vehicles, surgical robots, industrial robots and other artificial Intelligence (AI) entities are in common use across the globe. Such use may be personal, medical, military, commercial, or industrial. This research examines the current and future use of AI technologies and their potential impacts on major stakeholders in the criminal justice system. In this regard, this joint research of KIC and Université de Montréal is of great importance in helping to lead the way in applying AI to address criminal justice needs, such as identifying individuals and their actions in videos relating to criminal activity or public safety, DNA analysis, gunshot detection, and crime forecasting. I have no doubt that this publication will provide the valuable step in helping scholars and professionals around the world interested in AI for criminal justice purposes. It is my hope that this publication receives the widespread readership that it deserves, and that criminological partnership between Korea and Canada continues to thrive.

Once again, I would like to express my appreciation for the hard work of all the researchers and members in KIC and Université de Montréal who made this publication possible.

Korean Institute of Criminology
President In Sup Han

Preface

Perhaps no other technology currently under development invokes as much hope, hype and fear as Artificial Intelligence (AI). Governments and companies are pouring billions of dollars into research labs and startups that hope to disrupt entire sectors of the economy and improve humans' cognitive capacities. No area of human activity is left untouched by the advent of AI, as the uncontested Go champion Lee Sedol found out in 2016 when he lost a five-game match against AlphaGo, a program created by Google's subsidiary DeepMind that was awarded the highest rank of grandmaster by South Korea's Go Association following its 4-1 win.

Although current uses of AI have produced their most impressive results in the fields of language translation, image classification, and pattern recognition more generally, governments are increasingly exploring a broad range of opportunities to deploy AI in settings where it is expected that its predictive capacities will improve the quality of service delivery and the effectiveness of state interventions. One domain of application that has attracted a lot of media attention so far, but still offers very limited scientific research, is criminal justice, which is defined in this report as the complex web of interactions and institutions that bring together offenders, police officers, court officials and corrections professionals.

We believe one of the reasons for this intense interest resides in the proliferation of science fiction dystopias built around

intelligent machines that can predict individual crimes before they occur and curtail the individual freedoms of citizens to maintain law and order at all costs. Such terrifying outcomes are unlikely to materialize, but it does not mean that criminal justice institutions will smoothly adopt AI technologies or that these new tools will yield all the benefits that their designers and promoters are advertising. Many disappointments and failures, some of which will generate unpredictable and unfair outcomes, can be expected. In other words, the future of uses of AI in criminal justice might very well prove more reminiscent of Franz Kafka than of George Orwell or Philip K. Dick.

Hence, this report attempts to map the more mundane reality that will most likely emerge and the multiple challenges that criminal justice institutions will have to address as a result of their experimentations with AI. After having provided a brief overview of the different types of machine learning technologies available and their expected impact on society at large, we examine actual and potential uses of AI by the four main categories of actors and stakeholders that interact in the criminal justice system: offenders, law enforcers, judges and corrections officers. Each chapter outlines the known uses of AI by each group, potential applications that have not yet been implemented but that can be expected in the near future, and the ethical or operational barriers these deployments will encounter, as well as their estimated impacts on the delivery of justice. It is always hazardous to make predictions about the future, so we refrain from science fiction scenarios that make for good entertainment but often fail to imagine the duller reality of criminal justice bureaucracies. In the final chapter, we summarize what we believe are the four main challenges

that should be addressed by policy makers, practitioners and researchers thinking about deploying AI systems in criminal justice settings: these challenges are of an ethical, technical, administrative and cultural nature. Although ethical dilemmas and the biases they are trying to avoid occupy most of the conversation on AI, the three other interconnected challenges also all deserve our attention.

AI is not the first—nor the last—technology aiming at disrupting the criminal justice system and claiming to be able to make its institutions more effective and efficient. Many of these technologies failed to deliver their expected benefits. In order to understand why such promising innovations keep on faltering, the final recommendation of this report is to encourage ethnographic studies seeking to understand how the new assemblages of humans and AI-powered machines operate in day-to-day practice, in order to move beyond the current fetishism of algorithms.

Professor Benoît Dupont
Canada Research Chair in Cybersecurity – Université de Montréal

Executive Summary

News headlines remind us every day that artificial intelligence (AI) is bound to become one of the most disruptive technologies ushered in by the Digital Revolution. World Chess and Go champions are being defeated by machines that beat their opponents with relentless effectiveness, while AIs managing the power usage of data centres generate impressive energy savings and medical algorithms seem able to detect cancerous tumours before they appear on scans. In the near future, autonomous vehicles promise to significantly reduce the number of road fatalities and universal translators to enable better communications across languages, all powered by machine learning technologies that will optimize every aspect of human activities. Billions of dollars are currently being invested by governments, venture capital firms and Internet giants such as Microsoft, Facebook, Amazon and Apple to embed AI solutions into their services and products.

The disruption will also bring its share of pain, with the most negative predicted impact being the destruction of millions of jobs. The most pessimistic studies estimate that almost half of the jobs in developed economies are at risk of automation. Physical repetitive work is obviously being singled out, but knowledge work and professional services such as law and medicine are also becoming vulnerable. In some extreme cases, AI will also be used by individual offenders and criminal groups to harm an unprecedented number of victims. In response, criminal justice organizations are already considering the use

of AI technologies to improve the effectiveness and efficiency of their procedures, and some experimental applications are currently being deployed by law enforcement agencies, courts and correctional services. This is certainly not the first wave of technological innovation to transform the delivery of justice through the ages, but the potential biases it introduces and its lack of explainbility and accountability represents a major challenge for democratic values.

This exploratory report offers an overview of the role AI is bound to play in criminal justice, relying on a broad range of examples gathered from around the world. It adopts a sequential approach that reflects how a crime unfolds, from its commission by offenders to its detection by law enforcement investigators, then its judgement by criminal courts, and finally the enforcement of a sentence by correctional services.

In order to understand the underlying technical concepts making AI such a disruptive technology for criminal justice agencies, chapter 2 seeks to explain the history and features of AI, with a particular emphasis on differences with other forms of computer programming. This chapter maps the evolution of AI from rule-based systems that were introduced in computer science as early as the 1950s, which were then replaced by Machine Learning approaches in the 1980s, which developed the capacity to automatically improve with experience. Finally, Deep Learning is now flourishing as a subset of Machine Learning and relies on a multi-layered architecture inspired from the human brain that automatically finds relevant features in an ocean of unstructured data. Deep Learning has produced dramatic improvements in field such as image classification,

speech recognition and natural language processing. Despite its unparalleled potential, Deep Learning has also a number of pitfalls such as the potential to uncover features that people would prefer to remain private, to influence people on a large scale without these people realizing they have been manipulated, to reproduce and amplify the biases and discrimination embedded in the data it uses to make predictions, and a structural opacity with regards to the reasons why it has come to a particular conclusion. These limitations of Deep Learning technology, which fuels the current hype around AI, could therefore reinforce the status quo and sustain systematic discrimination.

Chapter 3 focuses on AI as a vector of crime. The democratization of AI means that members of the public have gained access to key resources needed to use and develop their own AI tools (data, software, and hardware), which may also empower malicious actors to use AI for nefarious purposes. The risks posed by criminal AI can be organized in three categories: existing criminal threats that expand due to the automation enabled by AI, new threats that are introduced by the capacity of AI to generate data mimicking the voice or picture of a person, hybrid threats that develop due to better targeted, more effective and less attributable attacks. Among the criminal activities facilitated or enabled by AI, this report highlights social engineering attacks (phishing, vishing, and astroturfing), generative attacks relying on the creation of extremely realistic looking images, videos, or soundbites (deep fakes), and more technical cyber-attacks where AI systems are used to discover and exploit unknown software vulnerabilities. Adversarial attacks, where AI systems can be subverted or poisoned, are also discussed.

Chapter 4 examines how law enforcement agencies around the world have begun using AI-powered technologies to detect, investigate, prevent and at times even try to predict crimes. There is a long history of the use of technology in criminal investigations, but the use of AI has the power to facilitate unprecedented levels of surveillance and social control. First, AI is an attractive technology to detect crime due to its pattern recognition and object classification capabilities: AI can for example learn to identify the location where an image or a video has been shot, or to associate particular tattoos with specific gang affiliations or meanings. Face recognition technology and its live-tracking capacity also relies heavily on AI, with China making extensive use of it in its urban centres. Other crime detection use cases include body-worn cameras, speech recognition technology (for phone intercepts for example), gunshot detection systems, and DNA and digital forensic analysis at scale. AI is also being leveraged by law enforcement to try to prevent and predict crime, with products such as PredPol claiming to be able to pinpoint the location of future occurrences, thereby enabling the dispatch of a proactive and deterrent police presence. The scientific evidence to back the effectiveness of this predictive approach remains inconsistent, at best, while the risks of unfair profiling for certain vulnerable groups (visible minorities in particular) are significant.

Chapter 5 focuses on courts and corrections, showing how AI is being incorporated into judicial and carceral decision-making processes. This report identifies a few key areas such as risk assessment decisions in bail and sentencing hearings where AI technologies are strategically marketed. It provides a case study of a particular assessment tool developed in the US to limit

over-incarceration, showing how AI can also be used to neutralize the human biases that have disproportionately afflicted some minority groups. Other assessment tools, such as the COMPAS software, have been scrutinized by journalists and researchers, who have discovered that the accuracy of their predictions is controversial and suggests systemic racial biases against black defendants. Although the designers and marketers of these AI products dispute such findings, there is at the moment very scant independent evidence allowing us to make robust assessments on their accuracy—or lack thereof. The lack of transparency around the algorithms that power such tools and the difficulty to review them only compound the caution that should be exercised when considering their adoption.

The final chapter considers four main categories of challenges raised by the deployment of AI tools in criminal justice settings, because of their potential impact on individual freedoms. These challenges are not only ethical, but also address the effectiveness of AI, the complexities of its procurement, and the vagaries of its appropriation by criminal justice professionals. These four challenges are closely interconnected and amplify each other. They need to be thoroughly addressed before AI becomes routinely embedded into criminal justice procedures. The central challenge that has attracted the most attention so far is ethical: although the benefits of AI are potentially very significant, the automation of decision-making in a justice context raises a number of moral dilemmas related to fundamental principles such as fairness and equality before the law. On a more technical level, there are also lingering uncertainties on the suitability of cutting-edge Machine Learning approaches to unstable domains where generalizations have to be made from limited data collected in

a dynamic context. Humans might still retain an edge over machines to investigate and assess certain criminal risks. The acquisition of AI systems by criminal justice agencies also create ethical and performance implications of their own if they are not handled properly. The companies that develop AI solutions for this market are reluctant to provide access to the "secret sauce" of their algorithms, but transparency should be non-negotiable in a criminal justice context, where the human right stakes are so high. Finally, AI systems will not be adopted seamlessly by criminal justice professionals: as the history of previous technologies has shown, human users will always retain high levels of agency that will take various forms, from domestication to resistance, and even sabotage.

1. INTRODUCTION

There is news everyday about the awe-inspiring possibilities brought on by artificial intelligence (AI). Already, AI systems have come to exceed the skills of humans at several challenging games. In 2011, a system named 'Watson' developed by IBM, defeated the world champions of the television game show Jeopardy.[1] In 2016, an artificially intelligent computer system developed by Google known as 'DeepMind' defeated Lee Sedol, one of the world's best 'Go' Players.[2] This victory was remarkable, given that Go is an extremely complex game. It heavily relies on the intuition of the player and was therefore thought to be extremely hard to master for computers.[3] In December 2017, DeepMind reached another milestone with its AlphaZero system (an upgraded version of AlphaGo), which taught itself to play chess in less than four hours and beat the world champion chess program in a 100-game match up.[4]

---

1 Jo Best, "IBM Watson: The inside story of how the Jeopardy-winning supercomputer was born, and what it wants to do next", TechRepublic (9 September 2013), online: https://www.techrepublic.com/article/ibm-watson-the-inside-story-of-how-the-jeopardy-winning-supercomputer-was-born-and-what-it-wants-to-do-next.

2 David Silver & Demis Hassabis, Cade Metz, "In Two Moves, AlphaGo and Lee Sedol Redefined the Future", Wired (16 March 2016), online: https://www.wired.com/2016/03/two-moves-alphago-lee-sedol-redefined-future/.

3 Ibid.

4 Samuel Gibbs, "AlphaZero AI beats champion chess program after teaching itself in four hours", The Guardian (7 December 2017), online: https://www.theguardian.com/technology/2017/dec/07/alphazero-google

The increasing capacities of artificial intelligence and its seeming competence at tasks formerly restricted to the human realm raise significant questions for the impact this technology may have on crime and criminal justice. AI technology could affect not only how crimes are committed, but also how law enforcement operates and how the criminal justice system functions. Of course, these drastic changes are not restricted to the administration of justice, as all sectors of human activity will be disrupted by AI. Many experts and analysts agree: A study by economist Carl Benedikt Frey and machine learning expert Michael A. Osborne claims that 47% of the US work force is at risk of automation.[5] Especially at risk, according to this study, are workers in transportation and logistics, the service industry, office and support workers as well as some forms of manual labor. For example, The 3.5 million truck drivers in the U.S. will likely soon be replaced by self-driving trucks, if the findings of these researchers hold true.[6] Waymo, a Google initiative, already operates test vehicles able to drive autonomously on the

---

-deepmind-ai-beats-champion-program-teaching-itself-to-play-four-hours.

5 Carl Benedikt Frey & Michael A Osborne, "The future of employment: How susceptible are jobs to computerisation?" (2017) 114 Technological Forecasting and Social Change 254 at 44.

6 Dominic Rushe, "End of the road: will automation put an end to the American trucker?", The Guardian (10 October 2017), online: https://www.theguardian.com/technology/2017/oct/10/american-trucker-automation-jobs; Finn Murphy, "Truck drivers like me will soon be replaced by automation. You're next", The Guardian (17 November 2017), online: https://www.theguardian.com/commentisfree/2017/nov/17/truck-drivers-automation-tesla-elon-musk; Paul A Eisenstein, "Millions of jobs are on the line when autonomous cars take over", NBC News (5 November 2017), online: https://www.nbcnews.com/business/autos/millions-professional-drivers-will-be-replaced-self-driving-vehicles-n817356.

roads of Arizona.[7] And 78% of predictable physical work, such as welding or assembly lines, can supposedly be automated.[8] Even knowledge work or professional services such as law and medicine are supposedly at risk of being affected by AI. Tools are being developed that are able to swiftly scan through thousands of documents and select the relevant ones[9] or spot issues in contracts with an average accuracy of 94%, compared to an average accuracy of 85% of human lawyers.[10]

Elon Musk, the entrepreneur behind the electric car manufacturer Tesla and the space company SpaceX, warns about the risks of artificial intelligence unleashed on the world, even if it occurs by accident.[11] Ray Kurzweil, on the other hand, believes that AI will surpass human general intelligence by 2029 – but that this will empower humanity, rather than threaten it.[12] These

---

7 Andrew J. Hawkins, "Waymo is first to put fully self-driving cars on US roads without a safety driver", The Verge (7 November 2017), online: https://www.theverge.com/2017/11/7/16615290/waymo-self-driving-safety-driver-chandler-autonomous.

8 Michael Chui, James Manyika & Mehdi Miremadi, "Where machines could replace humans--and where they can't (yet)", McKinsey Quarterly (July 2016), online: https://www.mckinsey.com/business-functions/digital-mckinsey/our-insights/where-machines-could-replace-humans-and-where-they-cant-yet.

9 Erin Winick, "Lawyer-bots are shaking up jobs", MIT Technology Review (12 December 2017), online: https://www.technologyreview.com/s/609556/lawyer-bots-are-shaking-up-jobs/.

10 "AI vs. Lawyers", LawGeex Blog (26 February 2018), online: https://blog.lawgeex.com/ai-more-accurate-than-lawyers/.

11 Maureen Dowd, "Elon Musk's Billion-Dollar Crusade to Stop the A.I. Apocalypse", Hive - Vanity Fair (26 March 2017), online: https://www.vanityfair.com/news/2017/03/elon-musk-billion-dollar-crusade-to-stop-ai-space-x.

12 Christiana Reedy, "Kurzweil Claims That the Singularity Will

voices summon images of both utopian and dystopian futures brought about by the development of AI. Either of these scenarios would of course have a tremendous impact on the conduct of society and the role of criminal justice institutions. These eventualities are hinged on a common supposition. The accounts and articles all believe in the supernatural capability of artificial intelligence to emulate and perhaps even improve on a part of what it takes to be human.

Both industry and academia have taken notice. Many large tech companies are heavily investing in AI research. The American consulting company McKinsey estimates that the private sector invested 20-30 billion USD in artificial intelligence in 2016.[13] Google, for example, acquired DeepMind in 2014.[14] This is the company responsible for AlphaGo and AlphaZero. Most other tech giants, such as Microsoft, Facebook, Apple and Amazon make heavy use of artificial intelligence in their products as well.[15]

Such massive investments are not limited to Silicon Valley. China is also pouring billions of dollars into the development

---

Happen by 2045", Futurism (5 October 2017), online: https://futurism.com/kurzweil-claims-that-the-singularity-will-happen-by-2045.

13 Jacques Bughin et al., "Artificial Intelligence: The Next Digital Frontier?", McKinsey Global Institute (June 2017) online: https://www.mckinsey.com/~/media/McKinsey/Industries/Advanced%20Electronics/Our%20Insights/How%20artificial%20intelligence%20can%20deliver%20real%20value%20to%20companies/MGI-Artificial-Intelligence-Discussion-paper.ashx at 7.

14 "DeepMind", DeepMind (website) online: https://deepmind.com.

15 Christina Mercer & Thomas Macaulay, "How tech giants are investing in artificial intelligence", Techworld (27 November 2018), online: https://www.techworld.com/picture-gallery/data/tech-giants-investing-in-artificial-intelligence-3629737.

and deployment of AI products at scale.[16] The startup world around artificial intelligence is equally flourishing. In December 2017, AngelList, a platform for connecting startups with investors, listed almost 4000 startups in AI.[17] According to Pitchbook, a financial research company, venture capitalists invested over 10 billion USD in AI startups in 2017, almost doubling the number from 2016.[18] Element AI, a Canadian company focused on helping firms implement artificial intelligence, raised 102 Million USD.[19] Interest is no less intense in academia. In 2017, almost 20,000 papers were published on the topic of AI.[20]

However, there is also a growing voice of critics of the irrational exuberance around AI. In a popular blog post, AI expert Filip Piekniewski predicts the coming of an "AI winter", a period of significant cooling in research in artificial intelligence.[21]

---

16 Kai-Fu Lee, ed, AI Superpowers: China, Silicon Valley, and the New World Order, (New York, NY: Houghton Mifflin Harcourt, 2018).

17 Alok Aggarwal, "The Current Hype Cycle in Artificial Intelligence", Scry Analytics (20 January 2018) online: https://scryanalytics.ai/the-current-hype-cycle-in-artificial-intelligence.

18 Dana Olsen, "2017 Year in Review: The top VC rounds & investors in AI", PitchBook News & Analysis (20 December 2017), online: https://pitchbook.com/news/articles/2017-year-in-review-the-top-vc-rounds-investors-in-ai.

19 Ingrid Lunden, "Element AI, a platform for companies to build AI solutions, raises $102M", TechCrunch (November 2016), online: http://social.techcrunch.com/2017/06/14/element-ai-a-platform-for-companies-to-build-ai-solutions-raises-102m.

20 Yoav Shoham, Raymond Perrault, Erik Brynjolfsso & Jack Clark, "Artificial Intelligence Index: 2017 Annual Report", AI Index (November 2017) online: http://cdn.aiindex.org/2017-report.pdf at 9.

21 Filip Piekniewski, "AI winter is well on its way", Piekniewski's Blog (28 May 2018), online:

Scientist Gary Marcus lists 10 challenges with deep learning in a paper that we will examine in greater detail in the last chapter of this report. According to him, deep learning (one of the sub-fields of AI) far exceeds human capacity in certain tasks, such as classifying input. However, other tasks, such as understanding language, are out of the scope for the current methods. Further, he points to the problem of deep learning algorithm being unable to respond well to stimuli outside of the data used to train the algorithm.[22]

Several studies have shown that modern artificial intelligence can fail in ways that might seem completely unintuitive to humans. Adding a certain pattern of noise over a picture, which does not in any way change the way the picture appears to a human, can make an AI classify a dog as an ostrich[23], or a stop sign as a yield sign.[24] Often, slight changes in the images an artificial intelligence is shown, such as adding an elephant to a picture,[25] will cause the recognition of other objects to fail in completely unexpected ways. Computer scientist Melanie Mitchell believes this is due to the "barrier of meaning". Humans have general, common-sense knowledge for understanding the

---

https://blog.piekniewski.info/2018/05/28/ai-winter-is-well-on-its-way.

22 Gary Marcus, "Deep Learning: A Critical Appraisal" (2018) arXiv Working Paper, arXiv:1801.00631 [cs.AI], online: https://arxiv.org/abs/1801.00631 at 15-16.

23 Christian Szegedy, et al., "Intriguing properties of neural networks" (2013) arXiv Working Paper, arXiv:1312.6199 [cs.CV], online: http://arxiv.org/abs/1312.6199.

24 Kevin Eykholt et al, "Robust Physical-World Attacks on Deep Learning Models" (2017) arXiv Working Paper, arXiv:1707.08945 [cs.CR] online: http://arxiv.org/abs/1707.08945.

25 Amir Rosenfeld, Richard Zemel & John K Tsotsos, "The Elephant in the Room" (2018) arXiv Working Paper, arXiv:1808.03305 [cs.CV] online: http://arxiv.org/abs/1808.03305.

world, which allows us to generalize and recognize new situations. Artificial intelligence lacks this common sense. According to her, this means that the current approach might not give us artificial intelligence that is trustworthy in its decision-making, and that we have to take a step back first before we rely on it.[26]

Faced with these two radically different viewpoints, it can be hard to determine what AI is and how it will affect the world. Is it a revolution that will make entire classes of work obsolete, cause mass unemployment and eventually surpass humans in cognitive ability? Or is it, as some of the critics claim, merely a statistical system that is able to emulate humans in some narrow tasks but that fails when exposed to the complexity of the world?

We explore some answers to these questions in this report, with a particular emphasis on criminal justice applications. The use of human-made technological tools to enact our notions of (retributive, punitive, or restorative) criminal justice dates back to as long as humans have existed: just think of all the technologies used for investigating wrongful behavior and for punishment throughout history. As this report demonstrates, AI has ushered in a new era in the delivery of criminal justice around the world marked by automated empirical analysis based on large datasets, which can be used to nudge humans or potentially make decisions for us altogether.

In this exploratory report, we offer an overview of the role of

---

[26] Amir Rosenfeld, Richard Zemel & John K. Tsotsos, "The Elephant in the Room" (2018) arXiv Working Paper, arXiv:1808.03305 [cs.CV] online: http://arxiv.org/abs/1808.03305.

AI in criminal justice relying on numerous examples spanning the globe. We adopt a chronological approach that traces how a crime unfolds, including (i) its commitment, (ii) its detection and finally (iii) the response to it by criminal courts and correctional services. First, we focus on the possibility for malicious actors to employ AI to commit reprehensible acts, though this has yet to be seen. Second, we assess the use of AI by law enforcement, including the new ability of police forces to detect and predict crime. Third, we examine the relationship between AI and criminal proceedings to show how AI is being deployed to assess the various risks associated to offenders at the pre-trial and post-conviction stages. Finally, we conclude with analysis of the four overarching categories of challenges posed by AI in the context of criminal justice: ethics, effectiveness, procurement, and appropriation. We urge caution to all entities seeking to implement AI in their criminal justice systems: these interrelated categories of issues must be explicitly and thoroughly addressed in order for AI systems to iteratively, fairly and transparently be a part of criminal justice decisions.

2. DEFINING ARTIFICIAL INTELLIGENCE: WHAT IS ARTIFICIAL INTELLIGENCE AND WHY DOES IT MATTER FOR CRIMINAL JUSTICE?

In order to understand how AI is and can be used in criminal justice, it is critical to gain an understanding in what exactly artificial intelligence is. First, we will explain the main concepts related to AI, and how it is different compared to other forms of computer programming. Then, we will delve into what artificial intelligence is not. We will then give an overview of the various ways in which AI has, and is likely to, disrupt the sectors it enters. Finally, we will explain the different ways artificial intelligence might be used in the world of criminal justice. This chapter serves as a useful introduction to understand the capabilities of artificial intelligence and which ones can transfer to the delivery of criminal justice.

2.1. What is AI?

2.1.1. Artificial intelligence

The American Association for the Advancement of Artificial Intelligence describes artificial intelligence as "the scientific understanding of the mechanisms underlying thought and intelligent behavior and their embodiment in machines."[27] This casts a very broad net, since it includes any intelligent seeming behavior a machine can perform. A simple chat interface that

[27] Robert Atkinson, "'It's Going to Kill Us!' and Other Myths About the Future of Artificial Intelligence" (2016) Information Technology 50 at 3.

asks you questions but only allows you to answer yes or no, for example, exhibits signs of intelligence. Another example is an electric drier that stops when it senses that clothes are dry.[28]

AI is generally split into two categories: *General Artificial Intelligence* and *Narrow Artificial Intelligence*. General Artificial Intelligence (or strong AI) is thought to be a computer system exhibiting human or superior intelligence in all fields. It would be able to take knowledge from one field and transfer it to another.[29] A number of tests have been suggested to determine whether an AI system exhibits strong artificial intelligence. The most famous is probably the Turing test, which asks judges to determine whether they are speaking to a computer or a human over a chat interface.[30] Another test that has been suggested is the Wozniak Coffee test – can a machine go into an unknown house and make a cup of coffee?[31] General Artificial Intelligence could have tremendous effects on humanity and potentially replace all human labor. However, it is likely a long way off. Experts disagree on whether it will happen in our lifetimes, and if the current path of artificial intelligence will get us there.[32]

---

28 Ibid.

29 The Privacy Expert's Guide to Artificial Intelligence and Machine Learning (Future of Privacy forum, 2018) at 5; "What is AGI?", (11 August 2013), online: Machine Intelligence Research Institute https://intelligence.org/2013/08/11/what-is-agi/.

30 Ben Goertzel, Matt Iklé & Jared Wigmore, "The Architecture of Human-Like General Intelligence" in Pei Wang & Ben Goertzel, eds, Theoretical Foundations of Artificial General Intelligence (Paris: Atlantis Press, 2012) at 140.

31 Ibid at 141.

32 Peter Voss, "From Narrow to General AI", Intuition Machine 3 October 2017), online: https://medium.com/intuitionmachine/from-narrow-to-general-ai-e21b568155b9; James Vincent, "This is when AI's top researchers think

Even though AlphaGo is amazing at playing Go, it still cannot transfer this superior knowledge mastery to another domain (or even make a cup of coffee).[33]

All human achievements in artificial intelligence so far therefore fall into the category of Narrow AI. This is artificial intelligence that deals with solving a predefined problem, such as playing a board-game, identifying images or driving a car.[34] Narrow AI is very useful in its own right, and can have large effects on society by making workers more efficient and automating tasks. However, it is not concerned with a fully conscious, human-level intelligence.

2.2. A history of approaches to Narrow AI

This section will elaborate on which methods have been used in order to create intelligent systems. They can be separated into 3 eras, or approaches: Rule-based methods, Machine Learning and Deep Learning.

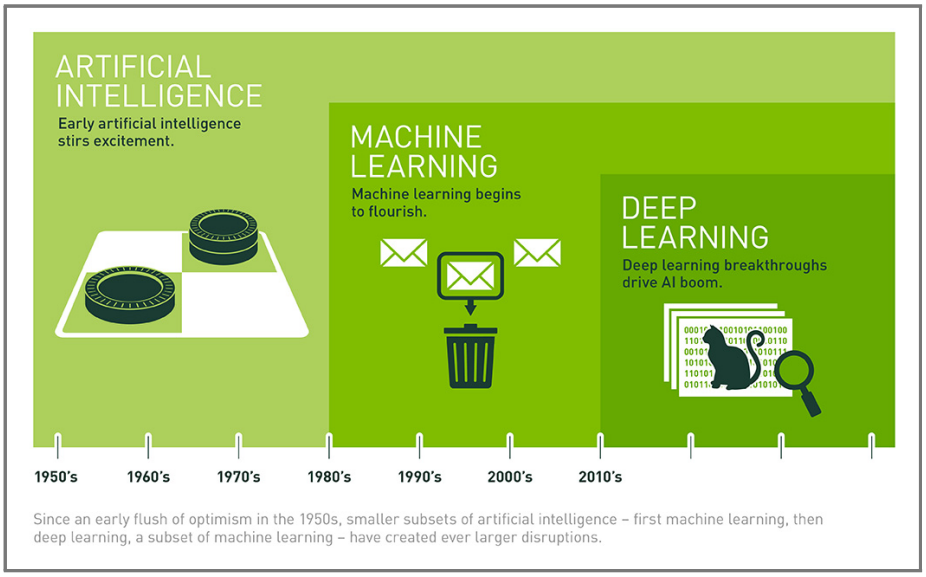
---

artificial general intelligence will be achieved", The Verge (27 November 2018), online: https://www.theverge.com/2018/11/27/18114362/ai-artificial-general-intelligence-when-achieved-martin-ford-book.

33 Atkinson, supra note 27 at 7.

34 Ibid.

Figure 1 - A timeline of artificial intelligence approaches[35]



2.2.1. Rule-based systems

To construct expert systems, a programmer will precisely encode knowledge of the problem he or she wants to solve into the computer. This results in an expert system, able to provide expert assistance in a limited domain automatically.[36] While Expert Systems can lead to impressive results in a number of areas, they suffer from a number of difficulties. First, as their name implies, they depend on the domain knowledge of an expert to obtain their knowledge. For example, a programmer building a chess engine would encode their own knowledge of chess into the computer. However, this could never surpass the level of chess

[35] Michael Copel, "The Difference Between AI, Machine Learning, and Deep Learning?", The Official NVIDIA Blog (29 July 2016), online: https://blogs.nvidia.com/blog/2016/07/29/whats-difference-artificial-intelligence-machine-learning-deep-learning-ai/.

[36] Bruce Buchanan, "A (Very) Brief History of Artificial Intelligence" 26 AI Magazine (2005).

knowledge of the person implementing the system, since it is merely reimplementing the knowledge of the creator.

Further, much of the knowledge we use in everyday life is implicit, and thus very hard to explicitly transfer into code. For example, it would be very hard for a human to formalize all the knowledge and muscle movements that go into riding a bike, or all the thoughts that go into determining whether an animal is a cat or a dog. Trying to formalize these instincts is likely to take a lot of time and is unlikely to capture the full complexity of the task performed by the brain.

Due to this difficulty of fully encoding knowledge into an algorithm, expert systems also have a problem generalizing to new information. As long as an issue falls exactly into the same class as the creator of the Expert System intended, the result will be good. However, as soon as the input falls outside of the specified parameters, the system will be unable to determine an outcome. A simple example will illustrate this point. A simple expert system could be to ask whether an animal has whiskers to determine whether it is a cat or a dog. If it has whiskers, it is a cat, otherwise a dog. This system works for many cases, but immediately fails if a cat has lost its whiskers. Even in this simple situation, it does not generalize well.

2.2.2. Machine Learning

Machine Learning (ML) works in another way. Instead of trying to encode his knowledge into the system, the programmer will show the algorithm a number of examples and a label for the data. The machine will then *itself* figure out what these examples have in common. The more examples it is shown, the better the

algorithm will become – it is thus capable or improving itself. Hence, a popular definition for ML is:

"The field [that] is concerned with the question of how to construct computer programs that automatically improve with experience."[37]

For our cat-or-dog example, this would work the following way. The programmer would select a large number of images of dogs, and a large number of images of cats. He or she would then show these to the computer and tell it which animal a picture represents. By looking at all of the data and identifying patterns, the computer then would build a model of what makes an animal a dog or a cat. After this, the computer is presented with an image that it has not previously seen and is then able to use the model to predict the species.

As will be described later in more detail, traditional ML algorithms typically require a human to decide which features of the real world it should look at.[38] This requires a lot of time and domain expertise and makes it very hard to use traditional ML for the analysis of unstructured data, such as speech and images.[39] There are hundreds of different algorithms to perform machine learning. Some of the differences will be explained below. Examples of algorithms are Linear Regression, Random Forests and Support Vector Machines. However, one set of algorithms, known as

---

37 Tom Mitchell, Machine Learning, (New York: McGraw-Hill Education, 1997).

38 Yann LeCun, Yoshua Bengio & Geoffrey Hinton, "Deep learning" (2015) 521:7553 Nature 436 at 1.

39 Ibid.

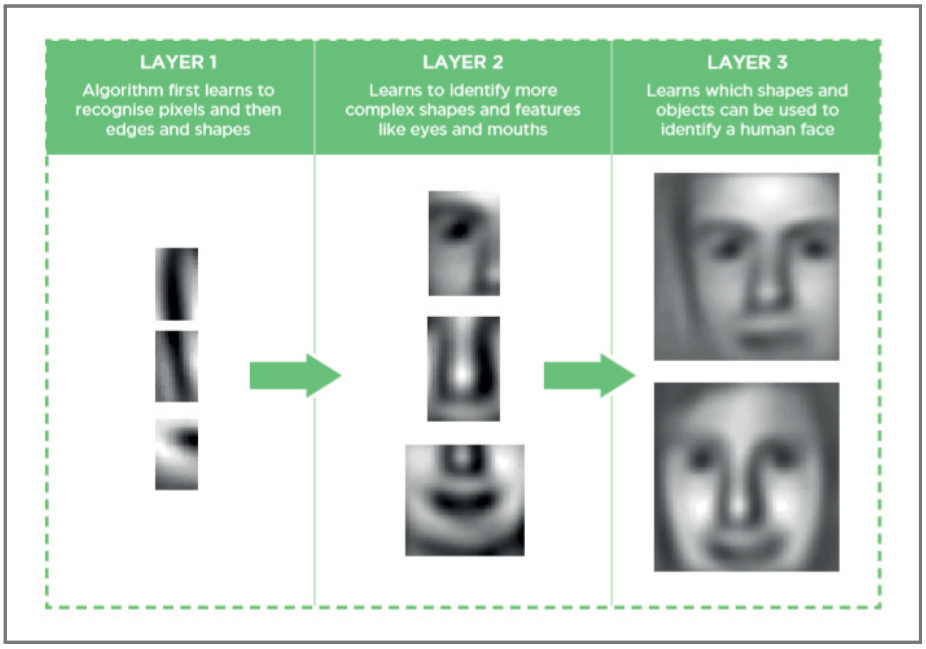
Artificial Neural Networks, have recently moved into the spotlight as maybe the most powerful yet.

2.2.3. Deep Learning

Deep Learning (DL) is thus a subset of machine learning algorithms. Typically, artificial networks that have more than two "hidden layers" are described as Deep Learning systems. The difference between traditional ML and DL is that the latter is structured into hierarchical layers. Instead of manually extracting features from the data, the engineer can feed the data directly to the Deep Learning algorithm, which will automatically find the relevant features. Each layer moves to a higher level of abstraction.[40] For cats and dogs, for example, the first layer could recognize basic visual patterns, the second could focus on whiskers, tails and paws, while the third would detect the higher-level features of dogs versus cats. Today, researchers construct models with tens of these layers. This means that they are able to learn much more sophisticated models of reality compared to regular ML.

[40] Ibid.

Figure 2 - Hierarchical representations in Deep Neural Networks[41]



Deep Learning has, in the recent years, produced dramatic improvements in the state of the art of several fields of artificial intelligence, such as image classification, speech recognition and understanding natural language.[42]

The three main reasons for the great leaps achieved by deep learning are as follow:

- *Large collections of data*: deep learning systems require a huge amount of data to be trained, which has become

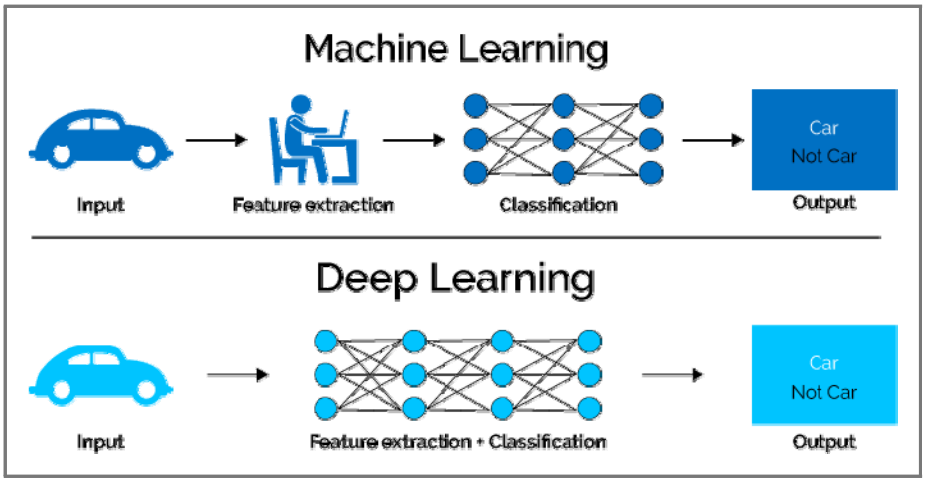
41 Sambit Mahapatra, "Why Deep Learning over Traditional Machine Learning?", Towards Data Science (21 March 2018), online: https://towardsdatascience.com/why-deep-learning-is-needed-over-traditional-machine-learning-1b6a99177063.

42 LeCun, Bengio & Hinton, supra note 38 at 1.

feasible with technical improvements in storage capacity and the creation of databases containing billions of data points;

- *More powerful technology*: training a deep neural net requires a huge amount of computation. Some of the models take days or even weeks to train. However, it turns out this computation can be performed very efficiently on Computer Graphics Cards. This has made incredibly complex models trainable in reasonable times;
- *Better algorithms*: the advances in deep learning algorithms in recent years have been staggering. Researchers such as Yoshua Bengio, Yann LeCun and Geoffrey Hinton developed and refined methods that made the deep learning revolution possible.[43]

Figure 3 - The difference versus traditional machine learning and deep learning[44]



[43] Terrence Sejnowski, The Deep Learning Revolution (Cambridge, Massachusetts: The MIT Press, 2018) at 141.
[44] Mahapatra, supra note 41.

2.3. The methods of Machine Learning

There are two main ways of implementing ML: supervised and unsupervised learning.

2.3.1. Supervised Learning

Supervised learning is a form of machine learning where a correct answer is provided to the machine at the training stage. For example, an image could be provided together with a label to specify whether the image is that of a dog or a cat. Or for a real estate application, a number of properties of a house could be provided, together with the price of the house. The algorithm would ultimately try to predict this label with the properties available to it.

All machine learning algorithms follow a similar process:

- *Data*: The programmer has to provide the algorithm with a dataset. This could be, for example, a set of one million house listings and their price. The price, in this example, would be the target that the algorithm attempts to predict. The more data, the better the algorithm can become. In fact, using more data with a "stupid" algorithm will usually beat a better algorithm with less data.[45] A big advantage of the large firms in machine learning is the amount of data that they hold. Google, for example, holds and uses enormous data-sets in training their models.[46] For

[45] Pedro Domingos, "A few useful things to know about machine learning" (2012) 55:10 Communications of the ACM 78 at 6–7.

[46] Tom Simonite, "AI and 'Enormous Data' Could Make Tech Giants Like Google Harder to Topple", Wired (13 July 2017), online: https://www.wired.com/story/ai-and-enormous-data-could-make-tech-

example, Google generates data from the log-in process to its various services. To verify that they are not bots, users are asked to click on images containing certain elements, such as cars or signposts. This human-generated interpretation data can then be used to train AI systems.[47] Beyond this, large tech companies employ thousands of workers that manually go through and label pictures for self-driving cars;[48]

- *Features*: the computer, at this stage, does not know how to deal with this data. It has to be turned into a number of features, or a numerical representation of the data. This is called feature engineering. It is a complex task, requiring a lot of time and knowledge in the area.[49] For our previous example of predicting the price of a house, the relevant features could be the number of bedrooms, the total area of the house, the location and the number of windows. Color, on the other hand, might have very little impact on the price, and therefore be a bad feature. One of the big advantages of deep learning is that this kind of feature engineering does not have to be performed. The network will instead itself learn the structure of the data in several layers of abstraction, as described before. This means that this expensive and time-consuming process can often be skipped;

---

giants-harder-to-topple/.

47 "'I'm Not A Robot': Google's Anti-Robot reCAPTCHA Trains Their Robots To See", AI Business, (25 October 2017), online: https://aibusiness.com/recaptcha-trains-google-robots/.

48 Dave Lee, "Why Big Tech pays poor Kenyans to programme self-driving cars", BBC (3 November 2018), online: https://www.bbc.com/news/technology-46055595.

49 Domingos, supra note 45 at 5-6.

- *Algorithm*: the features are then fed to an algorithm. This algorithm can have different goals: Mainly *regression* or *classification*. In *Regression*, the algorithm takes in the data and tries to guess a numerical value. In our example, it could try to predict the value of a house based on a number of features. The closer the algorithm lands to the actual price of the house, the better. *Classification* tries to put the example into a class. This could be, for example, deciding whether an image is of a cat or a dog. Here, the measure of success is how many of the images the algorithm correctly classifies;
- *Evaluation*: there has to be a way to evaluate the algorithm. This is typically used by the computer internally to determine how the algorithm it currently runs is performing;
- *Training*: once the computer learns how it is currently performing, it will subtly tweak the algorithm to perform better on the next try. This process is known as training. After training, the engineer will often go back to change the features or algorithm used to further improve the performance of the model.

2.3.2. Unsupervised learning

Unsupervised learning is a class of machine learning where no labels are provided. Instead, the computer itself tries to figure out what distinguishes one piece of data from another. In our example of cats and dogs, this would be the engineer providing the algorithm with images of both cats and dogs, and the computer itself realizing that there are two different animals in the dataset, and what distinguishes them. Unsupervised

learning does not perform as well as supervised learning. However, it is an active area of research and has several advantages over supervised learning. A big advantage is that the data does not have to be labelled, making enormous troves of unstructured data accessible to analysis. Therefore, many see unsupervised learning as the approach of the future.

One important use of unsupervised learning is that of anomaly detection. Here, a network is trained to learn the structure and general appearance of a stream of data. It is then able to tell if one data point looks different from the rest. This can be used, for example, to detect problems in production lines or possible cyber fraud attempts in a large number of financial transactions.

2.3.3. Reinforcement learning

There are some other types of ML that are starting to surface. One is reinforcement learning, which sets an agent loose in an environment and tries to get it to achieve a certain goal, such as driving a car or playing a game. At first the algorithm starts out randomly. However, if by chance it achieves a winning condition, this behavior is reinforced. This is done until the algorithm reliably learns how to achieve the set goal. Reinforcement learning has been instrumental in learning to play everything from board games to computer games.

2.3.4. Generation

A relatively recent AI technique is the one associated with generative adversarial networks. It is a technique that uses artificial intelligence to not just classify, but also generate data. In technical terms, this means that one network tries to trick

another one into believing its images are real, and not fake. The two networks evolve together until they both get very good at their jobs. At this point, the generator is able to output data that almost looks real. There are also other architectures that perform well in generating data, such as Recursive Neural Networks. Generative Adversarial Networks and other types have been used to create images of faces[50], compose music and produce extremely realistic sounding speech. There are also methods for transferring one piece of generated (fake) content into another that is a true representation of reality. This can be used, for example, to turn any image into the style of a famous artist[51] or to generate realistic videos of celebrities doing or saying things they have never said or done.[52]

---

50 Tero Karras et al, "Progressive Growing of Gans for Improved Quality, Stability, and Variation" (2018) arXiv Working Paper, arXiv:1710.10196 [cs.NE], online: https://arxiv.org/abs/1710.10196 at 26.

51 Leon A Gatys, Alexander S Ecker & Matthias Bethge, "A Neural Algorithm of Artistic Style" (2015) arXiv Working Paper, arXiv:150806576 [cs, q-bio], online: http://arxiv.org/abs/1508.06576; "Deep Dream Generator", Deep Dream Generator (Website), online: https://deepdreamgenerator.com/.

52 James Vincent, "Watch Jordan Peele use AI to make Barack Obama deliver a PSA about fake news", The Verge, (17 April 2018), online: https://www.theverge.com/tldr/2018/4/17/17247334/ai-fake-news-video-barack-obama-jordan-peele-buzzfeed.

Figure 4 - Images completely generated by GANs, based on a collection of images of celebrities.[53]

2.4. Risks of artificial intelligence

While artificial intelligence has many upsides, there are also a number of potential pitfalls the use of AI might fall into. It is important that these be addressed before AI gets deployed to make sensitive decisions on behalf of governments and corporations.

2.4.1. General AI

As mentioned before, General Artificial Intelligence is still likely to be far off. However, science fiction authors and academic researchers alike have reflected on the impact such a system could have on society. The big issue is that we cannot be sure that such an artificial intelligence will share the ethics and respect for human rights that citizens aspire to in democratic

[53] Karras et al., supra note 50.

societies. If given a task for example, they might pursue this task single-mindedly and let no other consideration stand in their way. Nick Bostrom uses the example of an AI tasked with creating paperclips, which ends up consuming the entire universe to generate more paperclips.[54] A number of researchers are working in this area to determine how we might ensure that AI will remain benevolent or constrained to a box where it can do no harm.[55]

2.4.2. Narrow AI

Even the development of advanced narrow artificial intelligence gives rise to a number of risks, some of which will be described in general terms below, before we explore how they apply to criminal justice institutions in the following chapters of this report. Most of these risks derive from the fact that AI can be a very efficient tool to accomplish certain goals. However, these goals might not align with the goals and interests of the persons they affect, either because they have been poorly framed or because the AI designers have different interests altogether and experience little or no legal or market constraints.

2.4.2.1. Privacy

Privacy can be defined as the right to choose when and whom to disclose personal information to. Modern artificial intelligence tools coupled with the massive collection of private data seriously threaten this right. As Kosinski et al. showed in 2013,

---

[54] Nick Bostrom, "Ethical Issues in Advanced Artificial Intelligence" (2003) Science Fiction and Philosophy: From Time Travel to Superintelligence at 5.Machine Learningge-f7cac935a5b4>6

[55] Vincent Müller, ed, Risks of Artificial Intelligence (Florida: Chapman and Hall/CRC Press, 2015) at 5.

data that seem completely unrelated can be tied together to create an in-depth picture of the person behind these data crumbs. In this particular study, 68 Facebook likes were enough to accurately predict several personal traits, such as personality types, sexuality, skin color and political beliefs.[56]

Another highly publicized example of a similar privacy challenge posed by AI occurred in 2012. A young teenage woman received coupons for products related to pregnancy from the large retailer chain Target. However, she had not disclosed the fact that she was pregnant to Target, or even to her parents for that matter. Target used big data analysis techniques to create profiles of its customers by tying purchases recorded in their loyalty card system to actual preferences and future needs. Based on her purchasing patterns of certain skin care products and health supplements, they were able to predict the intimate details of her pregnancy.[57]

2.4.2.2. Nudging

These profiles are mostly used to target ads to people. However, they have other intended or unintended uses. By creating comprehensive profiles of people and using the knowledge they have accumulated on how particular personal features interact or correlate, companies are able to target and influence people to further their own goals, even when these goals diverge from

---

56 Michal Kosinski, David Stillwell & Thore Graepel, "Private traits and attributes are predictable from digital records of human behavior" (2013) 110:15 Proceedings of the National Academy of Sciences 5802.

57 Kashmir Hill, "How Target Figured Out A Teen Girl Was Pregnant Before Her Father Did", Forbes (16 February 2012), online: https://www.forbes.com/sites/kashmirhill/2012/02/16/how-target-figured-out-a-teen-girl-was-pregnant-before-her-father-did/.

their customers' (and society's) interests. Much has been written about filter bubbles. These are the result of large tech companies, such as Google and Facebook, optimizing their algorithms to keep people on their sites for as long as possible. This usually favors content that the person already agrees with. This, in turn, creates bubbles where the users will only be exposed to information from their own perspective, which poses a threat to the independent opinion-making process, and by extension democracy.[58]

In 2018, a company known as Cambridge Analytica came under fire for having supposedly used a massive number of Facebook user profiles to influence the 2016 United States presidential election. Cambridge Analytica is said to have used the information it had collected on the personality traits of Facebook users to micro target ads that swayed a significant number of votes or suppressed them.[59] Artificial intelligence offers completely new possibilities of analyzing and influencing the population, which obviously represents a big risk for the stability and legitimacy of democratic governments.

2.4.2.3. Discrimination

Another risk that has already manifested itself is that of discrimination. Artificial Intelligence is very good at learning

---

58 "Measuring the Filter Bubble: How Google is influencing what you click", DuckDuckGo Blog (4 December 2018), online: https://spreadprivacy.com/google-filter-bubble-study/.

59 Carole Cadwalladr, "'I made Steve Bannon's psychological warfare tool': meet the data war whistleblower", The Guardian (18 March 2018), online: http://www.theguardian.com/news/2018/mar/17/data-war-whistleblower-christopher-wylie-faceook-nix-bannon-trump.

from data. However, if this data is biased, these biases will be reproduced by the AI. For example, an automated analysis tool for job applications in the technology sector might spot a historical trend to prefer men over women and therefore value traits associated with men higher than those associated with women.[60] Word embeddings, which try to learn the semantic meaning of words, often associate certain terms with women, and others with men, reproducing gender stereotypes. For example, nurse might be associated with women while doctor is associated with men.[61] Further, facial recognition software might fail to detect people of certain ethnic groups if the data used at the learning stage was exclusively drawn from another group.[62] Bots replicating conversations between users might be taught to make racist remarks and adopt a discriminatory set of values in its interactions with other users.[63] Men might be showed ads for jobs that attract a higher salary than those shown to women, reflecting the wage gap in many occupations and reproducing inequality in professional opportunities.[64]

---

60 Jeffrey Dastin, "Amazon scraps secret AI recruiting tool that showed bias against women", Reuters (10 October 2018), online: https://www.reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKCN1MK08G.

61 Tolga Bolukbasi et al, "Man is to Computer Programmer as Woman is to Homemaker? Debiasing Word Embeddings" (2016) arXiv Working Paper, arXiv:160706520 [cs, stat], online: http://arxiv.org/abs/1607.06520.

62 "Is facial recognition technology racist?", The Week UK (27 July 2018), online: https://www.theweek.co.uk/95383/is-facial-recognition-racist.

63 James Vincent, "Twitter taught Microsoft's friendly AI chatbot to be a racist asshole in less than a day", The Verge (24 March 2016), online: https://www.theverge.com/2016/3/24/11297050/tay-microsoft-chatbot-racist.

64 Julia Carpenter, "Google's algorithm shows prestigious job ads to

2.4.2.4. Opacity

These risks are amplified because it is often impossible to explain how an artificial intelligence system comes to a conclusion. In some cases, the algorithm is protected behind the veil of intellectual property secrecy. Companies might refuse to reveal details of their algorithm, and merely deliver the result, making analysis impossible. In other cases, especially when using deep learning algorithms, the complexity of the process at play might in itself make it very hard to explain to a human. A lot of efforts are being made by researchers to create an explainable AI, which is likely to be a requirement for using artificial intelligence in society on a large scale.[65] If AI is being used to make important decisions without being explainable, and therefore reviewable, the population might be unable to understand how these decisions are being reached or inclined to systematically contest and appeal them.

men, but not to women. Here's why that should worry you.", Washington Post (6 July 2015), online: https://www.washingtonpost.com/news/the-intersect/wp/2015/07/06/googles-algorithm-shows-prestigious-job-ads-to-men-but-not-to-women-heres-why-that-should-worry-you/.

65 Mouhamadou-Lamine Diop, "Explainable AI: The data scientists' new challenge", Towards Data Science (14 June 2018), online: https://towardsdatascience.com/explainable-ai-the-data-scientists-new-challenge-f7cac935a5b4; David Gunning, "Explainable Artificial Intelligence (XAI): Technical Report", (2016) Defense Advanced Research Projects Agency DARPA-BAA-16-53; Sandra Wachter, Brent Mittelstadt & Chris Russell, "Counterfactual Explanations without Opening the Black Box: Automated Decisions and the GDPR", (2017) arXiv Working Paper, arXiv:1711.00399 [cs.AI], online: https://arxiv.org/abs/1711.00399.

2.5. Conclusion

This chapter has attempted to give an overview over the incredibly vast and flourishing field of AI. It is important to understand the technology behind artificial intelligence in order to appreciate the impact it might have on the criminal justice system. This chapter's first takeaway is that the current batch of AI is what is defined as narrow artificial intelligence. It is trained to perform a certain task. While it can be very good at this task, it is not capable of expanding this knowledge to other fields. It also does not have a general understanding of how the world works, colloquially known as common sense. While concerns for AI replacing humans as the most intelligent beings on Earth are likely to be important in the future, they are not the issues that will predominate in the current use of artificial intelligence.

Machine learning is the practice of building self-improving algorithms. They sift through data in order to identify patterns and build a model of the data. This model could be, for example, what a class of images have in common and how to distinguish them (classification) or how factors interact to arrive at a numerical conclusion, such as temperature or price (regression). To create these algorithms, it is crucial to have a large set of high-quality data. Data is therefore poised to become "the new oil".

Traditional machine learning requires feature engineering, which requires domain knowledge and time. Deep Learning, which is the class of algorithms driving the current hype, is able to automatically extract features in different layers of abstraction from data. It is thus able to create very sophisticated models of huge amounts of data, with minimal human intervention.

Deep Learning has made large advances in a number of use cases, such as self-driving cars, the analysis of data such as speech, videos and images and the playing of games.

There are a number of risks that users of AI should be aware of. It can be an incredibly powerful tool in many instances. By inferring attributes based on other data, AI can reveal attributes about people that they might want to keep secret or that they are not even aware of themselves. It can also be used to nudge people into certain directions on a massive scale, and thereby undermine democratic principles if not used properly. Since it depends on and learns from data, AI risks perpetuating biases in this data. This can reinforce the status quo and sustain systematic discrimination. This discrimination might be hard to detect since the models built by AI can be very hard to understand.

3. AI AS A VECTOR OF CRIME: THE ADVENT OF 'CRIMINAL AI'

In early 2018, a user of the internet platform Reddit posted a tool he called "FakeApp", available to download for free. This tool allowed users to use a (usually quite large) number of images to "photoshop" or edit a face of a person into another video, including realistic depictions of expressions and behavioral details. This tool was downloaded over 100,000 times.[66] The technique was used to create montages of films, such as Nicolas Cage appearing in movies he was not in for comedic effect. [67] However, a large number of the created videos were of people transposed onto pornographic videos. Users created pornographic videos featuring Hollywood stars[68] and even their friends or ex-relationships, using data obtained from social media.[69] The technology was also used to create a fake video of President

---

66 Kevin Roose, "Here Come the Fake Videos, Too", The NY Times (8 June 2018), online: https://www.nytimes.com/2018/03/04/technology/fake-videos-deepfakes.html.

67 Usersub, "Nick Cage DeepFakes Movie Compilation", online: https://www.youtube.com/watch?time_continue=25&v=BU9YAHigNx8.

68 Alec Banks, "What Are Deepfakes & Why the Future of Porn is Terrifying", Highsnobiety (20 December 2018), online: https://www.highsnobiety.com/p/what-are-deepfakes-ai-porn.

69 Samantha Cole & Emanuel Maiberg, "People Are Using AI to Create Fake Porn of Their Friends and Classmates", Motherboard (26 January 2018), online: https://motherboard.vice.com/en_us/article/ev5eba/ai-fake-porn-of-friends-deepfakes; Rebecca Ruiz, "Deepfakes are about to make revenge porn so much worse" Mashable (24 June 2018), online: https://mashable.com/article/deepfakes-revenge-porn-domestic-violence/.

Trump deriding the climate choices of Belgium. Only after clarification by the authors did the public realize that the video was fake.[70] This example clearly illustrates how powerful and disruptive AI can be in any area of human activity. As discussed in the previous chapter, artificial intelligence can be very advantageous for much of society, but there are also tremendous risks if the technology is used for malevolent purposes. This section will therefore focus on the use of artificial intelligence as a crime enabling technology.

Compared to the remaining of this report, this chapter will appear more speculative. While artificial intelligence has rapidly spread over the criminal justice landscape, its use by criminal actors remains thankfully rare—or has not reached a critical mass that would attract a sufficient level of attention. However, many researchers and observers believe this is about to change.[71] In this chapter, we examine how AI can be currently used in crime and discuss future possible uses that have been considered in the literature. We focus on the AI capabilities that are available today or are likely to become available in the near future, and not on the speculative and very distant capabilities of future technologies such as 'general' AI. We do not pretend to be able to forecast how offenders will leverage AI and will refrain from doomsday scenarios, as there is always a large gap between what is possible and what is probable. However, it is

70 Oscar Schwartz, "You thought fake news was bad? Deep fakes are where truth goes to die", The Guardian (12 November 2018), online: https://www.theguardian.com/technology/2018/nov/12/deep-fakes-fake-news-truth.

71 "The Malicious Use of Artificial Intelligence", The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation (Website) online: https://maliciousaireport.com/.

important to be aware of the new challenges the criminal justice system might face given these novel technologies.

This section does not deal with agency issues in artificial intelligence. As we already stated, our focus on 'narrow' AI, which is a tool, makes us assume that the responsible actor is the person who designed or operates the AI system. We also choose not to include uses of artificial intelligence that accidentally causes negative results such as car accidents, which might engage the criminal liability of the AI operator or owner in certain jurisdictions. We instead focus on actors that purposefully use artificial intelligence to cause harm.

3.1. The democratization of artificial intelligence

With the recent excitement and hype surrounding artificial intelligence, members of the public have gained access to a number of the key resources needed to use and develop their own artificial intelligence tools. A broad access to cutting edge technologies is generally a positive thing, as it accelerates the adoption of innovative practices. However, this may also empower a small group of malicious actors to use artificial intelligence for nefarious purposes. There are several resources needed to create an artificial intelligence tool, as we've seen in the previous chapter. AI requires data, expertise, tools and hardware. The following sections describe how these resources are becoming more easily accessible to the public.

3.1.1. Data

In the age of artificial intelligence, data is the new oil. Large

sets of high-quality data are crucial to train machine learning algorithms.[72] Tech giants such has Google and Facebook have access to massive quantities of data about their users.[73] As previously discussed, they also have the resources to employ thousands of workers to label data for them.[74] While they are typically generous with sharing their algorithms, the data is typically guarded closely, meaning that the tech giants have a significant advantage in artificial intelligence research and applications.[75] However, this does not mean that it is impossible for malevolent actors to obtain data to train their algorithms. Firstly, there many public datasets available on the internet,[76] containing, for example, anonymized medical data,[77] economic indicators[78] or millions of images tagged with words describing their content.[79] Secondly, much of the data that people upload to the large social networks is publicly accessible. As such, it

---

72 Domingos, supra note 45. upra noet 45. n (Website): 5b4>6

73 Dylan Curran, "Are you ready? This is all the data Facebook and Google have on you", The Guardian (30 March 2018), online: http://www.theguardian.com/commentisfree/2018/mar/28/all-the-data-facebook-google-has-on-you-privacy.

74 Dave Lee, "Why Big Tech pays poor Kenyans to programme self-driving cars", BBC (3 November 2018), online: https://www.bbc.com/news/technology-46055595.

75 Daniel Faggella, "The AI Advantage of the Tech Giants: Amazon, Facebook, and Google", TechEmergence (24 November 2018), online: https://www.techemergence.com/the-ai-advantage-of-the-tech-giants-amazon-facebook-and-google-etc/.

76 Stacy Stanford, "The 50 Best Public Datasets for Machine Learning", Data Driven Investor (2 October 2018), online: https://medium.com/datadriveninvestor/the-50-best-public-datasets-for-machine-learning-d80e9f030279.

77 Alistair EW Johnson et al, "MIMIC-III, a freely accessible critical care database" (2016) 3 Scientific Data.

78 "Quandl", Quandl (Website), online: https://www.quandl.com.

79 "ImageNet", Image-Net (Website) online: http://image-net.org/index.

is often possible to obtain the data simply by going to the website or using authorized tools that scrape data from these websites. For example, a malevolent actor could use the official twitter API (Application Program Interface) to obtain data about what a user has tweeted.[80] In other cases, there might be ways of circumventing official restrictions on the data collection. Cambridge Analytica obtained the personal data of 87 Million Facebook users by the creation of the quiz app "This Is Your Digital Life". If a user, or any of their friends, used this app, their data was collected and later given to the company. Allegedly, it was then used to create political and psychological profiles of the users.[81] There are also multiple unauthorized data scrapping tools and services available on illicit marketplaces.

Beyond the large platforms, there are other ways of obtaining the personal data of individuals. So-called data brokers operate tracking networks that monitor users as they browse from website to website. This data is then assembled and sold to advertisers or anyone else willing to pay.[82] Through the hacking of websites and databases, hackers are able to obtain personal data on millions (and sometimes billions) of users. This data is often sold on criminal online marketplaces.[83] "Have I been

---

80 "GET statuses/user_timeline", Twitter (Website) online: https://developer.twitter.com/en/docs/tweets/timelines/api-reference/get-statuses-user_timeline.html/

81 Robinson Meyer, "My Facebook Was Breached by Cambridge Analytica. Was Yours?", The Atlantic (10 April 2018), online: https://www.theatlantic.com/technology/archive/2018/04/facebook-cambridge-analytica-victims/557648/.

82 Yael Grauer & Emanuel Maiberg, "What Are 'Data Brokers,' and Why Are They Scooping Up Information About You?", VICE Motherboard (27 March 2018), online: https://motherboard.vice.com/en_us/article/bjpx3w/what-are-data-brokers-and-how-to-stop-my-private-data-collection.

pwned", a website that lets people find out whether their data has been compromised lists almost 6 billion leaked accounts.[84]

3.1.2. Software and Expertise

Once the data has been collected, the next step is to utilize it to train an algorithm. This requires that the developer has access to software and the expertise to utilize the software. Both of these are now available to the public. Machine Learning is by design a very open field. The latest research is immediately published online in an open-access format, for example on the e-Print service ArXiv.[85] The leading frameworks used in the industry are also publicly available.[86] There are numerous online tutorials providing a quick and easy entry to ML.[87] This does not mean that learning ML is easy – there are a number of challenges that make the learning hard, even for trained engineers.[88] However, machine learning is increasingly becoming

---

[83] Tom Holt, "Exploring the social organisation and structure of stolen data markets", (2013) 14:2-3 Global Crime 155; Alice Hutchings and Tom Holt, "A crime script analysis of the online stolen data market", (2015) 55:3 The British Journal of Criminology 596; "McAfee Labs 2017 Threats Predictions Report", McAfee (Website), online: https://www.mcafee.com/enterprise/en-us/assets/reports/rp-threats-predictions-2017.pdf at 42.

[84] "Have I Been Pwned: Check if your email has been compromised in a data breach", Have I Been Pwned (Website) online: https://haveibeenpwned.com/.

[85] "arXiv.org e-Print archive", arXiv.org (Website), online: https://arxiv.org/.

[86] "PyTorch", PyTorch (Website), online: https://www.pytorch.org; "TensorFlow",TensorFlow (Website) online: https://www.tensorflow.org/.

[87] "fast.ai", fast.ai (Website), online: https://www.fast.ai/; "Google Launches Free Course on Deep Learning: The Science of Teaching Computers How to Teach Themselves", Open Cult (Website), online: http://www.openculture.com/2017/07/google-launches-free-course-on-deep-learning.html.

more accessible to average computer users and to criminal organizations that can hire computer experts.

3.1.3. Hardware

Another requirement for the development of artificial intelligence is access to powerful hardware. The new deep learning models rely on the massive parallel computing power of Graphical Processing Units (GPU), that allow researchers to train models much faster than traditional processors.[89] They can be purchased for several hundred dollars.[90] Recently, companies such as Google have even started to develop their own hardware to enable even more powerful models, known as TPUs (Tensor Processing Units).[91]

If one does not want to buy a graphics card, or requires more than one GPU, another possibility is to rent servers with powerful

---

88 Janakiram MSV, "Why Do Developers Find It Hard To Learn Machine Learning?", Forbes (1 January 2018), online: https://www.forbes.com/sites/janakirammsv/2018/01/01/why-do-developers-find-it-hard-to-learn-machine-learning/.

89 Colin Barker, "How the GPU became the heart of AI and machine learning", ZDNet (13 August 2018), online: https://www.zdnet.com/article/how-the-gpu-became-the-heart-of-ai-and-machine-learning/; Bernard Fraenkel, "For Machine Learning, It's All About GPUs", Forbes (1 December 2017), online: https://www.forbes.com/sites/forbestechcouncil/2017/12/01/for-machine-learning-its-all-about-gpus/; Fidan Boylu Uz, "GPUs vs CPUs for deployment of deep learning models", Microsoft Azure (11 September 2018), online: https://azure.microsoft.com/en-us/blog/gpus-vs-cpus-for-deployment-of-deep-learning-models/; LeCun, Bengio & Hinton, supra note 38 at 4.

90 Tim Dettmers, "Which GPU(s) to Get for Deep Learning", Tim Dettmers (5 November 2018), online: http://timdettmers.com/2018/11/05/which-gpu-for-deep-learning/.

91 "Cloud TPUs - ML accelerators for TensorFlow", Google Cloud (Website), online: https://cloud.google.com/tpu/.

GPUs. Amazon, Microsoft and Google offer servers for rent that are specifically configured to accommodate the types of computation required for Deep Learning. These companies also offer the possibility of renting machines with several GPUs, which allow for the creation of more complex models and the integration with systems designed to support Deep Learning tasks.[92]

3.2. Harmful uses of artificial intelligence

We have now demonstrated that AI is established to the point were any dedicated developer is able to enter the field using publicly available resources. As mentioned, such accessibility is generally a positive thing, however, it also potentially allows malicious actors to leverage the technology. There are several properties of AI which might make it attractive for malicious actors. Like many technologies, it can serve dual purposes and can be used both for beneficial and harmful ends. AI can emulate many acts performed by humans, and in some cases even exceed human performance in terms of efficiency and scalability. This means that crimes that previously required human skills and time can be performed on a much larger scale, targeting thousands of victims simultaneously.[93] AI can also increase the distance between the offender and the victims. This could make criminals harder to track and decrease psychological

---

92 "Amazon Deep Learning AMIs", Amazon Web Service (Website) online: https://aws.amazon.com/machine-learning/amis/; "Cloud AI | Cloud AI", Google Cloud (Website), online: https://cloud.google.com/products/ai/; jonbeck7, "Azure Windows VM sizes - GPU", Microsoft (Website), online: https://docs.microsoft.com/en-us/azure/virtual-machines/windows/sizes-gpu.

93 Supra note 71 at 16-17.

inhibitions.[94] Additionally, artificial intelligence, like any technological system, is bound to suffer from a number of technical vulnerabilities that will inevitably be exploited by criminal interests.

Therefore, there are three impending consequences regarding the risks posed by AI:

1. Existing threats could expand: due to the scalability of artificial intelligence, offenders could use the technology to target an increasing number of victims;
2. Entirely new threats could be introduced: AI is able to generate data such as audio files mimicking the voice of real people. These could be used to carry out entirely new types of attacks and be exploited for novel criminal activities;
3. The nature of threats could change: due to the capabilities of artificial intelligence, crimes could become more effective, targeted and difficult to attribute.[95]

Artificial intelligence therefore significantly changes the kinds and the amount of harm that can be directed against computer users.

3.3. Approaches of malevolent artificial intelligence

This section provides an overview of the various criminal strategies that could be facilitated by malevolent uses of AI. This

[94] Ibid at 17.
[95] Ibid at 18-22.

section is not meant to be exhaustive as the nature of criminal innovation is always unpredictable, but seeks to highlight a number of areas that could be affected by the availability of artificial intelligence.

3.3.1. Social Engineering

Social engineering has been defined as "any act that influences a person to take an action that may or may not be in their best interest."[96] It is an effective attack strategy targeting human rather than technical vulnerabilities that can be extremely hard to protect against, for individuals and companies alike.[97] In this subsection, we describe the numerous approaches in social engineering that could be significantly expanded and facilitated by artificial intelligence.

3.3.1.1. Phishing

Instead of using the voice, people may also use the method of 'phishing', which can be defined as the 'practice of sending emails appearing to originate from reputable sources with the goal of influencing or gaining personal information'.[98] It is likely the most widespread type of social engineering.[99] Typically, an attacker will create an email that purports to originate from a trustworthy source, such as a financial institution, tech support

---

96 "Social Engineering Defined", Security Education (Website), online: https://www.social-engineer.org/framework/general-discussion/social-engineering-defined/.

97 Ian Mann, Hacking the human: Social engineering techniques and security countermeasures, (London: Routledge, 2008).

98 "Phishing", Security Through Education (Website), online: https://www.social-engineer.org/framework/attack-vectors/phishing-attacks-2/.

99 Ibid.

service or a government institution. These emails will then be sent out in bulk. A person who clicks on a link will be taken to a counterfeit but convincing website where they are asked to enter their personal information.[100] The email might also contain an attachment which, once clicked, infects the victim's computer with malware. There are many ways an attacker might try to convince the user that the email is real, such as by altering the email address so that it seems legitimate or buying web domains that are very similar to the official domain names of the institutions being targeted.

A more personalized variant is called spear-phishing. Instead of sending an email to users in bulk, spear-phishing operations target specific users with meticulously crafted emails. These emails might be based on data obtained from social media or any other open source intelligence the attacker has been able to gather on the target.[101] For example, an email containing a link to a CV might be sent to a recruiter. In order to view the CV, the user is asked to log into their Microsoft account, through a page that mirrors exactly the look and feel of the real Microsoft portal. However, once users enter their details, the log-in credentials are instead harvested by the attacker, who are then able to compromise their victims' accounts. While very effective, spear-phishing requires attackers to perform a significant amount of background research and to create credible messages, limiting its use to high-value targets.[102]

---

100 Ibid; "Phishing", Know4Be (Website), online: https://www.knowbe4.com/phishing.

101 Ibid; "Spear Phishing", Know4Be (Website) online: https://www.knowbe4.com/spear-phishing/.

102 Supra note 71 at 19.

There is a big risk that artificial intelligence might enable criminals to combine the scale of regular phishing attacks with the targeted nature and effectiveness of spear-phishing. A system could be designed that would crawl a large number of targets' online presence, such as social media feeds. Profiles of these users could then be created, that would include which interests they have, which companies they have relationships with, and mapping patterns of online activity. Based on this information, a highly persuasive email might be created or selected by the machine. This could be done at a massive scale, unconstrained by the need for human operators. Additionally, the artificial intelligence system would be able to learn what works based on response or click rates, and subtly alter each message to circumvent phishing filters deployed by the victims' mail platforms. A recent study showed how effective such strategies could be and how easily they could be organized. Using a Machine Learning algorithm, a group of researchers were able to identify the interests of a group of targets by analyzing their Twitter activity. They then used the algorithm to word and send them personalized messages that contained a potentially malicious link, drawing on the content of messages that had been identified as resonating with the victims' interests. They also timed the fake messages with the period of the day when the victims seemed most active on the social platform, to maximize the chances of engagement. They then tracked how many users clicked on the embedded links that could have been malicious, had the researchers been criminal hackers instead. Between 33 and 66% of the targets clicked on the links, eclipsing the 5 to 14% usually achieved with mass phishing.[103]

[103] John Seymour & Philip Tully, "Weaponizing data science for social engineering: Automated E2E spear phishing on Twitter" (Paper delivered

3.3.1.2. Vishing

Vishing (a portmanteau of the words 'voice' and 'phishing') is the "practice of eliciting information or attempting to influence action via the telephone."[104] An attacker might manipulate its mark by claiming to work for the victim's bank, to be a Microsoft support employee or to represent a tax agency.[105] The scams can have devastating consequences – supposedly, victims of phone-based scams lost on average 720 USD in 2017.[106] Due to the propensity of people to trust phone calls, these attacks can be hard to defend against.[107] Even tech-savvy people can fall for the more advanced methods.[108] However, these frauds often require a lot of preparation and a skilled and convincing operator to pull them off.[109] The attacks can also take some time to perform, which limits the rate of victimization.

---

at Black Hat USA 2016, DEF CON 24, 2016), online: 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 at 8.

104 "Vishing", Security Through Education (Website), online: https://www.social-engineer.org/framework/attack-vectors/vishing/.

105 Rasha AlMarhoos, "Phishing for the answer: Recent developments in combating phishing", (2007) 3:3 I/S: A Journal of Law and Policy for the Information Society 595.

106 "The top frauds of 2017", Consumer Information, (1 March 2018), online: https://www.consumer.ftc.gov/blog/2018/03/top-frauds-2017.

107 "New Phishing Techniques To Be Aware of: Vishing and Smishing", MakeUseOf (Website), online: https://www.makeuseof.com/tag/new-phishing-techniques-aware-vishing-smishing/.

108 Brian Krebs, "Voice Phishing Scams Are Getting More Clever", Krebs on Security (Website), online: https://krebsonsecurity.com/2018/10/voice-phishing-scams-are-getting-more-clever/ation.roughgetting-more-clever/<.

109 "Let's Go Vishing", (22 December 2014), online: Security Through Education. https://www.social-engineer.org/general-blog/lets-go-vishing>.

This might change with artificial intelligence. The same techniques used to create a helpful chatbot, such as Apple's Siri[110] or Amazon's Alexa[111], can also be used to create a computer system able to imitate a human. Google has already proven that artificial intelligence can be used to create phone call operators that are virtually indistinguishable from real humans in tone and phrasing. This system, known as Duplex, is able to call restaurants and hair dressers to book a table or make an appointment without the employees at the other end of the line noticing they are interacting with a machine.[112] By using AI methods of realistic voice generation and natural language processing to respond to queries, criminal hackers[113] could thus create automated targeting operations. Even if they are not as effective as human operators, these systems could be deployed at a much larger scale, targeting thousands of individuals per day. This is thus an area where AI could increase the scale of crime. Brian Krebs describes for example how there are already systems using artificial intelligence to target individuals using a vishing stratagem. He describes a person's experience of being called by the employee of a Credit Alert Service. The caller sounded very realistic and was able to answer simple questions. However, after some more complicated enquiries, the caller was seamlessly

110 "Siri", Apple (Website), online: https://www.apple.com/siri/.

111 "Ways to Build with Amazon Alexa", Amazon (Website), online: https://developer.amazon.com/alexa.

112 "Google Duplex: An AI System for Accomplishing Real-World Tasks Over the Phone", Google AI (Blog), online: http://ai.googleblog.com/2018/05/duplex-ai-system-for-natural-conversation.html.

113 We deliberately use the term 'criminal hacker' to avoid the usual confusion between the majority of technology enthusiasts who like to tinker with software and hardware and the small minority of this group that uses their technical expertise to deliberately break the law.

switched out for a real human who attempted to finalize the fraudulent exchange. This shows how voice recognition and generation can be used to automate vishing operations.[114]

Artificial intelligence could even be used to create new attack vectors in vishing. Lyrebird, a Montreal-based AI startup launched in 2017 allows a user to train a synthetic version of their voice by recording a few sentences of their real voice.[115] Malicious actors could use this technology to generate voice messages that sound like they come from close relatives or friends (by training the machine with publicly-available videos or fake calls made to the persons whose voices need to be counterfeited), tricking the user to give out information.[116] This new capacity could alter the trust we place in a voice.[117]

3.3.1.3. Astroturfing

Another practice that might be exacerbated by AI is astroturfing. It consists of creating fake grassroot movements that seem to be genuine and wide-spread but in fact stem from very few actors.[118] There are several firms which offer

---

114 Krebs, supra note 108.

115 Francesc Cristiani, "How Lyrebird Uses AI to Find Its (Artificial) Voice", Wired (15 October 2018), online: https://www.wired.com/brandlab/2018/10/lyrebird-uses-ai-find-artificial-voice/; "Lyrebird: Ultra-Realistic Voice Cloning and Text-to-Speech", Lyrebird.a (Website), online: https://lyrebird.ai/.

116 Malicious Use of Artificial Intelligence, supra note 71 at 20.

117 Abhimanyu Ghoshal, "I trained an AI to copy my voice and it scared me silly", The Next Web (22 January 2018), online: https://thenextweb.com/insights/2018/01/22/i-trained-an-ai-to-copy-my-voice-and-scared-myself-silly/.

118 Thomas P Lyon & John W Maxwell, "Astroturf: Interest Group Lobbying and Corporate Strategy" (2004) 13:4 J Econ Manag Strategy 561; Kevin Grandia, "Bonner & Associates: The Long and

astroturfing as a service and provide software that allow employees to manage several online personas.[119] Astroturfing can be used by corporations to review their products in order to make them seem more desirable. Some claim that up to one third of online reviews are fake.[120] Astroturfing can also be used for political manipulation, by for example tweeting or sharing a certain viewpoint. A study showed for example that astroturfing techniques could be very effective in raising doubts about the origins of global warming.[121] Thus, fringe political views can be made to seem mainstream and to appear on the "trending" section on such social media websites as Twitter. Bots were allegedly used leading up to and after the 2016 U.S. presidential election to shift the public view towards voting for Trump, to make his base seem stronger than it was, or to discourage certain voters from voting at all.[122] In a consultation by the Federal Communications Commission (FCC) in the U.S., millions of briefs in favor of abolishing net neutrality were apparently filed by fake accounts, many under the names of dead people. A data scientist discovered 1.3 million comments

---

Undemocratic History of Astroturfing", Huffington Post (26 August 2009), online: https://www.huffingtonpost.com/kevin-grandia/bonner-associates-the-lon_b_269976.html.

119 Grandia, supra note 118; David Streitfeld, "Book Reviewers for Hire Meet a Demand for Online Raves", The New York Times (25 August 2012), online: https://www.nytimes.com/2012/08/26/business/book-reviewers-for-hire-meet-a-demand-for-online-raves.html.

120 Streitfeld, supra note 119.

121 Charles Cho et al, "Astroturfing Global Warming: It Isn't Always Greener on the Other Side of the Fence" (2011) 104:4 J Bus Ethics 571.

122 Jon Swaine, "Russian propagandists targeted African Americans to influence 2016 US election", The Guardian (17 December 2018), online: https://www.theguardian.com/us-news/2018/dec/17/russian-propagandists-targeted-african-americans-2016-election.

that followed extremely similar linguistic constructions and were thus likely fake.[123]

Artificial intelligence could potentially drastically increase the efficiency of astroturfing. Twitter, for example, uses anti-bot mechanisms to detect and ban fake accounts.[124] This means that attackers have to "herd" accounts by registering them, adding pictures, occasionally tweeting and following other users.[125] Artificial intelligence could be used to automate this process. It could also be used to automatically generate messages that disseminate the same information but are unique enough to not be detected as similar. Finally, AI could be used to better target messages so they become more convincing to certain people based on their socio-demographic characteristics or psychological traits.[126] The practice of astroturfing could be used to slander or harass people at an unprecedented scale.

---

123 Jeff Kao, "More than a Million Pro-Repeal Net Neutrality Comments Were Likely Faked", Hacker Noon (23 November 2017), online: https://hackernoon.com/more-than-a-million-pro-repeal-net-neutrality-comments-were-likely-faked-e9f0e3ed36a6.

124 Brian Krebs, "Buying Battles in the War on Twitter Spam", Krebs on Security (Website) online: https://krebsonsecurity.com/2013/08/buying-battles-in-the-war-on-twitter-spam/.

125 "Astroturfing, Twitterbots, Amplification - Inside the Online Influence Industry", The Bureau of Investigative Journalism (7 December 2017), online: https://www.thebureauinvestigates.com/stories/2017-12-07/twitterbots.

126 Matt Chessen, "The Madcom Future: How Artificial Intelligence Will Enhance Computational Propaganda, Reprogram Human Culture, and Threaten Democracy...and What Can Be Done About It", The Atlantic Council (1 September 2017), online: https://www.scribd.com/document/359972969/The-MADCOM-Future at 13.

3.3.2. Generation

As previously mentioned, artificial intelligence can be used to generate extremely realistic-looking data. This can be used for social engineering purposes, but also for new attack vectors. Humans have learned that images can be easily manipulated using tools such as Adobe Photoshop. However, with AI, even media such as sound and video can be counterfeited in convincing ways and on a massive scale. As mentioned before, this is a possibility that is being actively exploited in the wild. It therefore might be the most visible malicious use of artificial intelligence. The trend started in early 2018, when a user of the internet forum Reddit created and publicly released a tool he called FakeApp, which received over 100,000 downloads.[127] It allows any user with a sufficiently strong graphics card to generate fake videos using deep learning networks that rely on a technology known as autoencoders.[128] The user simply supplies a low number of pictures or videos of a targeted person. The neural network then 'learns' the face of that person. Next, the user supplies another video and designates a target face. The neural network will then generate a new video, rendering the face of the target person onto the face of the person in the target video. This includes the adaptation of facial expressions and can be very realistic looking.[129]

---

127 Roose, supra note 66.

128 Gaurav Oberoi, "Exploring DeepFakes", Hacker Noon (5 March 2018), online: https://hackernoon.com/exploring-deepfakes-20c9947c22d9; Alan Zucconi, "Understanding the Technology Behind DeepFakes", Alan Zucconi (14 March 2018), online: https://www.alanzucconi.com/2018/03/14/understanding-the-technology-behind-deepfakes/.

129 Ibid.

Figure 5 - A screenshot of a video created by the podcast RadioLab where Barack Obama is made to look like he is saying words he never uttered.[130]

The program found widespread uses, mainly for humorous purposes such as a public service announcement by President Obama, advising the public not to trust videos,[131] or videos featuring the actor Nicolas Cage playing all roles in a movie.[132] The technology of DeepFakes has many beneficial uses such as helping education, art and autonomy.[133] However, the most publicized and malicious use involved the creation of adult material. A large number of videos showing famous movie and music stars were published on social media websites and adult websites, before subsequently being banned. Users also seemed

130 "This PSA About Fake News From Barack Obama Is Not What It Appears", BuzzFeed News (17 April 2018), online: https://www.buzzfeednews.com/article/davidmack/obama-fake-news-jordan-peele-psa-video-buzzfeed.

131 Ibid.

132 Usersub, supra note 67.

133 Robert Chesney & Danielle Keats Citron, "Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security", (2019) 107 California Law Review (forthcoming) at 15-17.

to attempt creating videos showing their friends or previous love relationships.[134]

The DeepFake tool highlights a number of issues that could arise with the use of artificial intelligence. The first is the easy spread of the technology.[135] Before the advent of AI, creating a fake pornographic video involving any person might be possible using special effects technology, but that remained extremely expensive and required a lot of skills. However, AI enables one sufficiently-skilled individual to create a tool performing this task, and then make it available to almost anyone with a very moderate level of technical expertise.[136] Further, the tool highlights the potential breakdown of several trust vectors in society. Many of the videos created by the DeepFake tool are already very realistic looking. However, there are often artifacts giving the public a feeling that something is off. It can be assumed that the creation of fake, highly realistic videos will one day be possible, which will undermine trust in video material as proof that something is true.

Beyond pornographic material, The DeepFake tool can therefore be used by anyone to create a video of a person performing any act or saying anything. A German comedian created a fake video of a Greek minister pointing the finger to the crowd in 2016, causing media to speculate for days whether the video was real or fake and likely shaping public opinion. Creating videos of public figures could be used by politicians or entrepreneurs to discredit their opponents or competitors,

134 Cole & Maiberg, supra note 69.
135 Malicious Use of Artificial Intelligence, supra note 71 at 17.
136 Chesney & Citron, supra note 133 at 8-9.

or by nation states to attack the democratic processes or destabilize their adversaries. A small number of these well-executed attacks could potentially precipitate a break-down of public trust in the mass media. How could one trust any video recording if they could all be fake? Not only could it mean that some public figures have to constantly defend against allegations of wrongdoing or corruption captured on tape, but it could also provide a plausible deniability defense to personalities for things they actually did say or do.

Public figures are not the only ones that are exposed to such attacks. Most people these days will have a number of images of them publicly available online. A malicious person could use the DeepFake tool to create videos or audios of these people aiming to destroy their reputation. This could be anything from a disgruntled former partner, an angry employee or simply someone wishing to cause harm. There are many things that could cause irreparable harm to an individual, such as the starring in pornographic content or the uttering of racist remarks. Even if the person denies the accuracy of the video, this might not be enough – rumors travel fast—especially on social media—and once opinions are formed, they can at times be very difficult to change. Further, with the advent of instant internet searches for individuals, a negative story might result in a person having trouble finding a job or restoring their reputation for the rest of their life.

3.3.3. Cybersecurity

In our highly connected society, a large attack vector stemming from AI is that of cybersecurity. Writing and maintaining secure software and platforms is a task that depends on highly trained

experts that are in very short supply. Further, many companies might not have the resources or incentive to secure their systems, resulting in very high rates of avoidable vulnerabilities. Recently, the sophistication of cyber-attacks has been on the increase, due in part to the leakage of very sophisticated toolsets developed by intelligence agencies. Cybercriminals have also taken stock of our growing dependence on digital technologies and data and have developed new business models such as ransomware as a response. The ransomware business model abandons the theft of personal data that used to be resold to third parties on online criminal marketplaces. Instead, the value is extracted from the victim herself, who pays the offenders to regain access to her precious personal information.[137] This section will look at the way criminal hackers could use artificial intelligence to further improve the scale and effectiveness of their attacks.

3.3.3.1. Vulnerability discovery

Many computer viruses depend on the exploitation of a system vulnerability. This could be a bug in an operating system (such as Windows) or a software (such as Adobe Reader) or even a web technology (such as WordPress, a tool for online publishing) that allows a hacker to gain access to a system and steal information or execute their own code. Vulnerabilities, once discovered, have to be patched quickly by software providers so that as little damage as possible can be caused by them. Vulnerabilities that are used by a virus to infect a machine

[137] Masarah Paquet-Clouston, Bernhard Haslhofer & Benoît Dupont, "Ransomware payments in the bitcoin ecosystem", (Paper delivered at the 17th Annual Workshop on the Economics of Information Security (WEIS), 2018) online: https://arxiv.org/abs/1804.04080.

before a company has patched them are referred to as *zero-day* exploits. The StuxNet virus leveraged four of these vulnerabilities. These can be extremely valuable on the black market, leading many companies to offer bug bounties to researchers that disclose vulnerabilities to them first.

There are several methods to discover these vulnerabilities. Static Analysis requires a researcher to analyze the code of the program, manually or semi-automatically. Fuzzing feeds the program billions of random permutations to see when it fails. In penetration testing, a researcher pretends to be a hacker and discover the vulnerability by trying to enter the system.[138] These techniques can be used by researchers to discover and patch vulnerabilities in their own software, but also by attackers looking to find and exploit vulnerabilities.[139] The discovery of vulnerabilities requires a skilled analyst.[140]

The deployment of Artificial Intelligence could lead to an increase both in the quality and quantity of attacks. Researchers have shown promising approaches to further automating parts of vulnerability discovery using artificial intelligence.[141] Until

---

138 B Liu et al, "Software Vulnerability Discovery Techniques: A Survey" (Paper delivered at the Fourth International Conference on Multimedia Information Networking and Security, 2012), online: https://ieeexplore.ieee.org/document/6405650.

139 Malicious Use of Artificial Intelligence, supra note 71 at 16.

140 Daniel Votipka et al, "Hackers vs. Testers: A Comparison of Software Vulnerability Discovery Processes" (Paper delivered at the 2018 IEEE Symposium on Security and Privacy, San Francisco, CA, 2018), online: https://ieeexplore.ieee.org/document/8418614.

141 Gustavo Grieco & Artem Dinaburg, "Toward Smarter Vulnerability Discovery Using Machine Learning". (Paper delivered at the Proceedings of the 11th ACM Workshop on Artificial Intelligence and Security, Toronto, Canada, 2018); Steven Harp et al, "Automated

now, fuzzing has been hard to set up in use. Artificial Intelligence could be used to learn the data structures that a program relies on and then inject fake data automatically. This could increase the number of people able to perform these attacks and thus the number of vulnerabilities discovered.[142] In 2017, researchers at Microsoft demonstrated how neural networks could be used to make fuzzing simpler, more efficient and more generic.[143]

A weak password could also be a sort of vulnerability, since it allows a hacker to access the account of a user.[144] Researches have demonstrated that artificial intelligence can be very strong at guessing passwords. It can be trained on millions of leaked passwords to detect patterns and then apply these to guess the passwords of specific users.[145]

Vulnerability Analysis Using AI Planning" (Paper delivered at the 2005 AAAI Spring Symposium, Stanford, CA, 2018), online: https://www.researchgate.net/publication/221250445_Automated_Vulnerability_Analysis_Using_AI_Planning at 8.

142 FortiGuard SE Team, "Predictions: AI Fuzzing and Machine Learning Poisoning", Fortinet Blog (15 November 2018), online: https://www.fortinet.com/blog/industry-trends/predictions--ai-fuzzing-and-machine-learning-poisoning-.html.

143 "Neural fuzzing: applying DNN to software security testing", Microsoft Research (13 November 2017), online: https://www.microsoft.com/en-us/research/blog/neural-fuzzing/; Mohit Rajpal, William Blum & Rishabh Singh, "Not all bytes are equal: Neural byte sieve for fuzzing", (2017) arXiv Working Paper, arXiv:1711.04596 [cs.SE], online: https://arxiv.org/abs/1711.04596 at 10.

144 Julie J.C.H. Ryan, "How do computer hackers 'get inside' a computer?", Scientific American, online: https://www.scientificamerican.com/article/how-do-computer-hackers-g/.

145 Briland Hitaj et al, "PassGAN: A Deep Learning Approach for Password Guessing" (2017) arXiv Working Paper, arXiv:170900440 [cs, stat], online: http://arxiv.org/abs/1709.00440.

3.3.3.2. Exploitation

Even after the vulnerability is discovered, the work of the attacker is not finished. He will try to find a way to use the exploit to get access to one or many target machines. This can be done, for example, through the creation of a computer virus, that tries to autonomously attack as many computers as possible using the exploit. It can also be used to perform a regular cyber-attack against a server. Here, the attacker himself runs commands to move laterally toward other machines.

On the defense side, machine learning is used to monitor for these kinds of attacks. Anti-virus programs often use two ways of identifying malware: Signature-based technologies and behavioral analysis. Signature-based analysis tries to identify a virus based on the digital fingerprint of its code. It relies on the anti-virus vendor identifying malware and adding it to a database of malicious signatures.[146] Behavioral analysis identifies what a program tries to do rather than which code is it based on.[147] It often uses ML technologies.[148] Artificial Intelligence could be used to circumvent these systems. Researchers have showed that it is possible to create AI systems that automatically create malware that evades common anti-virus programs.[149] Attackers could use AI to ever so slightly alter a program until

---

146 John Cloonan, "Advanced Malware Detection - Signatures vs. Behavior Analysis", Infosecurity Magazine (11 April 2017), online: https://www.infosecurity-magazine.com:443/opinions/malware-detection-signatures/.

147 Ibid.

148 Malicious Use of Artificial Intelligence, supra note 71 at 33.

149 Hyrum S Anderson et al, "Learning to Evade Static PE Machine Learning Malware Models via Reinforcement Learning" (2018) arXiv Working Paper, arXiv:180108917 [cs], online: http://arxiv.org/abs/1801.08917.

it appears benign to anti-virus filters.

Likewise, server systems often run protective software known as Intrusion Detection Systems, that check for strange behavior on servers or traffic and report this to administrators. They often use machine learning technologies.[150] For example, if a server suddenly starts transferring massive amounts of data to an IP-address in Russia, this might indicate that a hack is underway. However, it might also just be a sign of Russian users following a popular link to access the website. A hacker could use AI to try to circumvent these systems by hiding their activity under the guise of human-looking behaviors. Mimicry attacks, that try to slip under the radar, have been demonstrated to be efficient.[151] Using machine learning to automate these seems a natural evolution.

3.3.3.3. Post-Exploitation & Data Theft

After the exploit, the attacker will often use the established access to install their own backdoor that they can use to re-enter the server, getting deeper access to the system and looking around the server for potentially sensitive information and downloading this information.[152] Other hackers might use the

---

150 P García-Teodoro et al, "Anomaly-based network intrusion detection: Techniques, systems and challenges" (2009) 28:1–2 Computers & Security 18; Alex Shenfield, David Day & Aladdin Ayesh, "Intelligent intrusion detection systems using artificial neural networks" (2018) 4:2 ICT Express 95.

151 David Wagner & Paolo Soto, "Mimicry Attacks on Host-Based Intrusion Detection Systems" (Paper delivered at the 9th ACM conference on Computer and communications security, Washington DC, 2002), online: https://dl.acm.org/citation.cfm?id=586145 at 10.

152 Ivan Novikov, "How AI Can Be Applied To Cyberattacks", Forbes (22 March 2018), online: https://www.forbes.com/sites/forbestechcouncil/2018/03/22/how-ai

access to gain further access to the operations of the company or destroy services to cause financial damage. Throughout this process, the hacker has to take care to stay hidden and erase any traces that might tell the operator he has been on the server and lead him to getting caught. It is a complicated process, requiring a lot of patience, skill, and knowledge of the computer system. The attacks are often constrained by the speed of human reaction – based on the information the hacker sees on the server, he will have to react in a different way.

While this is more far-fetched than the other applications, it could potentially be possible for hackers to train an Artificial Intelligence system to automate parts of these steps as well. There are already frameworks, created for security auditing of computer systems, that allow people to unleash an entire barrage of attacks on a computer system.[153] Artificial intelligence might enhance the capability of these systems to automatically infer which attacks are appropriate, or which data might be sensitive and should therefore be given priority. Such a system could be used in parallel to intelligently exploit many systems simultaneously, without requiring human intervention. While this is already possible to some extent, Artificial Intelligence might be able to enhance these capabilities.

3.3.4. Exploitation of deployed artificial intelligence

Most analysts see artificial intelligence as having a large effect on most, if not all, sectors of society. This might lead to another

-can-be-applied-to-cyberattacks/.

153 "Penetration Testing Software, Pen Testing Security", Metasploit (Website), online: https://www.metasploit.com/.

attack vector opening up for malicious users. As mentioned before, the current crop of artificial intelligence systems suffers from a number of weaknesses. If they are implemented in a large sector of society, they risk enabling new attacks that exploit this fragility. Depending on the way AI is implemented, and how much control it is given over people and processes, this could cause tremendous damage to society.

3.3.4.1. Adversarial attacks

Adversarial attacks are attacks that exploit the fact that AI does not operate like human intelligence. Artificial intelligence in general, and convolutional neural networks in particular, identify patterns based on a set of features that might be very unintuitive for humans. By slightly altering the input, one can completely change the way the AI system interprets a pattern. It has been shown that a picture of a puppy can be altered in ways that are imperceptible to humans. These effects can also be implemented in real-world scenarios – a team of researchers showed that an altered 3d-printed turtle could be classified as a gun in a video feed, no matter the orientation of the turtle. Researchers have even shown that the addition of stripes to traffic signs can alter the meaning of that sign for the AI running on autonomous vehicles.

It is important to note that an attacker typically requires access to a neural network in order to generate adversarial examples. However, often pretrained networks are used, which means that the models are readily available on the internet.[154] Recent research

[154] Arelis Guzmán, "Top 10 Pretrained Models to get you Started with Deep Learning (Part 1 - Computer Vision)", Analytics Vidhya (27 July 2018), online:

also shows that adversarial examples can be created by first training another neural network to mimic the target network.[155]

There are many potential attacks that might be carried out by exploiting this weakness. The malicious conversion of a yield sign to a go sign could be a recipe for disaster in traffic. Likewise, a system set up for detecting weapons might be confused by a gun designed to resemble a more innocuous object and interpreted as such by a neural network. Neural networks designed to detect anti-virus software is also vulnerable to malware crafted using adversarial techniques.[156] If a model directing autonomous weapon systems is targeted, the results could be that civilians are harmed.[157] The creation of neural networks that are resistant to adversarial attacks is an active area of research,[158] however until reliable countermeasures are implemented, the increasing use of AI opens society to new attack vectors.

3.3.4.2. Poisoning of artificial intelligence systems

Another attack against AI systems relies on the poisoning approach. Instead of subverting the algorithm itself by

---

https://www.analyticsvidhya.com/blog/2018/07/top-10-pretrained-models-get-started-deep-learning-part-1-computer-vision/.

155 Nicolas Papernot et al, "Practical Black-Box Attacks against Machine Learning" (2016) arXiv Working Paper, arXiv:160202697 [cs], online: http://arxiv.org/abs/1602.02697.

156 Kathrin Grosse et al, "Adversarial Perturbations Against Deep Neural Networks for Malware Classification" (2016) arXiv Working Paper, arXiv:160604435 [cs], online: http://arxiv.org/abs/1606.04435.

157 Malicious Use of Artificial Intelligence, supra note 71 at 20.

158 Kao, supra note 123; Xiaoyong Yuan et al, "Adversarial Examples: Attacks and Defenses for Deep Learning" (2017) arXiv Working Paper, arXiv:171207107 [cs, stat], online: http://arxiv.org/abs/1712.07107.

manipulating data or objects on the outlier of its model, poisoning relies on attacking the training data used to create the AI system. If this data is of poor quality, the resulting machine learning system will not operate correctly. The addition of quite few poisoned examples can be enough to severely damage the performance of an AI system.[159] Poisoning attacks rely on the attackers having control over some of the data used to train the AI. This makes the attack unfeasible in many instances. However, due to the large requirements of data for machine learning, data will often be crowd-sourced. Another issue is that of online learning. This is a common approach in anomaly detection. Here, the system is constantly trained to analyze a baseline of activity in a system. Only if an event falls outside of this baseline will the detector notice the anomaly. This could be exploited by attackers. Over time, they could inject patterns that are still within the allowed parameters, but close to the edge of what is allowed. This will extend the baseline to cover more situations. After extending the baseline this way for some time, the attackers can launch their attack without being detected.[160]

---

159 Battista Biggio, Blaine Nelson & Pavel Laskov, "Poisoning Attacks against Support Vector Machines" (2012) arXiv Working Paper, arXiv:12066389 [cs, stat], online: http://arxiv.org/abs/1206.6389.

160 Benjamin IP Rubinstein et al, "ANTIDOTE: understanding and defending against poisoning of anomaly detectors" (Paper delivered at the 9th ACM SIGCOMM Conference on Internet Measurement, 2009), online: https://people.eecs.berkeley.edu/~tygar/papers/SML/IMC.2009.pdf; Nitika Khurana, Sudip Mittal & Anupam Joshi, "Preventing Poisoning Attacks on AI based Threat Intelligence Systems" (2018), arXiv Working Paper, arXiv:1807.07418 [cs.SI], online: https://arxiv.org/abs/1807.07418v1; Maria Korolov, "Hackers get around AI with flooding, poisoning and social engineering", CSO Online (16 December 2016), online: https://www.csoonline.com/article/3150745/security/hackers-get-

3.4. Conclusion

This section has looked at the possibility of malicious actors using AI as a criminal tool or as a target. While AI is an active area of research, and has typically been restricted to the research community, a recent wave of democratization has meant that advanced AI tools are becoming widely available. This is a positive development that unfortunately also opens the door for offenders to exploit artificial intelligence. Data can be obtained from many sources online, such as the hacking of websites or the massive collection of personal data from social media platforms. Due to the openness of the ML community, both the algorithms needed and the skills required can be found online. The creation of ever more powerful hardware in the form of graphics cards and the possibility of easily renting these resources online make these required infrastructures of artificial intelligence available to malicious actors as well.

Like many technological developments, AI is characterized by its dual use – it has applications both for socially beneficial and malicious ends. It can be used to make crimes more *efficient*. This can be seen in the cases of social engineering and cyber security. These attacks are typically resource-intensive to execute, and are therefore usually restricted to high value targets or to victims that can generate attractive gains. The automation of some of these aspects could open the door for criminal hackers to industrialize and personalize their attacks at the same time—a concerning increase in capacity. Artificial intelligence could for example lead to phishing being executed at the same scale but in a more targeted manner by automatically creating emails that

around-ai-with-flooding-poisoning-and-social-engineering.html.

users are more likely to respond to. Finally, cyber-attacks might be carried out in automated ways, with AI predicting how best to gain access to a server and how to proceed in the most effective and stealthy manner.

Artificial intelligence also creates the possibility of developing completely *new forms of attacks*. Artificial intelligence can, for example, be used to generate accurate simulations of a user's voice. This is not something people are accustomed to, and therefore we still assume that the voices of our friends and relatives actually belong to them. This is something that might not hold up in the future. Further, realistic looking videos can be generated by simply using a few images of a person's face. These could be used to undermine the reputation of a person or even for blackmail. AI can also be used to create botnets that defraud users or misrepresent the true extent of certain views in the population. Criminal hackers will also leverage AI to develop new capacities such as the automated discovery of critical vulnerabilities and the circumvention of existing intrusion detection systems.

Finally, a new class of attacks might spring from the widespread deployment of artificial intelligence. AI makes it likely that a growing share of our life will be automated in the near future. Current versions of AI systems remain fragile to poisoning attacks. This could be used to devastating effects, for example in connection with the deployment of self-driving vehicles.

Again, this assessment remains largely speculative, and it is still uncertain how and when these tools will be used by offenders. However, criminal groups have proven willing to quickly adopt

new technologies when they provided them with new profitable opportunities. In that context, how should society adapt its control mechanisms to minimize the risks we have outlined in this chapter? For the social engineering and generation attacks, this seems to come down to two main courses of action: Countermeasures and education. It should be noted that the very same technologies that could be used by offenders can also be used to detect these attacks, for example by training AI systems to detect the slight accent in generated voices or videos. Education is likely to be an equally important measure. The public needs to be made aware of the fact that many of the old assumptions may no longer hold up. For example, videos might be faked, and emails and phone calls asking them for their information could be generated by machines to separate them from their money. Depending on the nature and extent of forthcoming attacks, this could be a painful adjustment period for many people. This adaptation process will even be potentially more arduous in the case of cybersecurity. Even a single vulnerability or attack can cause billions of dollars worth of damage. If AI can be used to quickly generate many of these, the number of data leaks will grow exponentially.

4. ARTIFICIAL INTELLIGENCE IN LAW ENFORCEMENT

Imagine that the year is 2054. Touch screen technology is commonplace. Ads are tailored and customized based on a person's life, decisions, whereabouts, and user history. Cars can drive on their own. Home appliances can be controlled with one's voice. Biometric recognition such as palm prints and identifying facial scans, is commonplace. The police are able to predict who is likely to commit a crime and apprehend that person before they do so. It is no accident that every element in that description refers to the plot of the 2002 American science fiction film called *Minority Report*. Indeed, all descriptive elements in the preceding paragraph are true at the time of writing except for the year and the statement that police habitually apprehend a person *before* they commit a crime.

As we shall see below, law enforcement around the world have begun using AI-powered technology to investigate and at times even try to predict crimes. While there is a long history of the use of technology in criminal investigations, the use of AI has the power to transform the relationship between police officers and citizens and to facilitate unprecedented surveillance and social control. We take stock of the current tools in use that assist the police in detecting and investigating crime, and offer a taxonomy of such tech in terms of its AI capacity. We also canvas the emerging tools that promise to predict crime by determining crime hotspots and who is likely to be involved in

gun violence. We offer an overview of some of the ethical issues raised by AI, as well as ways forward for governments and law enforcement that want to add AI to their crime response toolbox. Our aim is to critically assess the moral and technical authority that AI is often presumed to display, and suggest a human-centric approach to the implementation of artificial intelligence tools by law enforcement.

A note on scope is in order here. When we use the term 'law enforcement', we refer to domestic police services (that respond to crime that occurs within a contained jurisdiction), and for the purposes of this report, this term should be understood separately from government entities working either in national security, foreign intelligence or administrative policing bodies (such as those working in immigration).

4.1. AI and crime detection

Artificial intelligence is being used in the detection and investigation of criminal activity in countries around the world. We define crime detection as the act of attempting to ascertain whether or not certain crimes are being or have been committed. Crime detection in that context is past- or present-oriented, while crime forecasting, which we will discuss in more detail in section 4.2, is future-oriented.

4.1.1. The history of technology and crime detection

The use of technology to detect the occurrence of crimes that have happened—or that are occurring in real-time—in fact precedes the existence of 'artificially intelligent' technology. For

example, consider the use of such tools as video cameras, security systems that detect or monitor physical spaces, lie detector tests, radar detectors, and forensic analysis including technology with the capability for DNA analysis or any other physical or physiological trace, which can all be used to corroborate findings that a crime has been committed. Crime scenes themselves can be understood as physical spaces where a crime has occurred and where evidence of criminal activity can be found. Crime scenes may also refer to non-physical spaces where digital traces of criminal activity can be observed, collected, and analyzed to corroborate the finding that a crime occurred. Examples of traces of digital crime includes emails of phishing schemes aimed to defraud people, online forum discussions where people may sell or buy criminally obtained objects or discuss the details of their intent to commit a crime, an IP address associated to a machine trying to breach a computer system, or a pattern of usage for a mouse or a keypad. These examples are of course by no means exhaustive.

Crime scenes—whether physical or digital in nature—can now be defined as technology-rich environments,[161] even before we consider the ways in which artificial intelligence is being used to detect crimes. As forensic science expert Julie Mennell writes, there is bound to be "an abundance of technology" at crime scenes, including the following subtypes of technology that:

1. Seek to deter crime being committed and/or to alert that a crime might be about to take place, such as intruder alarms;

[161] Julie Mennell, "Technology Supporting Crime Detection: An Introduction" (2012) 45:12 Measurement + Control 304 at 304.

2. Capture a crime being committed, such as closed-circuit television systems;
3. Is contained with the crime scene itself and that may contain additional (digital) evidence that relates to the crime, victim or perpetrator, such as mobile phones or computers;
4. Is brought to the scene by the crime scene investigator (including forensic scientists), which can facilitate the discovery, recovery, recording, analysis, and transmission of evidence, such as digital cameras, laser scanners, lab on a chip (LOC) technology;
5. Assists in the identification of victims and perpetrators, such as fingerprint capture and recognition technology, and even automatic number plate recognition.[162]

With this information in mind, it is therefore clear that the use of AI in the detection of crime by law enforcement is not necessarily as technologically disruptive as it may at first seem to be. In other words, technology is already being used to detect the occurrence of crime. Artificial intelligence is simply a new addition to the repertoire of capabilities in the technologies used by law enforcement to determine when a crime may be happening or has already happened. AI merely creates new information processing and analytical capacities for other technologies that have become routine in law enforcement.

[162] Ibid.

4.1.2. A taxonomy of AI capabilities

It is possible to categorize the various types of artificial intelligence available to law enforcement for detection functions in terms of the capability of the software. The types of AI capabilities identified in the process of writing this report are as follows:

6. Object classification
7. Object recognition (including face recognition)
8. Speech recognition
9. Gunshot detection
10. DNA analysis
11. Digital forensics

In the following section, we describe each of these types of tools used for crime detection in terms of how they generally work, how they fall into the already-existing subtypes of crime detection technologies above, and their law enforcement use case scenarios.

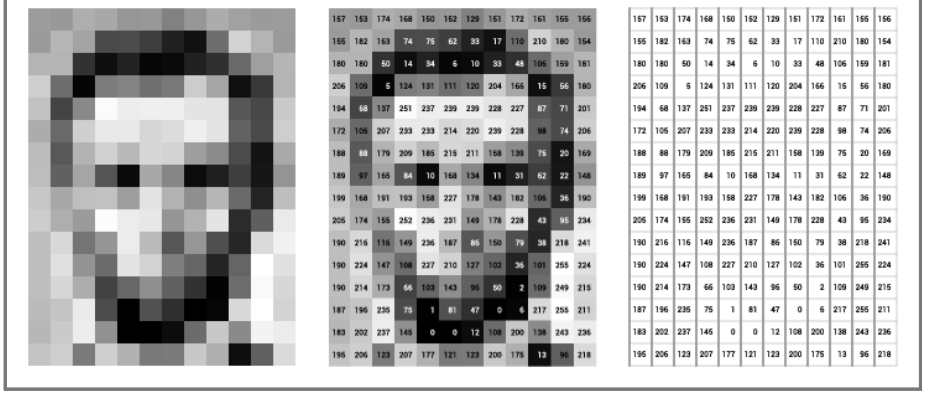
4.1.2.1. Object classification

Object classification software seeks to autonomously identify certain elements within images and videos, and label or categorize these elements much like humans do.[163] Object classification is a sub-domain within the field of computer vision, which can be understood as an application of artificial

[163] Nils J. Nilsson, The Quest for Artificial Intelligence (Cambridge, UK: Cambridge University Press, 2013) at 333.

intelligence. Systems that classify objects within imagery are able to work after researchers train a computer program or algorithmic model on a dataset of numerous images. Just as it occurs within machine learning more generally, elements within the imagery will be assembled into smaller parts such as pixels and groups of pixels, which will be labelled (often manually) on the basis of descriptors such as colour or textur e.[164] The program or model's learning process will then construct a decision tree that can classify the regions in the training set images as well as in future images. The program will subsequently be able to classify groups of pixels and therefore objects as part of the training categories.[165]

Figure 6 - Pixel Data Diagram of Abraham Lincoln[166]



164 Nilsson, ibid at 30; "Introduction to Computer Vision", Algorithmia Blog (2 April 2018), online https://blog.algorithmia.com/introduction-to-computer-vision/; Golan Levin, "Image Processing and Computer Vision", OpenFrameworks, online: https://openframeworks.cc/ofBook/chapters/image_processing_computer_vision.html.

165 Ibid.

166 Levin, ibid.

There are myriad reasons why law enforcement would want to—and do—use object classification in the detection of crimes. For example:

12. Law enforcement may gain access to an image of the commission of a crime and would seek to rely on machine learning to identify the location where an image was taken or recorded. Google's program called PlaNet does just this, and relies on convolutional neural networks for its geolocation capabilities;[167]

13. Police officers may also want to detect the possible existence of criminal activity depicted within an image. The image's contents may demonstrate the occurrence of a criminal act (e.g. the image depicts possible theft) and/or the existence of the image itself may constitute a crime (e.g. the image depicts child pornography).
    One well-known example of the latter is the PhotoDNA software developed by Microsoft and Hany Farid of Dartmouth College, which primarily aims to detect child pornography and works by a) creating a digital signature (known as a 'hash') associated with the image to prevent image alterations, and b) converts the image to black and white, resizes it, breaks into a grid, and quantifies its shading.[168] It then compares an image's hash against a

[167] "Google Unveils Neural Network with 'Superhuman' Ability to Determine the Location of Almost Any Image", MIT Technology Review (24 February 2016), online: https://www.technologyreview.com/s/600889/google-unveils-neural-network-with-superhuman-ability-to-determine-the-location-of-almost/.

[168] Jennifer Langston, "How PhotoDNA for Video is being used to fight online child exploitation", Microsoft On the Issues (12 September 2018), online:

database of images that have been identified as illegal, and matches can be manually reviewed by humans.[169] Microsoft claims that PhotoDNA cannot be used to recognize faces nor people or objects within the image.[170] PhotoDNA is used most notably by software giants such as Facebook,[171] Google,[172] Twitter,[173] and by the US-based National Center for Missing & Exploited Children.[174] Other examples of technology that seek to detect the commission of a crime within imagery include the European P-REACT Project,[175] the loss-prevention product offered by the US-based company StopLift,[176] and the Chinese software SenseTime;[177]

---

https://news.microsoft.com/on-the-issues/2018/09/12/how-photodna-for-video-is-being-used-to-fight-online-child-exploitation/.

169 Ibid.

170 Ibid.

171 "Meet the Safety Team", Facebook Safety (9 August 2011), online: https://www.facebook.com/notes/facebook-safety/meet-the-safety-team/248332788520844/

172 Rich McCormick, "Google scans everyone's email for child porn, and it just got a man arrested", The Verge (5 August 2014), online: https://www.theverge.com/2014/8/5/5970141/how-google-scans-your-gmail-for-child-porn.

173 Charles Arthur, "Twitter to introduce PhotoDNA system to block child abuse images", The Guardian (22 July 2013), online: https://www.theguardian.com/technology/2013/jul/22/twitter-photodna-child-abuse

174 "Partners", National Center for Missing & Exploited Kids (Webste), online: http://www.missingkids.org/supportus/partners.

175 Timothy Revell, "Computer vision algorithms pick out petty crime in CCTV footage", NewScientist (4 January 2017), online: https://www.newscientist.com/article/2116970-computer-vision-algorithms-pick-out-petty-crime-in-cctv-footage/.

176 "StopLift", Stoplift (Website), online: https://www.stoplift.com/.

177 "SenseTime: Our Company", SenseTime (Website), online: https://www.sensetime.com/ourCompany.

14. Law enforcement may use image recognition software to corroborate findings of criminal activity. For example, the Electronic Frontier Foundation found in 2016 that the Federal Bureau of Investigation in the US has invested in research that can identify and semantically analyze tattoos *en masse*, in order not only to "help law enforcement identify criminals and victims"[178] but also to map out people's relationships and identify their beliefs.[179] This is a task that can be clearly accomplished by human analysts, but such automation can introduce a new level of effectiveness to extract criminal intelligence from massive and publicly available data sets.

4.1.2.2. Object recognition (including face recognition)

Object recognition can be understood as a subset of computer vision. Rather than classify an element within imagery under a certain category, object recognition is focused on the identification of an individual instance within the imagery.[180] Examples include handwritten letters or digits, license plate numbers, specific vehicles, fingerprints, and a specific person's face. Object recognition works just as object classification does, but a key difference is that each object recognized can be

178 "Tattoo Recognition", FBI.gov, (25 June 2015), online: https://www.fbi.gov/audio-repository/news-podcasts-thisweek-tattoo-recognition.mp3/view.

179 Dave Maas, "FBI Wish List: An App That Can Recognize the Meaning of Your Tattoos", EFF Deep Links (16 July 2018), online: https://www.eff.org/deeplinks/2018/07/fbi-wants-app-can-recognize-meaning-your-tattoos

180 Moses Olafenwa, "Object Detection with 10 lines of code", Towards Data Science (16 June 2018), online: https://towardsdatascience.com/object-detection-with-10-lines-of-code-d6cb4d86f606.

uniquely identified as its own individual instance, rather than as a class of objects. For example, facial recognition software will work in various ways depending on the technology but it generally consists of: (i) the identification of key facial landmarks, such as the distance between a person's eyes and the distance from forehead to chin; (ii) the identification of these geometric measurements is turned into a facial signature or faceprint of sorts; which is then (iii) compared to a database of known faces; and finally (iv) matched with an image within the software's database.[181]

There are numerous examples of object recognition software that are used by law enforcement around the world. A notable example is Faception, the namesake of an Israel-based company and software that "can analyze faces from video streams (recorded and live), cameras, or online/offline databases, encode the faces in proprietary image descriptors and match an individual with various personality traits and types with a high level of accuracy."[182] The software has received criticism for allegedly facilitating what could amount to "facial-profiling" or profiling on the basis of one's biological characteristics[183] and for boldly claiming that it is able to classify a person as being endowed with a "high IQ", for being an "academic researcher", a "professional poker player", a "white-collar offender", "pedophile"

---

181 "How does facial recognition work?", Norton Security Center, online: https://us.norton.com/internetsecurity-iot-how-facial-recognition-software-works.html

182 "Faception", Faception (Website), online: https://www.faception.com/.

183 Gus Lubin, "'Facial-profiling' could be dangerously inaccurate and biased, experts warn", Business Insider (12 October 2016), online: https://www.businessinsider.com/does-faception-work-2016-10

and even a "terrorist."[184] A tool such as Faception clearly raises the possibility for discrimination, and at one point the company's website included computerized drawings of people who fell within their classifiers, and depicted stereotypical representations of white-collar offenders wearing sunglasses and terrorists with facial hair and head coverings that could be a reference to garments worn by some people who identify as Muslim:

Figure 7 - Faception's former depiction of criminals[185]

A country leading the way in its use of face recognition in public places is China. It seems as though almost every day the country

184 "Faception: Our technology", Faception Website, online: https://www.faception.com/our-technology.

185 Marcus Ranum, "It's Worse Than You Think: Robo-Profiling", Free Thought Blogs (16 March 2017), online: https://freethoughtblogs.com/stderr/2017/03/16/its-worse-than-you-think-robo-profiling/.

is the subject of a news piece documenting its use of technology that recognizes faces, whether related to smart locks,[186] the move towards a cashless society[187] or the use of face recognition in bathrooms.[188] With approximately 200 million cameras scattered throughout the country and with 400 million more coming online in 2020,[189] the country's partially AI-powered surveillance CCTV system is just one element in the government's bid for social control through its social credit system.[190] A thorough explanation of the numerous AI tools used by the Chinese government is outside the scope of this report. However, the appearance of these tools is worth noting as they may be considered as creating a supra-judicial system that conflates illegal with 'anti-social' behaviours and potentially automate the detection and sentencing of defined deviant behaviour at such a scale that would be difficult to oversee.

---

186 Meng Jing, "Chinese home sharing site Xiaozhu to roll out facial recognition-enabled smart locks in Chengdu pilot scheme", South China Morning Post (26 December 2018), online: https://www.scmp.com/tech/start-ups/article/2179495/chinese-home-sharing-site-xiaozhu-roll-out-facial-recognition-enabled.

187 "Facial recognition the future of cashless payment in China", Asia Times (20 December 2018), online: http://www.atimes.com/article/facial-recognition-the-future-of-cashless-payment-in-china/.

188 Masha Borak, "China's public toilets now have facial recognition, thanks to Xi Jinping", Tech in Asia (21 December 2018), online: https://www.techinasia.com/chinas-public-toilets-facial-recognition-xi-jinping.

189 Jon Russell, "China's CCTV surveillance network took just 7 minutes to capture BBC reporter", Tech Crunch (13 December 2017), online: https://techcrunch.com/2017/12/13/china-cctv-bbc-reporter/.

190 Megan Palin, "Big Brother: China's chilling dictatorship moves to introduce scorecards to control everyone", news.com.au (19 September 2018), online: https://www.news.com.au/technology/online/big-brother-chinas-chilling-dictatorship-moves-to-introduce-scorecards-to-control-everyone/news-story/6c821cbf15378ab0d3eeb3ec3dc98abf.

Notably, "Jaywalking, late payments on bills or taxes, buying too much alcohol or speaking out against the government, each cost citizens points" from their social credit score.[191] Having a higher score reaps benefits such as waived deposits on hotels and rental cars, VIP treatment at airports, discounted loans, priority job applications and fast-tracking to the most prestigious universities.[192] Punishments include losing the right to travel by plane or train, suspensions from social media and being excluded from government jobs.[193]

China's face recognition system was launched by the Ministry of Public Security in 2015 and is under development with a security company based in Shanghai.[194] The Chinese government has been framed by Forbes as seeking to build one of the world's largest face recognition databases in the world.[195] Otherwise, the planned scope and scale of the national project has been unclear. The overarching purpose of the country's facial recognition system seems to aim to identify people who have committed crimes or minor infractions (like jaywalking or stealing toilet paper) and to facilitate a hyper-efficient ease of economic transactions and daily-life interactions.[196] Already, the technology has been used

---

191 Ibid.Greene urgess (2018line: 2013), online:::Kids, online:irness. e judicial reasoning, whether e. Weresumed authority that r crime

192 Ibid.

193 Ibid.

194 Stephen Chen, "China to build giant facial recognition database to identify any citizen within seconds", South China Morning Post (12 October 2017), online: https://www.scmp.com/news/china/society/article/2115094/china-build-giant-facial-recognition-database-identify-any.

195 Bernard Marr, "The Fascinating Ways Facial Recognition AIs Are Used In China", Forbes (17 December 2018), online: https://www.forbes.com/sites/bernardmarr/2018/12/17/the-amazing-ways-facial-recognition-ais-are-used-in-china/#5842e21c5fa5.

to arrest wanted people who were attending a concert,[197] erroneously identify and publicly shame a person for jaywalking whose image had in fact appeared on an ad placed on the side of a moving bus,[198] and analyze the facial expressions of school children to see if they were paying attention in class.[199]

China's facial recognition framework has attracted significant scrutiny from North American and European policymakers and privacy advocates, who hold the technology to be an infringement of individual civil liberties or fundamental rights, and who fear that the norm of using such all-encompassing technology will spread to other jurisdictions.[200] Indeed, in

---

196 Ibid.

197 "China uses facial recognition to arrest fugitives", NHK World – Japan (26 December 2018), online: https://www3.nhk.or.jp/nhkworld/en/news/20181227_10/.

198 Ryan Daws, "Chinese facial recognition flags bus ad woman for jaywalking", IoT News (28 November 2018), online: https://www.iottechnews.com/news/2018/nov/28/chinese-facial-recognition-ad-jaywalking/.

199 Louise Moon, "Pay attention at the back: Chinese school installs facial recognition cameras to keep an eye on pupils "South China Morning Post (16 March 2018), online: https://www.scmp.com/news/china/society/article/2146387/pay-attention-back-chinese-school-installs-facial-recognition.

200 Casey Newton, "Microsoft sounds an alarm over facial recognition technology", The Verge (7 December 2018), online: https://www.theverge.com/2018/12/7/18129858/microsoft-facial-recognition-ai-now-google; Paul Mozur, "Inside China's Dystopian Dreams: A.I., Shame and Lots of Cameras", The New York Times (8 July 2018), online: https://www.nytimes.com/2018/07/08/business/china-surveillance-technology.html; Joyce Liu, "In Your Face: China's all-seeing state", BBC News (10 December 2017), online: https://www.bbc.com/news/av/world-asia-china-42248056/in-your-face-china-s-all-seeing-state; Simon Leplâtre, "En Chine, la reconnaissance faciale envahit le quotidian", Le Monde (9 December 2017), online: https://www.lemonde.fr/economie/article/2017/12/09/en-chine-la-re

December 2018 the U.S. Department of Homeland Security revealed that the Secret Service plans to test the use of facial recognition in and around the White House.[201] The use of AI for object and face recognition is a particularly powerful tool when combined with the omnipresent web of cameras that dot modern streetscapes[202] or that are increasingly added to the uniform of police officers as 'bodycams' to capture each police-citizen interaction.[203] However, this technology remains error-prone in operational environments where images are captured with a low resolution and where the elements and street furniture also degrade the quality of the data.

Furthermore, people's movements mean that their pictures are taken at a broad range of angles that degrade the quality of the analysis. The few real-life experiments that have been conducted in the UK and whose results have been disclosed publicly indicate extremely high levels of false positives: The South

---

connaissance-faciale-envahit-le-quotidien_5227160_3234.html; Thomas Assheuer, "Die Big-Data-Diktatur", Die Zeit (29 November 2017), online: https://www.zeit.de/2017/49/china-datenspeicherung-gesichtserkennung-big-data-ueberwachung.

201 Jay Stanley, "Secret Service Announces Test of Face Recognition System Around White House", ACLU Free Future (4 December 2018), online: https://www.aclu.org/blog/privacy-technology/surveillance-technologies/secret-service-announces-test-face-recognition.

202 David Barrett, "One surveillance camera for every 11 people in Britain, says CCTV survey", The Telegraph (10 July 2013), online: https://www.telegraph.co.uk/technology/10172298/One-surveillance-camera-for-every-11-people-in-Britain-says-CCTV-survey.html

203 Dan Greene & Genevieve Patterson, "The Trouble With Trusting AI to Interpret Police Body-Cam Video", IEEE Spectrum (21 November 2018), online: https://spectrum.ieee.org/computing/software/the-trouble-with-trusting-ai-to-interpret-police-bodycam-video

Wales Police trial of a facial recognition system conducted during the UEFA Champions League final in June 2017 generated 2,470 alerts for possible matches, 92% of which were incorrect. The London Metropolitan Police also tested a similar technology in 2016 and 2017 to manage a street carnival, with an error rate of 98% in the identification of possible suspects.[204] We examine other salient criticisms that apply to face recognition technology in the final subsection of this chapter.

4.1.2.3. Police body cameras

The decision for police to use AI-powered body cameras is another tool that promises benefits and poses various challenges. A leader in this industry is the U.S. company called Axon, formerly Taser International, also known for its Taser stun gun.[205] As a part of its decision to rebrand and to expand its business, Axon offered to provide free body cameras to any interested police department.[206] The company stated in June 2018 that it wanted to use AI to automate the police body camera video assessment and annotation process, and eventually help generate police reports from the recorded video of police-citizen encounters thanks to AI.[207] The purpose was to automate data

---

204 Matt Burgess, "AI is invading UK policing, but there is little proof it's useful", Wired (21 September 2018), online at https://www.wired.co.uk/article/police-artificial-intelligence-rusi-report.

205 reene urgess (2018line: 2013), online:::Kids, online:irness. e judicial reasoning, whether e. Weresumed authority that r crime Greene & Patterson, supra note 203.

206 Devin Coldewey, "Taser rebrands as Axon and offers free body cameras to any police department", Tech Crunch (5 April 2017), online: https://techcrunch.com/2017/04/05/taser-rebrands-as-axon-and-offers-free-body-cameras-to-any-police-department/.

207 Nancy Perry, "How Axon is accelerating tech advances in policing", Police One (Blog) (22 June 2018), online: https://www.policeone.com/police-products/body-cameras/articles

gathering and records management so that police officers can spend more time performing other tasks.[208] The company touted that more than 200,000 officers use their services, and that they have accumulated 30 petabytes of data ("10 times larger than the Netflix database")[209] that will be analyzed by its multifunctional AI system.[210] The company has also filed a patent for real-time face recognition in order to keep up with its competitors.[211]

In April 2018, Axon launched its AI and Policing Technology Ethics Board, made up of external efforts from various fields and with a hope to "provide expert guidance to Axon on the development of its AI products and services, paying particular attention to its impact on communities."[212] News articles state that the group is to meet twice a year to discuss the ethical implications of the company's products,[213] and the role of the board is to offer frank, honest advice.[214] It is not clear what, if any, impacts the board has had on the ethical development of Axon's products. But the decision to forge a path marked by a commitment to ethics

/476840006-How-Axon-is-accelerating-tech-advances-in-policing/

208 "TASER International's (TASR) CEO Rick Smith on Q4 2016 Results -Earnings Call Transcript", Seeking Alpha (28 February 2017), online: https://seekingalpha.com/article/4050796-taser-internationals-tasr-ceo-rick-smith-q4-2016-results-earnings-call-transcript?page=3.

209 Perry, supra note 207.

210 Greene & Patterson, supra note 203.

211 Ibid.

212 "Axon AI and Policing Technology Ethics Board", Axon (Website), online: https://ca.axon.com/info/ai-ethics.

213 James Vincent & Russell Brandom, "Axon launches AI ethics board to study the dangers of facial recognition", The Verge (26 April 2018), online: https://www.theverge.com/2018/4/26/17285034/axon-ai-ethics-board-facial-recognition-racial-bias.

214 "Axon AI and Policing Technology Ethics Board", supra note 52.

is laudable, and some have stated that they wished companies like Google (in light of its artificial and human intelligence lab called DeepMind) would follow suit and disclose who sits on the board, what the board discusses, and how often they meet.[215]

4.1.2.4. Speech recognition

Speech recognition is similar to object recognition in that the technology seeks to identify idiosyncratic elements of speech patterns, often with a view to identify the person speaking and to automatically transcribe the words being spoken. Regardless of the exact algorithms that can be used in this process, speech recognition software detects and measures sound waves and the frequency patterns of the speech signal.[216] Numerous obstacles must be overcome through this process, such as the existence of background noise and accounting for variations in the speed of speaking.[217] The software then classifies extracted blocks or sections of the speech using various—and at times multiple—techniques, such as statistical models or artificial neural networks.[218] The purpose is to classify small segments in terms of the type of sound that is made, and then classify larger segments of each sound to determine which word is being said.

---

215 Sam Shead, "Google's Mysterious AI Ethics Board Should Be Transparent Like Axon's", Forbes (27 April 2018), online: https://www.forbes.com/sites/samshead/2018/04/27/googles-mysterious-ai-ethics-board-should-be-as-transparent-as-axons/#12e80d0019d1.

216 Nitin Washan & Sandeep Sharma, "Speech Recognition System: A Review" 115:18 International Journal of Computer Applications 7, online: https://pdfs.semanticscholar.org/8f2c/b3f70bb75b6235514b192b83e413a0e23dd8.pdf.

217 Ibid.

218 Ibid.

One example of the operational use of voice recognition comes from Interpol—the International Criminal Police Organization. In mid-2018, it engaged in the final review of a project called the Speaker Identification Integrated Project.[219] The technology extends the capabilities of voice recognition software by taking collections of voice samples, analyzed for certain behavioral features, and creates 'voice prints' in order to match new voice data uploaded to its system (from police intercepts for example) to the voice data already on file for suspected criminals.[220] The technology can also filter voice samples by gender, age, language, and accent.[221] The Speaker Identification Integrated Project allows uploads and downloads of samples from 192 law enforcement agencies around the world.[222] The database will purportedly include samples not only from law enforcement but also "from YouTube, Facebook, publicly recorded conversations, voice-over-internet-protocol recordings, and other sources where individuals might not realize that their voices are being turned into biometric voice print."[223]

4.1.2.5. Gunshot detection

Gunshot detection software seeks to detect the occurrence of

---

219 Ava Kofman, "Interpol Rolls Out International Voice Identification Database Using Samples From 192 Law Enforcement Agencies", The Intercept (25 June 2018), online: https://theintercept.com/2018/06/25/interpol-voice-identification-database/

220 Ibid.

221 Michael Dumiak, "Interpol's New Software Will Recognize Criminals by Their Voices", IEEE Spectrum (16 May 2018), online: https://spectrum.ieee.org/tech-talk/consumer-electronics/audiovideo/interpols-new-automated-platform-will-recognize-criminals-by-their-voice.

222 Kofman, supra note 219.

223 Ibid.

gunfire and determine the precise location of the gunshot. Acoustic gunshot detection systems typically use a set of microphones distributed over large populated areas that detect and isolate the staccato sounds of gunfire, which can be then confirmed by humans who may notify law enforcement where the gunshot went off.[224] Gunshot detection can be understood as falling under the umbrella of AI because the designers of the software rely on machine learning in order to train their systems to identify the audio signature of gunfire and to isolate it from all the other sound interferences commonly found in urban settings.[225]

ShotSpotter is a US-based company that offers gunshot detection services to over 90 cities in the US[226] and has been approved for use in the major Canadian city of Toronto.[227] Law enforcement agencies have repeatedly justified their use of this software in public spaces to curb gun violence, especially in neighborhoods

---

224 Chris Weller, "There's a secret technology in 90 US cities that listens for gunfire 24/7", Business Insider (27 June 2017), online: https://www.businessinsider.com/how-shotspotter-works-microphones-detecting-gunshots-2017-6.

225 "Artificial intelligence-based system warns when a gun appears in a video", PhysOrg (Website) (7 July 2017), online: https://phys.org/news/2017-07-artificial-intelligence-based-gun-video.html

226 Matt Drange, "We're Spending Millions On This High-Tech System Designed To Reduce Gun Violence. Is It Making A Difference?", Forbes (17 November 2016), online: https://www.forbes.com/sites/mattdrange/2016/11/17/shotspotter-struggles-to-prove-impact-as-silicon-valley-answer-to-gun-violence/#11ee763731cb.

227 Jordan Pearson, "Toronto Approves Gunshot-Detecting Surveillance Tech Days After Mass Shooting", VICE Motherboard (25 July 2018), online: https://motherboard.vice.com/en_us/article/7xqk44/toronto-approves-shotspotter-gunshot-detecting-surveillance-tech-danforth-shooting.

where gunshots are common occurrences (and might not elicit calls to the police) or where citizens might feel intimidated and prefer to avoid interactions with the police.[228]

Another example of gunshot detection software—although it falls outside the scope of this report—is Boomerang III, a system developed by the US Department of Defense for use in the military. According to its description online, "Boomerang pinpoints the shooter's location of incoming small arms fire. Boomerang uses passive acoustic detection and computer-based signal processing to locate a shooter in less than a second."[229] Even if this technology has only been used in war environments so far, the trend of police militarization that has been observed in many Western democracies might lead to its rapid adoption by law enforcement agencies facing high homicide rates.[230]

4.1.2.6. DNA analysis

DNA analysis understood at its broadest consists of the application of genetic testing for crime-assessment and legal purposes.[231]

---

228 Jessica Patton, "What is 'ShotSpotter'? Controversial gunshot detector technology approved by Toronto police", Global News (20 July 2018), online: https://globalnews.ca/news/4344093/controversial-gunshot-detector-shotspotter-toronto-police/; "Chicago Signs $23 Million Multi-Year Agreement With Shotspotter to Extend Gunshot Detection Coverage Into Next Decade", ShotSpotter (Website) (5 September 2018), online: https://www.shotspotter.com/press-releases/chicago-signs-23-million-multi-year-agreement-with-shotspotter-to-extend-gunshot-detection-coverage-into-next-decade/.

229 "Boomerang III: State-of-the-Art Shooter Detection", Raytheon (Website), online: https://www.raytheon.com/capabilities/products/boomerang.

230 Peter B. Kraska, "Militarization and policing: Its relevance to 21st Century police", (2007) 1:4 Policing: A Journal of Policy and Practice 501.

The use of DNA as forensic material is a branch of forensic science that examines genetic material in criminal investigations. The most obvious reason law enforcement would want to collect and analyze genetic material at a crime scene concerns their desire to determine who was present when the alleged crime occurred, what their role may have been in the altercation, where the crime occurred and whether protagonists of the incident (either victim, witness or suspect) can be tied to previous solved or unsolved crimes.

Artificial intelligence plays a role in DNA analysis because of the new capacity it offers to significantly speed up the DNA sequence matching process, where collected DNA is matched with the DNA contained within a given database. Consider the decision on the part of police in California to use DNA data held by commercial genealogy websites in 2018.[232] In that instance, law enforcement found and arrested a person charged with numerous counts of rape and murder, and appear to have uploaded DNA data about the accused onto the website GEDMatch.[233] The DNA was obtained from a crime scene, and was purportedly used by the police to find one of his relatives.[234] It was not clear that the police had obtained authorization from the company to upload the accused's DNA and compare it with

---

231 "DNA Forensics: The application of genetic testing for legal purposes", GeneEd (Website), online: https://geneed.nlm.nih.gov/topic_subtopic.php?tid=37.

232 Antonio Regalado, "Investigators searched a million people's DNA to find Golden State serial killer", MIT Technology Review (27 April 2018), online: https://www.technologyreview.com/s/611038/investigators-searched-a-million-peoples-dna-to-find-golden-state-serial-killer/.

233 Ibid.

234 Ibid.

others on their website, and it is questionable whether DNA abandoned by the perpetrator of a crime is afforded constitutional protection in the US.[235] Cases like this call into question whether law enforcement should be required to obtain judicial authorization to upload the genetic material of perpetrators onto genealogy and DNA analysis website. Furthermore, it is not clear whether law enforcement should be able to rely on algorithms that are proprietary to private companies, and that are not free and open source and therefore escape technical and legal scrutiny. While there may be little legal protection over the privacy of DNA abandoned at a crime scene by a perpetrator, police organizations that may be interested in using AI-powered DNA matching and analysis tools ought to consider whether they are infringing upon the right to privacy of all other people whose DNA is stored in that database.

4.1.2.7. Digital forensics

Digital forensics, also called computer forensics, is the work of extracting and analyzing digital material found in electronic devices to turn it into evidence. There are numerous tools that comb through computers, mobile devices, and software looking for evidence of data that may be incriminating. Artificial intelligence is relevant here because it augments the capability of digital forensic analysis tools, which have generated massive quantities of data that no human being has the cognitive ability to process in reasonable amounts of time.

One key example is the software called Magnet AXIOM, made by Magnet Forensics based in Waterloo, Canada. The tool is called

235 Ibid.

"a digital investigations platform that allows examiners to acquire and examine relevant data from smartphones and computers, and visualize it for better analysis."[236] A core feature of the software is its use of Magnet.AI, which uses machine learning to conduct semantic or contextual content analysis of conversations on smartphones, computers, and chat applications.[237] The company claims that the tool has been optimized for cases of child exploitation, and seeks to categorize and flag language in conversations that could constitute child luring.[238] The company specifically highlights that this tool will alter how police conduct their interviews and engage in arrest proceedings.[239]

4.2. AI for crime prediction and prevention

Artificial intelligence is also being developed with the aim to predict and prevent crime, and not merely just to detect what has occurred or is unfolding. Interestingly, the use of technology to predict the future occurrence of crimes is not new. Consider the use of violence risk assessment tools in criminal justice and forensic psychiatry. One study demonstrated that there are over 200 tools available in numerous jurisdictions to inform initial sentencing, parole, and decisions regarding post-release monitoring and rehabilitation, but even in 2017 there were very little

---

236 Amira Zubairi, "Magnet Forensics launches Magnet.AI to fight child exploitation", Betakit (Website) (16 May 2017), online: https://betakit.com/magnet-forensics-launches-magnet-ai-to-fight-child-exploitation/.

237 Ibid.

238 "Introducing Magnet.AI: Putting Machine Learning to Work for Forensics", Magnet Forensics (Website), online: https://www.magnetforensics.com/blog/introducing-magnet-ai-putting-machine-learning-work-forensics/

239 Ibid.

relevant, reliable and unbiased data that could demonstrate the predictive accuracy of such forensic psychiatry data.[240]

Many of these tools are still in development and could be seen as consisting of vaporware—technology that makes promises, but are not mature enough to be launched commercially. One major company offering services in this field is PredPol, a US-based company that "grew out of a research project between the Los Angeles Police Department and UCLA."[241] The company claims to be a "Market Leader in Predictive Policing" and seeks to identify the times and locations where specific crimes are most likely to occur so that these areas can be patrolled to prevent those crimes from occurring.[242] The company states that it has patented its algorithm, which is based on the statistical analysis of three aspects of offender behavior: 1) Repeat victimization (in short, the company assumes that where a crime has occurred, it is more likely that another crime will occur soon after), 2) Near-repeat victimization (which assumes that crimes occur in proximity to one another), 3) Local search (which again assumes that crimes tend to cluster together). This algorithm is partly inspired by the statistical models that are being used to predict earthquake aftershocks.[243]

---

240 T. Douglas, J. Pugh, I. Singh, J. Savulescu, and S. Fazelb, "Risk assessment tools in criminal justice and forensic psychiatry: The need for better data" (2017) 42 Eur Psychiatry 134, online: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5408162/.

241 "Overview", PredPol (Website), online: https://www.predpol.com/about/.

242 Ibid.

243 Alexander Babuta, Marion Oswald, & Christine Rinik, "Machine learning algorithms and police decision-making: Legal, ethical and regulatory challenges" (2018) Whitehall Reports (21 September), at 5, online: https://rusi.org/publication/whitehall-reports/machine-learning-

The technology differs from what has been developed in other US cities, such as in Chicago where a Strategic Subject List seeks to algorithmically or probabilistically determine who is most likely to be a perpetrator or victim involved in future shootings.[244] PredPol does not assess who is likely to commit a crime, but nonetheless has been criticized for its use of machine learning, the Los Angeles Police Department's criminal data, and an outdated gang territory map to automate the process of classifying "gang-related" crimes.[245] This combination could create a feedback loop in which certain neighbourhoods or groups of people are labelled as criminal.[246] Additionally, in an article published in a French journal, the original designer of the seismographic algorithm that influenced the PredPol algorithm was asked to test the applicability of his model to crime data from Chicago and seriously challenged the transferability of this tool to the prediction of crime patterns. The output generated by this kind of approach does not seem much more effective than traditional hotspot maps at forecasting the location of future crimes.[247]

---

algorithms-and-police-decision-making-legal-ethical.

244 Jeff Asher & Rob Arthur, "Inside the Algorithm That Tries to Predict Gun Violence in Chicago", The New York Times (13 June 2017), online: https://www.nytimes.com/2017/06/13/upshot/what-an-algorithm-reveals-about-life-on-chicagos-high-risk-list.html

245 Ali Winston & Ingrid Burrington, "A pioneer in predictive policing is starting a troubling new project", The Verge (26 April 2018), online: https://www.theverge.com/2018/4/26/17285058/predictive-policing-predpol-pentagon-ai-racial-bias.

246 Randy Rieland, "Artificial Intelligence Is Now Used to Predict Crime. But Is It Biased?", Smithsonian Magazine (5 March 2018), online: https://www.smithsonianmag.com/innovation/artificial-intelligence-is-now-used-predict-crime-is-it-biased-180968337/.

247 Bilel Benbouzid, "À qui profite le crime? Le marché de la prediction du crime aux États-Unis" (2016) La Vie des Idées, online at https://laviedesidees.fr/A-qui-profite-le-crime.html.

As demonstrated by the work of companies like PredPol, there is a growing and largely unregulated market for software that seeks to assist law enforcement agencies with the prediction of criminal acts. Police organizations seeking to deploy tools that forecast the commission of crimes should proceed with caution and seek to obtain as much information as possible about the accuracy of any tool they wish to use, prior to expending resources on them.

4.3. Conclusion: Gaps in literature and ethical concerns

This section has sought to shed light on the use of AI by law enforcement. We demonstrate that there is a historical analogy for assistive technology in crime analysis. There is a usefulness of recognizing the taxonomy of current tactics in use, such as deploying software that analyzes sounds, objects, faces, and DNA, as well as the act of uncovering of digital traces of crimes found within technological devices themselves. There also remains an emerging and unregulated market of tools that use machine learning to predict crime 'hotspots', and other elements of criminal activity.

Yet there remain numerous unanswered questions about the impact that artificial intelligence may have on law enforcement's response to crime. We draw on the work of mathematician Hannah and Fry and researchers Alexander Babuta, Marion Oswald and Christine Rinik to guide the decision of any law enforcement agency considering the adoption of AI. We also raise numerous issues that demonstrate the need to carefully assess any overbroad use of AI, showing that there are instances where the benefits of AI seem to outweigh its costs, and conversely, the use of AI in other situations in fact pose significant challenges

that have yet to be overcome. We encourage law enforcement and governments to identify and explicitly state the priorities underpinning the use of AI. There are numerous solutions available to governments whose law enforcement agencies wish to use AI, which require proactive regulation imbued with a commitment to minimum standards regarding transparency, systems of oversight, and interdisciplinary collaboration.

4.3.1. Mapping out the issues of AI in law enforcement

British mathematician Hannah Fry, an expert on computer science and human behavior, has identified numerous ethical issues raised by the use of AI for crime analysis. Her expertise holds that artificially intelligent algorithms are bound to make mistakes, and that there are times when they *will* be unfair.[248] For all the positive impacts that AI may have on the criminal justice system, there will invariably be endless examples of unfairness engendered by algorithms. Consider the fact that the Strategic Subject List was initially intended to help victims of gun crime but was ultimately used by police as a 'hit list' to pursue gun violence offenders.[249] By recognizing the inevitable imperfection and replication of unconscious bias of its designers, we diminish the assumption that an algorithmic tool has innate, dispassionate authority.

There are numerous ways in which algorithmic tools can be unnecessary or unfair when deployed by law enforcement, which calls into question how and when AI should be used. Arguments in favor of AI for law enforcement may presume there is a causal link between the use of AI and decreased crime

248 Hannah Fry, Hello World: Being Human in the Age of the Machine (New York, NY: W.W. Norton, 2018) at 330-332.
249 Ibid at 331.

rates. Indeed, cities like Kent (UK) and Los Angeles as well as Alhambra (California, USA) observed a reduction in crime in certain city regions after running through a trial of PredPol, which correlated with police officers being dispatched to those certain areas right after crimes had occurred.[250] But as Fry observes, it is difficult to know whether technology should take credit for the detection or forecasting of crimes. PredPol would certainly want to take credit for any such crime reductions, yet dispatching police officers to certain geographic areas — with or without the use of algorithms — could be the causal factor in reducing crime in those city regions.[251]

The use of AI by law enforcement may engender confirmation bias of police officers looking for crime, which may in fact alter crime rates. According to Toby Davies, mathematician and crime scientist, police will detect more crime when they are in a certain place than they would have done otherwise.[252] In other words, if an equal amount of crime is happening in two places, the police will detect more crime in the place they were, rather than in the other place, where they were not. The result could be a feedback loop, where an algorithm predicts that more crime will happen in, for example, a poor neighbourhood. Police officers would be sent to that neighbourhood, where they detect crime. As the algorithm keeps predicting that the neighbourhood is a crime hotspot, more police officers are sent there, and more crime is detected in those areas. As mentioned above, feedback loops like this occur when AI systems are not challenged for the confirmation or unconscious bias that is

250 Ibid at 260-262.
251 Ibid at 262.
252 Ibid at 262.

bound to characterize their design and are likely to be a problem for people who are already in precarious economic or immigration positions. Another cause for concern is the fact that much crime detection and forecasting technology, and the algorithms within them, is proprietary. For experts like Fry let alone the average judge or person, it is not clear how technology like PredPol works. Without having the ability to assess how algorithms come to their findings, people accused of crimes could be deprived of procedural fairness or the right to due process.[253] We revisit this issue in our concluding chapter.

When it comes to facial recognition, a major issue concerns the possibility for false identification. Mathematically, false identifications are bound to happen with AI systems, with potential severe consequence for those wrongfully identified as a suspect. By contrast, forensic DNA analysis is based on the assessment of the highly variable genetic information of individuals and the probability that DNA sequences will match, and the chance of mismatch is lowered when a larger sample size and larger number of genetic markers is used.[254] Yet the

---

253 Danielle Keats Citron, "Technological Due Process" (2008), 85:6 Washington University Law Review 1249, online: https://openscholarship.wustl.edu/cgi/viewcontent.cgi?article=1166&context=law_lawreview; Ellora Israni, "Algorithmic Due Process: Mistaken Accountability and Attribution in State v. Loomis" Jolt Digest (31 August 2017), online: https://jolt.law.harvard.edu/digest/algorithmic-due-process-mistaken-accountability-and-attribution-in-state-v-loomis-1, "Taking Algorithms To Court Current Strategies for Litigating Government Use of Algorithmic Decision-Making", AI Now Institute (24 September 2018), online: https://medium.com/@AINowInstitute/taking-algorithms-to-court-7b90f82ffcc9; Frank Pasquale, "Secret Algorithms Threaten the Rule of Law", MIT Technology Review (1 June 2017), online: https://www.technologyreview.com/s/608011/secret-algorithms-threaten-the-rule-of-law/.

inverse is currently true for face recognition technology. Google's FaceNet scored an accuracy rate of 99.6 when tasked with identifying five thousand images of celebrities' faces,[255] but when it took part in the University of Washington's 'Megaface challenge' in 2015, it managed only a 75 percent identification rate.[256] This is because the chances of facial misidentification multiply dramatically when there are more faces to compare to (given current technical capabilities). The more faces the algorithm searches through, the greater the chance of it finding two faces that look similar. In the words of Fry, "similarity is in the eye of the beholder." [257] With this in mind, "facial recognition, as a method of identification, is not like DNA, which sits proudly on a robust statistical platform." [258] Further, face recognition technology can be fooled by twins,[259] siblings,[260] masks,[261] and specifically-designed fake eyeglass frames.[262]

---

254 R. Chakraborty, "Sample size requirements for addressing the population genetic issues of forensic use of DNA typing" (1992) 64:2 Human Biology 141; Sankar Subramanian, "The effects of sample size on population genomic analyses – implications for the tests of neutrality" (2016) 17:123 BMC Genomics.

255 Florian Schroff, Dmitry Kalenichenko & James Philbin, "FaceNet: A Unified Embedding for Face Recognition and Clustering", arXiv Working Paper, arXiv:1503.03832v3 [cs.CV], online: https://arxiv.org/pdf/1503.03832.pdf

256 "FGNet Results", MegaFace (Website), online: https://megaface.cs.washington.edu/results/fgnetresults.html.

257 Fry, supra note 248 at 277.

258 Fry, supra note 248 at 275.

259 Emmanuel Ocbazghi, "We put the iPhone X's Face ID to the ultimate test with identical twins — and the results surprised us" Business Insider (31 October 2017), online: https://www.businessinsider.com/can-iphone-x-tell-difference-between-twins-face-id-recognition-apple-2017-10.

260 Alex Hern, "Apple: don't use Face ID on an iPhone X if you're under 13 or have a twin", The Guardian (27 September 2017), online: https://www.theguardian.com/technology/2017/sep/27/apple-face-id-iphone-x-under-13-twin-facial-recognition-system-more-secure-touch-id.

That said, there some situations where the benefits of using face recognition technology outweigh the above costs. One example is the decision of the Canadian province of Ontario to use face recognition technology for people with gambling addictions, and who have voluntarily placed themselves on a self-exclusion list, allowing themselves to be recognized by algorithms upon entering a casino and politely asked to leave the building.[263]

Figure 8 - Actress Reese Witherspoon impersonating Russell Crowe using eyeglass frames[264]

Difficult trade-offs lay before governments and law enforcement organisations that want to implement algorithmic crime analysis

---

261 Thomas Brewster, "Apple Face ID 'Fooled Again' -- This Time By $200 Evil Twin Mask", Forbes (27 Novembere 2017), online: https://www.forbes.com/sites/thomasbrewster/2017/11/27/apple-face-id-artificial-intelligence-twin-mask-attacks-iphone-x/#7df1a8052775.

262 Mahmood Sharif, Sruti Bhagavatula, Lujo Bauer & Michael K. Reiter, "Accessorize to a Crime: Real and Stealthy Attacks on State-of-the-Art Face Recognition", Conference Paper (October 2016), online: https://www.cs.cmu.edu/~sbhagava/papers/face-rec-ccs16.pdf

263 "Self-Exclusion Program", Ontario Lottery and Gaming Corporation (Website), online: https://about.olg.ca/self-exclusion/facial-recognition/.

264 Eva Rinaldi, "Reese Witherspoon", Flickr (Website), online: https://goo.gl/a2sCdc; Eva Rinaldi, "Russell Crowe", Flickr (Website), online: https://goo.gl/AO7QYu.

tools: they must carefully balance the competing interests of individual privacy versus protection of the public as a whole, in order to ensure the fairness of algorithmic decision-making. Returning to the guiding map of issues identified by Fry, any entity that decides to utilize algorithms as a regulatory tool for crime must ultimate decide what its priorities are. "Is it keeping crime as low as possible? Or preserving the freedom of the innocent above all else? How much of one would you sacrifice for the sake of the other?"[265] And Fry reminds us: "Gary Marx, professor of sociology at MIT, put the dilemma well in an interview he gave to the Guardian: 'The Soviet Union had remarkably little street crime when they were at their worst of their totalitarian, authoritarian controls. But, my God, at what price?"[266] We, like Fry and numerous other experts, advise for careful implementation and explicit limitations on the reach of algorithmic decision-making in the detection and prediction of crime. The decision of a law enforcement agency to use AI and *how* the agency uses it is an important part of the regulatory ecosystem that enables or limits the power of this technology.

4.3.2. Ways forward

There are numerous solutions available to governments whose law enforcement agencies wish to use AI, which require proactive regulation imbued with a commitment to minimum standards regarding transparency, systems of oversight, and interdisciplinary collaboration. Law enforcement agencies

---

[265] Fry, supra note 248 at 290.

[266] Fry, supra note 248 at 290, citing Nate Berg, 'Predicting crime, LAPD-style', The Guardian (25 June 2014), online: https://www.theguardian.com/cities/2014/jun/25/predicting-crime-lapd-los-angeles-police-data-analysis-algorithm-minority-report.

exploring the use of AI, and the government behind these agencies have a growing corpus of work at their disposal on the best practices for algorithmic decision-making. We turn in particular to the findings of researchers Alexander Babuta, Marion Oswald and Christine Rinik, writing for the UK defence and security think tank Royal United Services Institute.[267] They sought to examine the legal, ethical and regulatory challenges posed by the deployment of machine learning algorithms for police decisions. [268] Their findings are worth enumerating here and apply to all law enforcement agencies regardless of jurisdiction: [269]

15. Clear guidance and codes of practice that constrain how law enforcement should trial and use algorithmic tools should be developed as a matter of urgency;

16. A regulatory framework is needed to establish minimum standards for the use of algorithmic tools by police forces, especially related to relevant data protection legislation, transparency and intelligibility of the AI system, and respect for human rights and administrative law principles;

17. Retroactive deconstruction of the algorithm should be a required element of all public procurement agreements, in order to assess the factors that influenced the model's predictions;

---

267 Banline: (Corporation (Website), online:Kids, online:irness. e judicial reasoning, whether e. Weresumed authority that r crimeBabuta, supra note 243.

268 Ibid.

269 (August 2018), online:ebsite) (4 December 2018), one:irness. e judicial reasoning, whether e. Weresumed authority that r crime Ibid.

18. A formalized system of scrutiny and oversight (be it a commission, taskforce, committee, board, etc.) is necessary to ensure adherence to this regulatory framework. These ethics boards should be multidisciplinary, and consist of a combination of practitioners, technical experts, academics, and perhaps average or lay people;

19. The approach should be collaborative and cut across disciplines to ensure representation from various experts and stakeholders. This board should provide recommendations to individual law enforcement agencies for practice, strategy and policy decisions relating to the use of algorithms.

There are several other general guiding principles on AI for further reading, including the "Principles for Accountable Algorithms and a Social Impact Statement for Algorithms" put together by members of the Fairness, Accountability, and Transparency in Machine Learning (FAT/ML) community.[270] AI Now Institute's latest report and publications, which were published in 2018,[271] a recent report published by Data & Society Research Institute,[272]

---

270 "Principles for Accountable Algorithms and a Social Impact Statement for Algorithms", FAT/ML (Website), online: http://www.fatml.org/resources/principles-for-accountable-algorithms.

271 Meredith Whittaker et al., "AI Now Report", AI Now Institute (December 2018), online: https://ainowinstitute.org/AI_Now_2018_Report.pdf; AI Now Institute, "After a Year of Tech Scandals, Our 10 Recommendations for AI", AI Now Institute (6 December 2018), online: https://medium.com/@AINowInstitute/after-a-year-of-tech-scandals-our-10-recommendations-for-ai-95b3b2c5e5.

272 Mark Latonero, "Governing Artificial Intelligence: Upholding Human Rights & Dignity", Data & Society Research Institute (10 October 2018), online:

and the Montréal Declaration for Responsible Development of Artificial Intelligence that was most recently updated in December 2018 constitute a few examples of high level documents outlining these principles, which we revisit in our concluding chapter.[273]

In short, given the high stakes issues at risk in the use of AI by law enforcement (privacy, presumption of innocence, freedom from punishment) we advocate for what some scholars have called a human-rights-by-design approach to technology,[274] where algorithms are designed such that the human (rather than the presumed authority of the machine) is considered first at every stage of design, deployment, and iterative improvement[275] when law enforcement is involved.

---

https://datasociety.net/output/governing-artificial-intelligence/.

273 "Official Launch Of The Montréal Declaration For Responsible Development Of Artificial Intelligence", Montréal Declaration for Responsible Development of Artificial Intelligence (Website) (4 December 2018), online: https://www.declarationmontreal-iaresponsable.com/blogue/d%C3%A9voilement-de-la-d%C3%A9claration-de-montr%C3%A9al-pour-un-d%C3%A9veloppement-responsable-de-l-ia.

274 Jon Penney et al., "Advancing Human-Rights-By-Design In The Dual-Use Technology Industry", Columbia Journal of International Affairs (August 2018), online: https://jia.sipa.columbia.edu/advancing-human-rights-design-dual-use-technology-industry.

275 Fry, supra note 248 at 333.

Table 1 - AI Software in use for Law Enforcement and Criminal Justice Services

Name of software	Creator	Purpose & capabilities	Capability of technology	Institution(s) using it
PhotoDNA[276]	Microsoft; Hany Farid	"PhotoDNA creates a unique digital signature (known as a "hash") of an image which is then compared against signatures (hashes) of other photos to find copies of the same image. When matched with a database containing hashes of previously identified illegal images, PhotoDNA is an incredible tool to help detect, disrupt and report the distribution of child exploitation material. PhotoDNA is not facial recognition software and cannot be used to identify a person or object in an image. A PhotoDNA hash is not reversible, and therefore cannot be used to recreate an image."	Video and image recognition	National Center for Missing and Exploited Children, Europol, Interpol
Optimizing the Use of Video Technology	University of Central Florida, National Institute of Justice	"A $1.3 million grant from the National Institute of Justice is funding a new two-year project that may revolutionize the way police monitor and analyze	Video and image recognition	University of Central Florida, Orlando Police Department,

Name of software	Creator	Purpose & capabilities	Capability of technology	Institution(s) using it
to Improve Criminal Justice Outcomes[277]		crime scene surveillance video footage with technology developed at the University of Central Florida. For the first time, UCF computer scientists will develop and test computer vision technology that will automate the process of monitoring and reviewing thousands of hours of video streams fed-in from multiple cameras."		Florida State Attorneys and Public Defenders
Live Facial Recognition [278]	Unknown	"Live Facial Recognition (LFR), is technology that can identify a person from a digital image. The technology is being used [by the Metropolitan Police force in London, UK] to assist in the prevention and detection of crime by identifying wanted criminals."	Facial recognition	The Metropolitan Police Force (London, UK)
Predictive Policing Research: Shreveport [279]	Shreveport Police Department	"The Shreveport Police Department received an award to implement a pilot that will use a randomized experimental design using experimental and control groups involving six of the	Crime forecasting	Shreveport Police Department (US)

Name of software	Creator	Purpose & capabilities	Capability of technology	Institution(s) using it
		highest-crime policing districts in Shreveport. The pilot will evaluate the "broken windows" theory of policing in an operational setting and employ a predictive model using leading indicators related to that theory such as juvenile complaints, loud music, disorderly persons, suspicious activity, loitering, disputes and prowlers. RAND will measure the efficacy of the pilot in terms of its ability to reduce tactical crimes such as shootings, robbery, burglary, auto break-ins, outside residential thefts, outside business thefts and auto thefts."		
Predictive Policing Research: Chicago[280]	Chicago Police Department; Illinois Institute of Technology	"CPD received an award to implement a pilot that in part will evaluate the ability of modified pattern-matching software currently used for medical diagnostic purposes to predict crime patterns. It will also evaluate the efficacy of a software tool that	Crime forecasting	Chicago Police Department (US)

Name of software	Creator	Purpose & capabilities	Capability of technology	Institution(s) using it
		quantifies and maps gang activity to predict emerging areas of gang conflict. RAND will evaluate the pilot in terms of accuracy of prediction, process and impact, using randomized, retrospective and quasi-experimental studies. CPD's research partner is the Illinois Institute of Technology (IIT)."		
TopMatch-3D Portable Scanner[281]	Cadre Research Labs	"Cadre's new portable scanner is designed to complement the TopMatch-3D High-Capacity desktop scanner for high-volume labs or those where in-field measurements are required. The Portable scanner runs off a laptop and provides 3D topographies in just 5 seconds. The Portable system can be used during triage / pre-screening to sort cartridge cases prior to full analysis or at a crime scene to provide immediate actionable intelligence regarding the number and possible makes of firearms involved."	Forensic firearm analysis	National Chief of Justice (US) commissioned research

Name of software	Creator	Purpose & capabilities	Capability of technology	Institution(s) using it
Horizon or Anti-Crime [282]	Proposed by a laboratory of the Mines-Telecom Institute, produced in partnership with Morpho, a subsidiary of the Safran group	AI powered by various data, such as INSEE, weather, geography, data of criminal interest, and even extractions of blogs or social networks (e.g. Facebook, Twitter), to aid police decisions.	Crime forecasting	France
PredPol[283]	Californian start-up led by academics	"A program called PredPol was created eight years ago by UCLA scientists working with the Los Angeles Police Department, with the goal of seeing how scientific analysis of crime data could help spot patterns of criminal behavior. Now used by more than 60 police departments around the country, PredPol identifies areas in a neighborhood where serious crimes are more likely to occur during a particular period."	Crime forecasting; Recidivism prediction	Kent Constabulary (UK) US police services (the company claims on its website it helps protect one out of 33 people in the US)

Name of software	Creator	Purpose & capabilities	Capability of technology	Institution(s) using it
WARRANT (automated warrant service triage tool)[284]	Research Triangle Institute	"Researchers at the Research Triangle Institute, in partnership with the Durham Police Department and the Anne Arundel Sheriff's Department, are working to create an automated warrant service triage tool for the North Carolina Statewide Warrant Repository. The NIJ-supported team is using algorithms to analyze data sets with more than 340,000 warrant records. The algorithms form decision trees and perform survival analysis to determine the time span until the next occurrence of an event of interest and predict the risk of re-offending for absconding offenders (if a warrant goes unserved). This model will help practitioners triage warrant service when backlogs exist. The resulting tool will also be geographically referenced so that practitioners can pursue concentrations of high-risk absconders — along with others who have active warrants — to optimize resources."	Crime forecasting; Recidivism prediction	North Carolina Statewide Warrant Repository (US)

Name of software	Creator	Purpose & capabilities	Capability of technology	Institution(s) using it
VALCRI (Visual Analytics for Sense-Making in Criminal Intelligence Analysis) [285]	Middlesex University London, funded by the European Commission	"At the cutting edge of intelligence-led policing, VALCRI is a semi-automated analysis system that helps find connections humans often miss. When pre-empting crime or investigating a case, it can be deployed by analysts to reconstruct situations, generate insights and discover leads. Through autonomous work or collaboration with a human team, VALCRI creatively analyses data from a wide range of mixed-format sources. It displays its findings with easy-to-digest visualisations, comes up with possible explanations of crimes, and paves the way for rigorous arguments. Protecting against human error and bias, VALCRI works with objective intelligence, speed and precision."	Crime forecasting; Data collection; Data visualization; Recreation of criminal events through spatial-tempora l constructions	Various
Analytics Enterprise[286]	Cellebrite	"Transform raw, disparate data into digital intelligence faster with Analytics Enterprise – a unified investigation tool	Facial recognition; Digital	UK forces: Metropolitan Police and

Name of software	Creator	Purpose & capabilities	Capability of technology	Institution(s) using it
Analytics Desktop[287]		that enables examiners, investigators and prosecutors to access, analyze and collaborate in real time on all centrally located forensic artifacts. Eliminate manual, time-intensive tasks through automatic data correlation and management. Focus on data from drone, mobile, cloud, computer and telco sources into a single view from a centralized digital forensics library. Perform complex analytics more simply and discover critical evidence hidden within text, images, videos and more. Streamline workflows and collaborate across the lab and the broader investigative team to accelerate investigations and reduce case cycle times." "Designed as a standalone application, Analytics Desktop eliminates time-intensive analysis, turning raw data into digital intelligence so you reveal	forensics; Video and image recognition; Face detection; Optical character recognition;	others unlisted online US federal (FBI, Secret Service) and municipal agencies Police forces in Turkey, the UAE, Bahrain, etc.

Name of software	Creator	Purpose & capabilities	Capability of technology	Institution(s) using it
		more leads in less time and shorten investigation cycle times. Be confident that you are delivering accurate, defensible intelligence that can expedite and successfully bolster a case when you search, filter and dig into digital artifacts across disparate sources."		
HART (Harm Assessment Risk Tool)[288]	University of Cambridge in collaboration with Durham Constabulary	"It has been developed to aid decision-making by custody officers when assessing the risk of future offending and to enable those arrestees forecast as moderate risk to be eligible for the Constabulary's Checkpoint programme. Checkpoint is an intervention currently being tested in the Constabulary and is an 'out of court disposal' (a way of dealing with an offence not requiring prosecution in court) aimed at reducing future offending."	Recidivism prediction	Durham Constabulary (UK)
Palantir Intelligence[289]	Palantir	"Palantir Intelligence integrates disparate data from disconnected data	Crime forecasting	Formerly New Orleans Police

Name of software	Creator	Purpose & capabilities	Capability of technology	Institution(s) using it
		silos at massive scale for low-friction interaction with intelligence officers. Search through every shred of enterprise data at high speed, pull out significant intelligence, and perform intuitive, multi-dimensional analysis to reveal unseen patterns, connections, and trends. Enterprise data sources, unstructured cable traffic, structured identity data, email, telephone records, spreadsheets, network traffic and more can all be searched and analyzed without the need for a specialized query language. Intelligence officers can rapidly turn mountains of data into actionable insight by asking the questions they need answered, in a language they understand."		Department (NOPD)
Faception[290]	Faception	"Faception is first-to-technology and first-to-market with proprietary computer vision and machine learning technology for profiling people and revealing their	Facial recognition	Unknown

Name of software	Creator	Purpose & capabilities	Capability of technology	Institution(s) using it
		personality based only on their facial image. … Faception can analyze faces from video streams (recorded and live), cameras, or online/offline databases, encode the faces in proprietary image descriptors and match an individual with various personality traits and types with a high level of accuracy. We develop proprietary classifiers, each describing a certain personality type or trait such as an Extrovert, a person with High IQ, Professional Poker Player or a threat. Ultimately, we can score facial images on a set of classifiers and provide our clients with a better understanding of their customers, the people in front of them or in front of their cameras."		
Entrupy[291]	Entrupy	"Entrupy has developed algorithms that allow it to analyze various materials ranging from canvas and leather to metal and wood. The device, which looks like a handheld scanner, takes	Image recognition; Verification of authenticity of consumer	Unknown

Name of software	Creator	Purpose & capabilities	Capability of technology	Institution(s) using it
		microscopic photographs of different areas of an item and runs them through a computer. Entrupy is accurate more than 98 percent of the time and returns results in less than 30 seconds."	products	
Axon Body 2, 3[292]	Axon	"Police officers wearing new cameras by Axon, the U.S.'s largest body camera supplier, will soon be able to send live video from their cameras back to base and elsewhere, potentially enhancing officers' situational awareness and expanding police surveillance."	Gunshot detection; Potential for facial recognition	Various police department (US and Canada)
Identify[293]	Veritone (US)	"Developed specifically for public safety and judicial agencies, these cloud-based applications enable agencies to rapidly extract actionable intelligence from video evidence used in investigations and the criminal justice process, increasing the speed and efficiency of their investigative and disclosure workflows. A natural extension to the aiWARE application	Motion detection; Video and image recognition; Facial recognition; Speech recognition; Geolocation;	Complete Discovery Source (eDiscovery company)

Name of software	Creator	Purpose & capabilities	Capability of technology	Institution(s) using it
		suite, agencies are now empowered to not only intelligently search to find pertinent evidence but identify suspects and redact sensitive materials within that evidence prior to distribution. The new suite includes Veritone IDentify, which leverages arrest records and person of interest databases to identify potential suspects quickly, and Veritone Redact, which swiftly redacts sensitive, personally identifiable or compromising information from video or photographic evidence."	Redaction of personally identifying information	
Internet Evidence Finder (IEF) and Magnet.AI[294]	Formerly Magnet Forensics, now a part of Deloitte	"Magnet AXIOM 1.1 features Magnet.AI, a contextual content analysis tool that uses machine learning to search through conversations on smartphones, computers, and chat apps. Magnet Forensics said the tool is specifically designed to help investigators tackle child exploitation cases, which often involve "luring," a process where a child predator gains his or her victim's trust."	Digital forensics	Unknown

Name of software	Creator	Purpose & capabilities	Capability of technology	Institution(s) using it
Traffic Jam[295]	Marinus Analytics (formerly Carnegie Mellon Robotics)	"Traffic Jam's FaceSearch gives detectives the tools to find specific missing [and trafficked] persons, and apprehend their exploiters more effectively. It uses the latest advancements in artificial intelligence, machine learning, computer vision, predictive modeling, and geospatial analysis—turning big data into actionable intelligence."	Digital forensics; Facial recognition	Unknown
Artificial Intelligence-Based Human Efface Detection[296]	Staqu Technologies	"Staqu has developed a proprietary AI technology stack which is comprised of advanced image analysis, language and text independent proprietary speaker identification engine, facial recognition and text processing, including name entity recognition, sentiment analysis and summarisation APIs which can be used by various other firms as well. ... The company worked with Rajasthan Police to introduce an AI-enabled app that helps law enforcement agencies by offering tools	Crime forecasting; video and image recognition; facial recognition; speech recognition	Rajasthan Police, Punjab (India), Dubai

Name of software	Creator	Purpose & capabilities	Capability of technology	Institution(s) using it
		like criminal identity registration, and tracking and missing persons' search."		

276 "PhotoDNA", Microsoft (Website), online: https://www.microsoft.com/en-us/photodna.

277 Zenaida Kotala, "Orlando Crime Scene Video Analysis Goes High-Tech With $1.3 Million Grant to UCF" Space Coast Daily,19 April 2016), online) http://spacecoastdaily.com/2016/04/orlando-crime-scene-video-analysis-goes-high-tech-with-1-3-million-grant-to-ucf/.

278 "Life Facial Recognition Trial", Metropolitan Police (Website), online: https://www.met.police.uk/live-facial-recognition-trial/.

279 "Predictive Policing Research", National Institute of Justice (Website), online: https://www.nij.gov/topics/law-enforcement/strategies/predictive-policing/Pages/research.aspx.

280 Ibid.

281 "The Future of Firearm Forensics is 3D", Cadre Forensics (Website), online: https://www.cadreforensics.com/.

282 Camille Polloni, "Police prédictive : la tentation de « dire quel sera le crime de demain", L'Obs (27 May 2015), online : https://www.nouvelobs.com/rue89/rue89-police-justice/20150527.RUE9213/police-predictive-la-tentation-de-dire-quel-sera-le-crime-de-demain.html.

283 Rieland, supra note 246.

284 "Predictive Policing Research", supra note 280.

285 "VALCRI", VALCRI (Website), online: http://www.valcri.org/.

286 "Analytics Enterprise", Cellebrite (Website), online: https://www.cellebrite.com/en/products/analytics-enterprise/.

287 "Analytics Desktop", Cellebrite (Website), online: https://www.cellebrite.com/en/products/analytics-desktop/.

[288] Marion Oswald, et al, "Algorithmic risk assessment policing models: lessons from the Durham HART model and 'Experimental' proportionality" (2018) 27:2 Information & Communications Technology Law 223, online : https://www.tandfonline.com/doi/pdf/10.1080/13600834.2018.1458455.

[289] "Intelligence", Palantir (Website), online: https://www.palantir.com/solutions/intelligence/.

[290] "Faception : Facial Personality Analytics", Faception (Website), online : https://www.faception.com/.

[291] "8 Companies Using AI for Law Enforcement", Nanalyze (Website), online: https://www.nanalyze.com/2017/11/8-companies-ai-law-enforcement/.

[292] Alex Pasternack, "Body camera maker will let cops live-stream their encounters", Fast Company (10 August 2018), online: https://www.fastcompany.com/90247228/axon-new-body-cameras-will-live-stream-police-encounters.

[293] "Veritone Announces New AI-Powered Law Enforcement Application Suite to Collectively Expedite Investigations and Evidence Disclosure", Business Wire (20 September 2018), online: https://www.businesswire.com/news/home/20180920005160/en/Veritone%C2%AE-Announces-New-AI-Powered-Law-Enforcement-Application.

[294] "Magnet Forensics Launches Magnet.AI to Fight Child Exploitation", StartUp Toronto (Website), online: https://startupheretoronto.com/sectors/technology/magnet-forensics-launches-magnet-ai-to-fight-child-exploitation/.

[295] "Pittsburgh-based tech company debuts first facial recognition technology designed to halt global human trafficking", Marinus Analytics (Website), online: http://www.marinusanalytics.com/articles/2017/6/27/face-search-debut.

[296] Kul Bhushan, "Meet Staqu, a startup helping Indian law enforcement agencies with advanced AI", Live Mint (26 June 2018), online: https://www.livemint.com/AI/DIh6fmR6croUJps6x7JW5K/Meet-Staqu-a-startup-helping-Indian-law-enforcement-agencie.html.

5. ARTIFICIAL INTELLIGENCE IN CRIMINAL PROCEEDINGS

Courts in various jurisdictions around the world are coming to incorporate artificial intelligence in their decision-making processes. Our research identifies a few areas in which AI has already come to be used in criminal proceedings: namely, risk assessment decisions in bail and sentencing hearings. We have found an emerging supply of technology that is strategically marketed as AI or that functions as AI. This technology generally assesses a level of risk associated with a person charged with a crime, or an incarcerated person who has been found guilty of committing a crime. Once again, it would be wise for decision makers in all jurisdictions to employ such risk assessment tools with much carefulness and forethought in light of their potential negative impact on basic principles of criminal justice such as the right to a presumption of innocence, the necessity of procedural fairness, and the necessity for decisions to be made without discrimination.

5.1. How AI is already being used in criminal proceedings

There are numerous instances of judicial systems that already employ artificial intelligence tools in criminal proceedings. Thus far, our findings demonstrate that the AI currently in use assesses the risk of future unwanted behavior on the part of an *accused person* — rather than examining other possible places of risk, such as examining the likelihood that a judge

or jury will respond in a certain given way depending on the facts before them. Judicial systems, particularly within the United States, have come to primarily rely on artificial intelligence in the context of decisions that relate to bail and first appearance in court (if applicable in that jurisdiction), as well as sentencing.

5.1.1. The use of AI in bail decisions

A description of what we mean by bail is useful here. Bail, also called pre-trial detention, can be understood as a pre-emptive safeguard used by courts to ensure that an accused person complies with criminal justice proceedings. The notion is rooted in the fear that a person, once charged with a crime, may miss their court hearings or may continue to commit crimes or cause harm. The concept of bail or pre-trial detention exists in numerous countries around the world. Some jurisdictions such as certain states in the U.S. use a bail system that allows for the accused to provide numerous types of collateral as a condition to being released from their pre-trial detention, such as cash, surety bonds that rely on a third party, the pledging of property, promises not to engage in illegal conduct, restraining orders, a combination of the above, and others not listed here.[297]

One method currently used to make bail decisions involves the use of 'bail schedules.' Bail schedules are a way that many jurisdictions have sought to streamline the process of bail determinations: they are a list of the set amounts that an

[297] "What is Bail? Understanding What Bail is & Different Types of Bal Bonds", Bail USA (Website), online: http://www.bailusa.net/what-is-bail.php.

accused person is required to pay. They are based on the nature of the offense that the accused is charged with.[298] For example, the bail schedule for the state of Alabama delineates the recommended range of bail amounts that judges should require based on the severity and classification of the charge as follows:

Figure 9 - State of Alabama bail schedule[299]

BAIL SCHEDULE

Recommended Range

Felonies:			
Capital felony	$50,000	to	No Bail Allowed
Murder	$15,000	to	$ 150,000
Class A felony	$10,000	to	$ 60,000
Class B felony	$ 5,000	to	$ 30,000
Class C felony	$ 2,500	to	$ 15,000
Drug manufacturing and trafficking	$ 5,000	to	$1,500,000
Class D felony	$1,000	to	$ 10,000
Misdemeanors (not included elsewhere in the schedule):			
Class A misdemeanor	$ 300	to	$ 6,000
Class B misdemeanor	$ 300*	to	$ 3,000
Class C misdemeanor	$ 300	to	$ 1,000
Violation	$ 300	to	$ 500
Municipal Ordinance Violations	$ 300	to	$ 1,000
Traffic-Related Offenses:			
DUI	$ 1,000	to	$ 7,500

As of 2018, numerous states within the U.S. have enacted reforms to their cash bail and bail schedule systems, and in their place, some of these states have begun implementing laws that require the use of risk assessment tools to influence bail

298 "Bail Schedule Law and Legal Definition", USLegal (Website), online: https://definitions.uslegal.com/b/bail-schedule/ at p. 2. Website), online: z, ":December 2018), one:irness. e judicial reasoning, whether e. Weresumed authority that r crime.

299 Alabama Rules of Criminal Procedure, Rule 7.2(b), online: http://judicial.alabama.gov/docs/library/rules/cr7_2.pdf.

decisions. New Jersey and California are two examples of states that have shifted from cash bail and fixed bail schedules towards risk assessment systems.[300] In the case of California, the underlying principle of the recent change in 2018 "is that a suspect will be evaluated on the basis of risk to public safety and the likelihood of not appearing in court, rather than on his or her ability to post a certain bail amount." [301] The hope is that rather than pay a certain amount of cash as a form of collateral to convince the court that an accused person will appear at their trial, judges will instead make pre-trial detention or release decisions based in part on empirical systems that determine whether a person is likely to flee or commit another alleged crime.

New Jersey is one of the states with the most experience thus far with using an automated risk assessment tool based on statistics and algorithms. The state uses the Public Safety Assessment (PSA), a pre-trial risk assessment tool developed by the Laura and John Arnold Foundation. This Foundation hopes to improve the criminal justice system in the U.S. For example, the Foundation has stated that its team created the PSA only after partnering with "leading criminal justice researchers" to determine where there was greatest need for improvement in the criminal justice system and that statistical risk assessment

---

300 Thomas Fuller, "California Is the First State to Scrap Cash Bail", The New York Times (28 August 2018), online: https://www.nytimes.com/2018/08/28/us/california-cash-bail.html; Rebecca Ibarra, "New Jersey's Bail Reform Law Gets Court Victory", WNYC (9 July 2018), online: https://www.wnyc.org/story/new-jerseys-bail-reform-law-gets-court-victory/.

301 Fuller, ibid.

tools were a viable solution to limit over-incarceration and the over-spending of taxpayer money associated with the pre-trial phase.[302] The Foundation initially piloted its use of the PSA in certain counties within Kentucky, North Carolina, California, and Arizona.[303] As of April 2018, the Foundation states that around 40 jurisdictions have launched or are in the process of implementing the PSA, [304] which demonstrates the staggering reach of this model for assessing risk in the pre-trial phase in the U.S. and possibly beyond.

5.1.2. New Jersey's Public Safety Assessment Tool

How does the Public Safety Assessment tool actually work? As is described in documents released by the state of New Jersey, the PSA uses nine risk factors to determine the likelihood that an accused person would engage in (i) new criminal activity or (b) violent criminal activity in the time before their trial, or (c) the likelihood of their failure to appear to their trial.[305] The nine factors, including any explanatory information, are listed below, with an answer of "yes" ostensibly increasing the likelihood of unwanted risk associated with the accused person:

---

302 "RFP: National Provider of Training & Technical Assistance", Arnold Foundation (Website), online: https://www.arnoldfoundation.org/wp-content/uploads/PSA-National-Provider-RFP.pdf at 5.

303 Ibid at 5.

304 Ibid at 5.

305 Ibid at 1.

Table 2 - New Jersey Public Safety Assessment Tool[306]

Factor	Explanation	Possible answers
1. Age at current arrest	The PSA calculates age based on the accused person or defendant's age in years at the time of the current arrest.	"Age is used to determine if the defendant is 20 or younger, 21 or 22, or 23 or older."
2. Current violent offense	The PSA categorizes an offense as violent when "a person causes or attempts to cause physical injury through use of force or violence against another person," with more caveats as described in the text of New Jersey court risk factor documentation.	"If any of the current offenses are violent, the answer to this risk factor is yes. Otherwise, the answer Is no."
2a. Current violent offense & 20 years old or younger	The PSA takes into consideration whether or not someone who was 20 or younger committed a violent crime.	"If one or more of the current offenses is violent as defined in risk factor 2 above and the defendant is 20 or younger at the time of the arrest as defined in risk factor 1 above, the answer to this risk factor is yes. Otherwise, the answer is no."
3. Pending charge at the time of the offense	The PSA assesses whether or not the accused person is already facing any other charge, which it defines in the context of New Jersey as "is a charge that has a future	"If the defendant had an Indictable or Disorderly Persons charge pending at the time the current offense allegedly occurred, the answer

[306] Ibid at 1-4.

Factor	Explanation	Possible answers
	pre-disposition related court date or is pending presentation to the grand jury, or has not been disposed of due to the defendant's failure to appear pending trial or sentencing, or that is in some form of deferred status (e.g., conditional discharge, conditional dismissal, pretrial intervention program)."	to this risk factor is yes. Otherwise, the answer is no."
4. Prior Disorderly Persons conviction	It also assesses whether the accused in the context of New Jersey has been charged with disorderly conduct, or any other misdemeanor under the laws of another state.	"If the defendant has pled guilty or been found guilty as an adult of one or more Disorderly Persons or misdemeanor offenses and the charge is not in deferred status or pending sentencing, the answer to this risk factor is yes. Otherwise, the answer is no."
5. Prior indictable conviction	The PSA assesses whether the accused person has been convicted of an indictable or felony offense, with some caveats.	"If the defendant has pled guilty or been found guilty as an adult of one or more Indictable or felony offenses and the charge is not in deferred status or pending sentencing, the answer to this risk factor is yes. Otherwise, the answer is no."

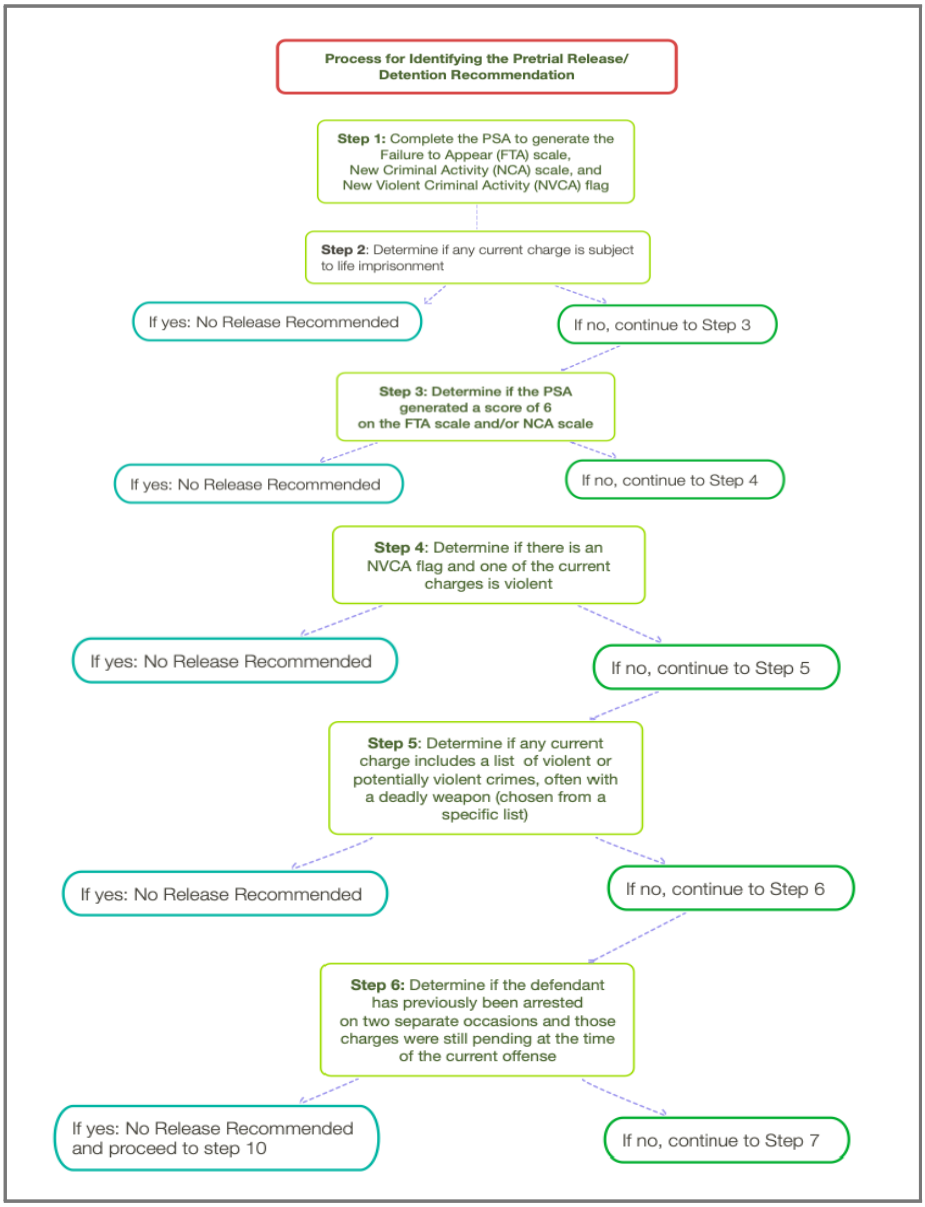
Factor	Explanation	Possible answers
5a. Prior conviction	It also assesses whether the accused person has been convicted of a "Disorderly Persons conviction" as defined in risk factor 4 or has a prior indictable conviction as defined in risk factor 5.	"If the defendant has a prior Disorderly Persons conviction as defined in risk factor 4 above or the defendant has a prior Indictable conviction as defined in risk factor 5 above, the answer to this risk factor is yes. Otherwise, the answer is no."
6. Prior violent conviction	The PSA takes into consideration whether the accused person has been convicted of a violent crime.	"The number of guilty dispositions for a prior violent charge is used to determine if the defendant has none, 1 or 2, or 3 or more prior violent convictions."
7. Prior failure to appear pre-trial in past 2 years	It also examines whether the accused person has failed to appear for a court hearing and the court acted by issuing a particular notice or a warrant for arrest, as per specific conditions and in the past 2 years.	"The number of failures to appear pretrial in the past two years is used to determine if the defendant had none, 1, or 2 or more prior failures to appear."
8. Prior failure to appear pre-trial older than 2 years	It also examines whether the accused person has failed to appear for a court hearing and the court acted by issuing a particular notice or a warrant for arrest, as per specific conditions and more than two years from the date of the current arrest.	"If the defendant failed to appear for court pretrial and an FTA notice/bench warrant for arrest was issued more than two years from the date of the current arrest, the answer to this risk factor is yes. Otherwise, the answer is no."

Factor	Explanation	Possible answers
9. Prior sentence to incarceration	The final risk factor considers whether the accused person was sentenced to incarceration, which it defines as including "any sentence to jail or prison of 14 days or more for an Indictable or Disorderly Persons offense imposed by a judge at the time of sentencing or re-sentencing", with other particular caveats or conditions.	"If the defendant previously received a sentence of incarceration to jail or prison of 14 days or more as a single sentence imposed by a judge, the answer to this risk factor is yes. Otherwise, the answer is no."

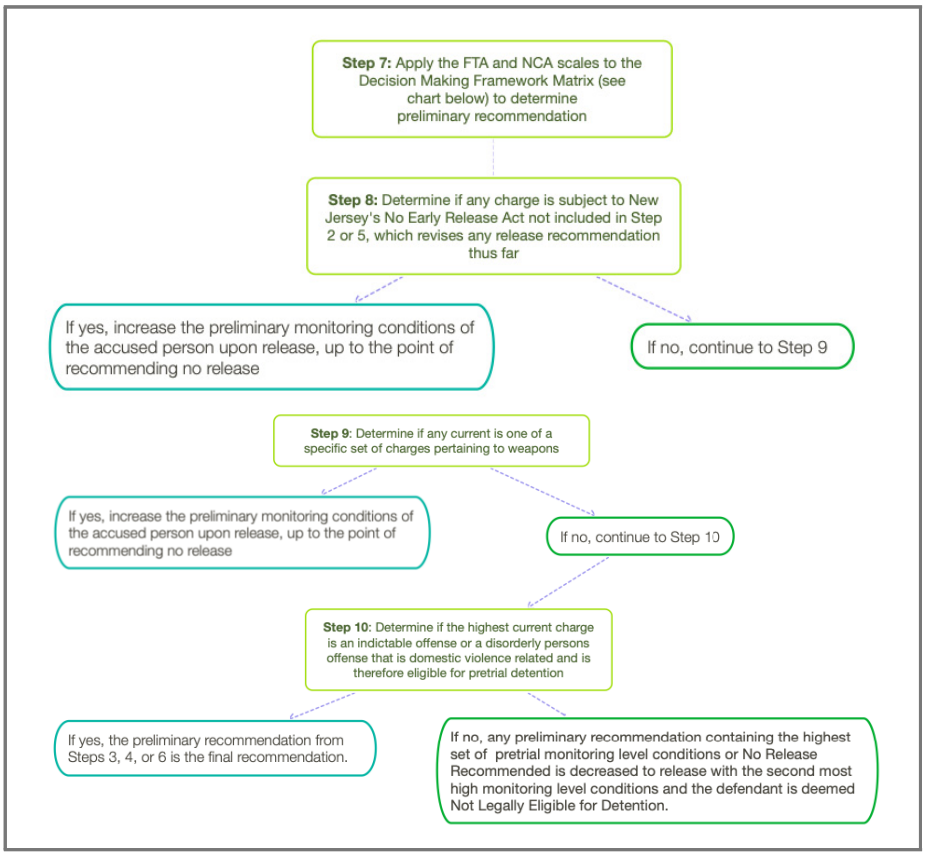
The Laura and John Arnold Foundation states that the above risk factors "are weighted and converted to separate FTA [failure to appear] and new criminal activity scales that range from 1 to 6, and a new violent criminal activity flag (i.e., binary indicator of yes/no)."[307] As one report indicates, the framework assumes that if the person has been charged with a violent offense, they are "flagged" to judges as having "a high potential for violence, and this case should be reviewed more carefully before making the release decision,"[308] a logical inference which presumes a link between what is alleged to have happened before and what may happen again. New Jersey and other states using the PSA then use the above risk factors to nudge a judge to release a person on bail using the 10-step process below, which we have depicted as a visualization based on New Jersey's Pretrial Release Recommendation Decision Making Framework dated March 2018.

307 Ibid at 12.
308 Ibid at 13.

Figure 10 - New Jersey's Pretrial Release Recommendation Decision Making Framework dated March 2018[309]



309 "Pretrial Release Recommendation Decision Making Framework (DMF)", New Jersey Courts (March 2018), online: https://www.njcourts.gov/courts/assets/criminal/decmakframwork.pdf?cacheID=JOvH2H8.



There are several things worth noting in the above table and visualization. It is highly commendable that the state of New Jersey has decided to release these documents to the public. It is also important to note that there are exclusion criteria as to what information is used in this actuarial determination, such as juvenile records, domestic violence restraining orders, "Petty Disorderly Persons" offenses, and local ordinance or municipal by-law offenses.[310] The PSA tool is also framed such that it

310 "Public Safety Assessment New Jersey Risk Factor Definitions - March 2018", New Jersey Courts, online: https://www.njcourts.gov/courts/assets/criminal/psariskfactor.pdf?cacheID=IDYJVkr.

seeks, at least on paper, to offer only recommendations to judges in their pre-trial detention or release decisions. It is also important to note that there is already precedent in U.S. law for making bail determinations on the basis of several of the factors listed above. Consider the Alabama Rules of Criminal Procedure, which state that the accused has the presumptive right to release on recognizance or on bond.[311] In order for a judge to impose any other conditions on the accused person, they "may" take into account circumstances such as the following:

1. "The age, background and family ties, relationships and circumstances of the defendant;
2. The defendant's prior criminal record, including prior releases on recognizance or on secured appearance bonds, and other pending cases;
3. Violence or lack of violence in the alleged commission of the offense."[312]

There are nonetheless several ethical issues that legislators and policymakers ought to consider in the use of algorithmic tools such as this one in decisions involving pre-trial detention or release. We lay them out below, after discussing a similar AI-powered tool used in the U.S. for sentencing decisions.

5.1.3. The use of AI in sentencing

Statistical and actuarial tools are also increasingly used by courts in sentencing decisions. Sentencing refers to a judge's

311 Alabama Rules of Criminal Procedure, supra note 3.
312 Ibid.

decision as to how a person, once convicted, ought to be punished. Regardless of the jurisdiction, sentences can range from anywhere between paying a small fine to spending a lifetime in jail, and in some jurisdictions, the application of a death sentence. Jurisdictions will vary in how they determine sentencing decisions, but they tend to be based on factors such as the severity or classification of the crime committed, whether the person has already been convicted of crime before, and any pre-existing guidelines where certain offenses have been determined by policymakers as deserving of specific punishments. Numerous jurisdictions already employ the use of reports such as pre-sentence report or victim impact statements, which offer judges information as they decide how to sentence an individual.

5.1.4. The use of COMPAS in sentencing decisions

A prominent example of AI in sentencing that recently came to the fore concerned the software called COMPAS (Correctional Offender Management Profiling for Alternative Sanctions). COMPAS received significant attention after being featured by news media in 2016, with reporters claiming that the software was imbued with unacknowledged bias particularly against black people and other people of color in the U.S.[313] The software later received a renewed wave of attention in 2017 after the U.S. Supreme Court refused to hear an appeal by a Wisconsin man, who had been sentenced to six years in prison after a judge consulted the results of COMPAS's risk assessment. The company behind COMPAS is Equivant (formerly known as

313 Julia Angwin, Jeff Larson, Surya Mattu & Lauren Kirchner, "Machine Bias", ProPublica (23 May 2016), online: https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing.

Northpointe). The company has locations in numerous states around the U.S. The former Northpointe website used to state: "As a criminal justice professional, you are responsible for implementing policies and practices that reduce recidivism and guard public safety. We provide, scientifically validated assessment tools, significant practical experience and technical knowledge necessary to help you accomplish your goals."[314] COMPAS was one risk assessment tool offered by Northpointe that sought to reduce the rate of re-offending and to "guard public safety."[315]

COMPAS was developed in 1998, and Northpointe's developers introduced the recidivism risk assessment component in 2000.[316] Among other things, the software specifically seeks to predict an accused person's risk of committing another crime within two years of assessment, based on 137 questions answered by the accused person or information obtained from their criminal record.[317] Reporters who wrote about COMPAS in 2016 were able to obtain information about its data through a public records request.[318] Below is a snapshot of one part of the questionnaire, with particular attention paid to whether or not the person conducting the interview with the accused believes him or her to be "a suspected or admitted gang member".

---

314 "Northpointe Suite: Automated Decision Support", Northpointe (Website, via Internet Archive), online: https://web.archive.org/web/20160307002839/http://www.northpointeinc.com/.

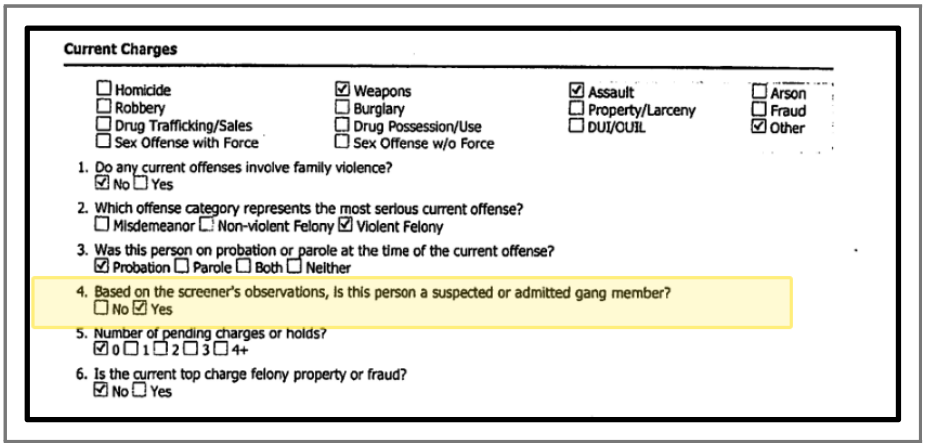
315 Ibid.

316 Julia Dressel & Hany Farid, "The accuracy, fairness, and limits of predicting recidivism" Science Advances (17 January 2018), online: http://advances.sciencemag.org/content/4/1/eaao5580.full.

317 Ibid.

318 Angwin, supra note 319.

Figure 11 - Snapshot of the questions used in COMPAS' determination[319]



Current Charges

☐ Homicide ☐ Robbery ☐ Drug Trafficking/Sales ☐ Sex Offense with Force
☑ Weapons ☐ Burglary ☐ Drug Possession/Use ☐ Sex Offense w/o Force
☑ Assault ☐ Property/Larceny ☐ DUI/OUIL
☐ Arson ☐ Fraud ☑ Other

1. Do any current offenses involve family violence?
☑ No ☐ Yes
2. Which offense category represents the most serious current offense?
☐ Misdemeanor ☐ Non-violent Felony ☑ Violent Felony
3. Was this person on probation or parole at the time of the current offense?
☑ Probation ☐ Parole ☐ Both ☐ Neither
4. Based on the screener's observations, is this person a suspected or admitted gang member?
☐ No ☑ Yes
5. Number of pending charges or holds?
☑ 0 ☐ 1 ☐ 2 ☐ 3 ☐ 4+
6. Is the current top charge felony property or fraud?
☑ No ☐ Yes

It is not fully clear how exactly COMPAS makes its predictions. Equivant has decided not to publicly release how its algorithm comes to a decision, though it has at times explained the theoretical rationale that underpins its work.[320] What we do know is that the COMPAS software relies on approximately 137 features in the aforementioned questionnaire to derive predictions regarding (i) a person's risk of being charged with the same crime within two years of assessment, (ii) failure to appear before court, or (iii) the probability that the person will commit a violent crime.

---

319 "Sample-COMPAS-Risk-Assessment-COMPAS-'CORE'", DocumentCloud (Hosting service), online: https://www.documentcloud.org/documents/2702103-Sample-Risk-Assessment-COMPAS-CORE.html

320 "Practitioner's Guide to COMPAS Software", Northpointe (Website, via Internet Archive), online: https://web.archive.org/web/20160507022911/http://www.northpointeinc.com/downloads/compas/Practitioners-Guide-COMPAS-Core-_031915.pdf; "Evidence-Based Practice Implementing the COMPAS Assessment System", Northpointe (Website, via Internet Archive), online: https://web.archive.org/web/20160506140944/http://www.northpointeinc.com/downloads/whitepapers/EVIDENCE-BASED_PRACTICE-implementing_COMPAS.pdf.

We also know that in response to claims that COMPAS is racially biased against black people, Equivant has attempted to prove that the overall predictive accuracy of its software across all ethnicities is an average of 68%, and claims that this meets the purported threshold for reliability in criminological studies of 70% and higher.[321] It has also stated that COMPAS is just one tool that could make up one part of decisions made within the context of criminal justice, and therefore warrants interpretation of the results it offers.[322]

A study published by scholars Julia Dressel and Hany Farid in January 2018 sought to assess the accuracy of COMPAS, and in so doing demonstrated that the software is actually accurate an average of 65% of the time.[323] This study also demonstrated that the recidivism predictions of COMPAS were no more accurate than predictions made by people with little or no criminal justice expertise or simple statistical analysis based on two features.[324] The study by Dressler and Farid confirms the finding Equivant attempted to debunk, namely ProPublica's conclusion that COMPAS's overall recidivism predictions were accurate an average of 65.1% of the time.[325]

---

321 William Dietrich, Christina Mendoza & Tim Brennan "COMPAS Risk Scales: Demonstrating Accuracy Equity and Predictive Parity", Volaris Groupe (Website), online: http://go.volarisgroup.com/rs/430-MBX-989/images/ProPublica_Commentary_Final_070616.pdf at 3.

322 "Practitioner's Guide to COMPAS Software", supra note 326 at 7.

323 Dressel & Farid, supra note 322.

324 Ibid.

325 Jeff Larson, Surya Mattu, Lauren Kirchner & Julia Angwin, "How We Analyzed the COMPAS Recidivism Algorithm", ProPublica (23 May 2016), online: https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm.

Figure 12 - An image from Northpointe's previous website[326]

Outside of the U.S. context, legal researchers in jurisdictions such as Australia have also begun exploring the use of AI in in sentencing, with the hopes of making the process of sentencing not only "more transparent and quicker" but also "fairer" and "more accurate."[327]

5.2. Gaps in literature and ethical concerns

There are numerous research questions that have yet to be explored when it comes to the use of automated risk assessment tools in the criminal justice system. Policymakers ought to be aware of these understudied areas, which give rise to ethical concerns pervading the use of tools that seek to "guard" public safety. First of all, it is unclear whether risk assessment tools

326 "COMPAS", Northpointe (Website, via Internet Archive), online: https://web.archive.org/web/20160315175056/http://www.northpointeinc.com:80/risk-needs-assessment

327 "Artificial intelligence to enhance Australian judiciary system", Swineburn University of Technology (Blog) (29 January 2018), online: http://www.swinburne.edu.au/news/latest-news/2018/01/artificial-intelligence-to-enhance-australian-judiciary-system.php.

exceed the average accuracy of judges who hold criminal justice expertise. Second, it is unclear whether jurisdictions outside of the United States ought to use these tools unless their bail and sentencing procedures also exhibit demonstrable deficits. Third, and finally, assuming that these tools may still be used in some jurisdictions, we draw on the work of Kelly Hannah-Moffat, whose analysis of recidivism and risk assessment tools in criminal proceedings makes clear that there are at least three primary concerns in such ventures: (i) accuracy and transparency of information and procedures, (ii) impact on dispositions and disparity, and (iii) the need to incorporate risk into any guidelines that govern the use of such technology, in particular elements of criminal proceedings.[328] The following subsections outline just some of the questions that should be asked by policymakers seeking to procure automated risk assessment software.

5.2.1. Is there evidence that these tools are more accurate than systems already in place? Is there evidence that the use of AI in legal proceedings will fulfill its promises?

First, the study by Dressler and Farid from 2018 lends credible weight to the argument that it is not clear whether the growing (and often unregulated) market for risk assessment software is actually meeting a need within the criminal justice system, and whether such software can in fact fulfill its promises. More

---

328 Kelly Hannah-Moffat, "Actuarial Sentencing: An 'Unsettled' Proposition" (2013) 30:2 Justice Quarterly 270-296, DOI: 10.1080/07418825.2012.682603; Mark H. Bergstrom & Richard P. Kern, "A View from the Field: Practitioners' Response to Actuarial Sentencing: An 'Unsettled' Proposition" (2013) 25:3 Federal Sentencing Reporter 185at 4.a note 32ne: Insight (onrd.logy (Blog), 18), one:irness. e judicial reasoning, whether e. Weresumed authority that r crime.

research could be useful to determine the accuracy of judges' decisions: for example, how often does a judge's decision not to detain someone correlate with the accused being absent from their trial? How often does a judge's detention or sentencing decision correlate with that person committing the same or worse crime in the time before or after their hearing? In other words, is there a need for technology to aid judges in bail and sentencing decisions due to the inaccuracy of human decision-making, which demonstrably causes harm to the justice system or to the public?

Without concrete and measurable answers to questions like this, it is difficult to justify the urgent use of AI in the criminal justice system. Otherwise, AI remains an appealing tool that may fascinate the intellectual curiosity of policymakers and both computer and data scientists alike, but this technological intervention and judicial nudging may use unreliable data to cause harm to those facing the criminal justice system, without concrete demonstration that such powerful statistical techniques and algorithms are needed in the first place. If indeed, tools like COMPAS use far more variables than needed to make assessments that could be reached with far fewer factors, and unless algorithmic tools like the PSA or COMPAS demonstrably surpass the risk assessment accuracy of humans, then all responsible policymakers acting in the public interest should restrain their use of such tools until a clearly defined trial period has ascertained that the gains in efficiency and accuracy outweigh their potential harm.

5.2.2. Should a specific AI tool that is created and/or used for one particular context be used to meet the different needs of another?

Second, all policymakers outside the U.S. ought to make themselves aware of the highly specific context in which tools like the PSA and COMPAS have arisen, and critically examine whether their jurisdictions have the same needs. Consider the fact that numerous states within the U.S. are undergoing significant reforms to their bail and sentencing systems, with the former having been criticized for perpetuating systemic discrimination against poor or low-income people,[329] and the latter undergoing significant change at both the state[330] and federal levels for at least the last ten years.[331] The American criminal justice system is also one of the most privatized in the world, with a whole industry designing and marketing a broad range of products and services to meet its growing needs.[332] With these facts in mind, is it understandable that some of the most prominent cases of AI within the criminal justice system

---

329 Matt Burgess, "UK police are using AI to inform custodial decisions – but it could be discriminating against the poor", WIRED (1 March 2018), online: https://www.wired.co.uk/article/police-ai-uk-durham-hart-checkpoint-algorithm-edit.

330 Honorable Michael A. Wolff, "Evidence-Based Judicial Discretion: Promoting Public Safety Through State Sentencing Reform" (2008) 83:5 New York University Law Review 1389, online: https://www.nyulawreview.org/wp-content/uploads/2018/08/NYULawReview-83-5-Wolff.pdf.

331 Lucia Bragg, "Federal Criminal Justice Reform in 2018" (2018) 26:10 LegisBrief, online: http://www.ncsl.org/research/civil-and-criminal-justice/federal-criminal-justice-reform-in-2018.aspx.

332 David Garland, The culture of control: Crime and social order in contemporary society (Oxford: Oxford University Press, Oxford, 2001).

have arisen first in the U.S. It is also logical that the companies and organizations creating these tools — be they the Laura and Arnold Johnson Foundation or Equivant — are based in the United States and are responding to the very specific needs of their own cultural realities and legal jurisdictions.

On the other hand, risk assessment tools have come to be used in the U.S. in the wake of a move away from its cash bail system; tools like the PSA have been framed as an antidote or solution to an egregiously unfair paradigm. It is not at all clear that other countries' pre-trial release system exhibit the same deficits or problems as they exist in the U.S. In that sense, the implementation of tools like the PSA outside of the U.S. may be creating problems rather than alleviating any pre-existing ones. Technologists, lawyers and policymakers in each jurisdiction should therefore tread incredibly carefully when they transfer or implement AI technology that has been created and optimized for the U.S. criminal justice system.

5.2.3. Is the technology being designed and deployed with demonstrated transparency, mitigation of harm on vulnerable populations, and with the requirement to enable informed consent as to the risks that it poses?

It is in the interest of every government and policymaker considering the use of AI in its criminal justice system to set the highest ethical standards for the actual deployment of such tools. We draw on the work of criminologist Hannah-Moffat to identify just a few of these ideal ethical ends and draw on the findings of practitioners who critically appraised her work to identify some of the means to these ends. Quite simply,

Hannah-Moffat argues that risk assessment tools used in the criminal justice system not only ought to be justifiable but must also facilitate due process (or procedural fairness) and must inform judges as to the caveats and risks inherent in using such technology, with particular commitment to counteracting any possibility for the reproduction of systemic discrimination. This is no small feat. However, two practitioners in the U.S., both directors at their respective state commissions on sentencing in Pennsylvania and Virginia, offered insights as to how other jurisdictions might seek to accomplish these very goals.

The state of Virginia, for example, engaged in a long process to develop its risk assessment tool, replete with pilot testing, independent evaluation, and a re-validation study with numerous stakeholders such as judges, state officials, legislators, corrections officials, prosecutors, public defenders, defense attorneys, criminologists and representatives of victim's organizations.[333] Other jurisdictions should also explore the possibility for any risk assessment reports to be presented before judges in open court, so that the findings of probation officers and report in general can be subject to cross-examination by both defense counsel and prosecutor, who, in Virginia, are given access to the report for at least a week.[334] As is imaginable, the Director of Virginia's Criminal Sentencing Commission states that there is ample evidence that judges rely substantially on the risk tools, which have been proven to alter sentencing practices across the state so much so that risk assessment tools have "altered the flow of offenders

[333] Bergstrom, supra note 334 at 4.
[334] Ibid.

into prison, jail and community based sanctions."[335] Decision makers who are convinced that they must implement risk assessment tools can do two things to assess, be aware of and mitigate their harms:

1. Ensure that any findings from a risk assessment are just one part of advisory decision-making guidelines. This would seek to reduce judicial over-reliance on these reports;
2. Include in all risk assessment reports specific reliable empirical data, such as statistics demonstrating how a person with the characteristics of the accused is over-represented in the criminal justice system. By addressing the risks of the risk assessment tool, policymakers counteract the reality that certain marginalized groups will score higher with risk assessment tools due to their exposure to discrimination and inequality, and not because they are more likely to recidivate;[336]
3. Provide robust and thorough training to all major players (defense lawyers, prosecutors, judges, probation officers) on any automated technology used in the criminal justice system. To learn again from the context of Virginia, this training is so thorough that it involves the "genesis of the instrument, the study and its findings, and the risk instrument and how all of its factors are to be correctly scored."[337] More than this, it also "necessarily includes coverage of the limits and strengths of actuarial risk tools

335 Ibid.
336 Ibid at 6.
337 Ibid at 4.

> so that they can correctly interpret and apply their findings." [338]

There are numerous other issues not raised here — such as whether risk assessment should use static, historic or indeterminate, contemporary factors to assess risk. In this section, we have attempted to identify the places within criminal procedure where AI has come to be used. Our research shows that AI has thus far come to be employed especially for offender risk assessment decisions in bail and sentencing decisions.

We have also identified numerous questions that warrant further exploration, that question whether AI is needed at all to enhance or improve judicial reasoning, whether it is appropriate to transplanting AI technology optimized for one jurisdiction to other areas, and whether courts have nonetheless identified and mitigated the risks associated with the AI system they choose to use.

Table 3 – Uses of AI in criminal proceedings

Name of software	Creator	Purpose & capabilities
Correctional Offender Management Profiling for Alternative Sanctions (COMPAS)[339]	Northpointe, Inc.	Recidivism calculation tool, based on a questionnaire to be answered by the accused and used in many US jurisdictions. COMPASS evaluates variables in five main areas: criminal participation, relationships / lifestyles, personality / attitudes, family and social exclusion.

[338] Ibid.

Name of software	Creator	Purpose & capabilities
Public Safety Assessment (PSA)[340]	Laura and John Arnold Foundation	Used during the pre-trial phase, the tool assesses the likelihood of the accused committing a new crime or not appearing in court. This tool is based on a smaller number of parameters, it only takes into account variables related to the age and criminal history of a defender.
Pretrial Risk Assessment (PTRA)[341]	US Office of Pretrial and Probation Services	Pre-sentence risk assessment instrument to reduce crime by releasing the accused pending trial and to reduce unnecessary pre-trial detention.
Virginia Pretrial Risk Assessment Instrument (VPRAI)[342]	Luminosity, Inc.	Pre-sentence risk assessment tool to identify the likelihood of not appearing in court and the risk of danger to the accused community pending trial.
PREDICTICE[343]	Predictice	Determines the likelihood of success of a case based on decisions made previously and anticipates the solution of a dispute. From some user-selected parameters, predictive justice software sorts through court decisions and delivers a prognosis based on statistics.
Mathematical quantification tools for legal and judicial risk[344]	Case Law Analytics	From case law analysis in a specific area, an algorithm produces representative decisions that would likely be taken by the jurisdictions whose decisions were used to construct the mathematical model. The quantifications of risk for four

Name of software	Creator	Purpose & capabilities
		disputes are already available: compensation for dismissal without real and serious cause (wrongful dismissal); compensatory benefits; contribution to maintenance and education of children (alimony); abrupt termination of established commercial relations.
Analytical algorithms[345]	Doctrine.fr	Doctrine's artificial intelligence enriches each legal decision with a timeline, links to comments, similar decisions, or references to the same theme.

339 "COMPAS Risk & Need Assessment System", Northpointe (Website), online: http://www.northpointeinc.com/files/downloads/FAQ_Document.pdf

340 "Public Safety Assessment: A risk tool that promotes safety, equity, and justice", Arnold Foundation (Blog), online: http://www.arnoldfoundation.org/public-safety-assessment-risk-tool-promotes-safety-equity-justice/.

341 "Risk Assessment", Pretrial Justice Center for Courts (Website), online: http://www.ncsc.org/Microsites/PJCC/Home/Topics/Risk-Assessment.aspx.

342 Marie VanNostrand & Kenneth Rose, "Pretrial Risk Assessment In Virginia", Virginia Department of Criminal Justice (Website), online: https://www.dcjs.virginia.gov/sites/dcjs.virginia.gov/files/publications/corrections/virginia-pretrial-risk-assessment-report.pdf.

343 "La justice prédictive (1/3) : l'enjeu de l'ouverture des données", Le Monde Internet Actu (Blog, 9 September 2017), online : http://internetactu.blog.lemonde.fr/2017/09/09/la-justice-predictive-13-lenjeu-de-louverture-des-donnees/.

344 "L'Intelligence artificielle au service de la quantification du risque juridique", Case Law Analytics (Website), online: http://caselawanalytics.co.m

345 "Le moteur de recherche juridique", Doctrine (Website), online: https://www.doctrine.fr.

6. CONCLUSION AND RECOMMENDATIONS

We have attempted to provide in this report an overview of the multiple existing and future applications of AI technologies in the criminal justice system, from the use of malicious AI by online offenders to their deployment by law enforcement organizations to detect, predict and investigate crimes. Court officers and correctional services are also increasingly relying on AI to make decisions on the risk levels, culpability, sentencing and release of offenders. As we have found, a growing number of tools promise to enable the processing of a deluge of data to support complex decision-making with the aim to enhance the security of modern societies. However, these AI tools also raise four main categories of challenges that are particularly critical in the field of criminal justice, because of their potential impact on individual freedoms: ethics, effectiveness, procurement, and appropriation. These four groups of issues are closely interconnected, affecting and amplifying each other, and need to be thoroughly addressed before AI becomes adopted on a large scale and routinely embedded into criminal justice procedures.

6.1. Ethical challenges

The central challenge created by the development and deployment of AI tools in a criminal justice setting is of an ethical nature. If AI can certainly generate many uncontroversial social

benefits such as more reliable medical diagnosis, less congested (and therefore polluted) thoroughfares, or better farming outcomes in developing countries, to name just a few, its application to a law enforcement or judicial context raises a number of moral dilemmas related to a clash with fundamental principles such as fairness and justice. In her seminal book, Virginia Eubanks has for example shown how these new algorithmic tools of social control can exclude and isolate the most vulnerable members of our societies, intruding into their lives and denying them basic services or singling them out for enhanced forms of intervention.[346]

We have outlined in previous chapters the practical manifestations of these ethical dilemmas in law enforcement and criminal proceedings, and we will therefore not reiterate these concerns here. Instead, we will focus on the ethical frameworks that are being elaborated as a response to minimize the negative social impacts of AI. Because they are formulated at a higher level of generality to be applicable to the broadest possible range of situations, these ethical frameworks represent a good starting point for criminal justice agencies that wish to adopt a transparent approach to the implementation of AI technologies. Although a few more are available, we focus on five frameworks that have been designed through a diversity of approaches, including efforts led by a regulatory authority (France), legislators (UK), scientists and engineers (Japan and IEEE), or an academic institution (Canada).

---

346 Virginia Eubanks, Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor (New York: NY, St Martin's Press, 2017).

Table 4 - Overview of existing AI ethical frameworks[347]

Framework	CNIL (French Data Protection Authority)	Japanese Society for Artificial Intelligence Ethical Guidelines	House of Lords – Select Committee on Artificial Intelligence	Ethically Aligned Design V. 2 (IEEE)	Montreal Declaration for a Responsible Development of Artificial Intelligence
Country	France	Japan	UK	International	Canada
Year released	2017	2017	2018	2018	2018
Stakeholders consulted	3000 people who took part in 45 debates organized by 60 partners (research centres, public institutions, trade unions, think tanks, companies)	Ethics committee made up of 12 members (AI researchers, a science fiction writer, a journalist, an STS researcher)	57 experts from academia, industry, government, and advocacy groups	Several hundred participants from academia, industry, civil society, and government	More than 500 citizens and experts from academia, industry, advocacy groups, and government

[347] Other frameworks include the European Union's Statement on Artificial Intelligence, Robotics and 'Autonomous' Systems, The Future of Life Institute's Asilomar AI Principles, the Allen Institute for Artificial Intelligence's Draft Principles of AI Ethics, or the ACM's Principles for Algorithmic Transparency and Accountability.

Framework	CNIL (French Data Protection Authority)	Japanese Society for Artificial Intelligence Ethical Guidelines	House of Lords – Select Committee on Artificial Intelligence	Ethically Aligned Design V. 2 (IEEE)	Montreal Declaration for a Responsible Development of Artificial Intelligence
Principles	Fairness (personal and collective outcomes) Continued attention and vigilance	Contribution to humanity Abidance of laws and regulations Respect for privacy Fairness Security Act with integrity Accountability and social responsibility Communication with society and self-development Abidance of ethics guidelines by AI	Common good and benefit to humanity Intelligibility and fairness Not to be used to diminish data rights or privacy of users Right for citizens to be educated to flourish alongside AI Ban on autonomous power to hurt, destroy or deceive human beings	Human rights Well-being Accountability Transparency Minimize the risks of misuse	Well-being Respect for autonomy Protection of privacy and intimacy Solidarity Democratic participation Equity Diversity inclusion Prudence Responsibility Sustainable development

Table sources:

France: CNIL, How can humans keep the upper hand? The ethical matters raised by algorithms and artificial intelligence (Paris : Commission Nationale Informatique & Libertés, 2017) online at https://www.cnil.fr/en/algorithms-and-artificial-intelligence-cnils-report-ethical-issues.

Japan : The guidelines are available online in English (http://ai-elsi.org/wp-content/uploads/2017/05/JSAI-Ethical-Guidelines-1.pdf) and in Korean (http://ai-elsi.org/wp-content/uploads/2017/09/ - -20170303-KoNIBP.pdf).

UK: Select Committee on Artificial Intelligence, AI in the UK: Ready, willing and able? (London: House of Lords, 2018), online at https://publications.parliament.uk/pa/ld201719/ldselect/ldai/100/100.pdf.

IEEE: IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems, Ethically aligned design v2, (Piscataway : IEEE 2018), online at https://ethicsinaction.ieee.org/.

Canada: Déclaration de Montréal IA responsable, Rapport de la Déclaration de Montréal pour un développement responsable de l'intelligence artificielle (Montréal: Université de Montréal 2018).

What unites these five frameworks is a core set of principles that emphasize transparency (thereby facilitating democratic debate and participation on the use of this technology), benefits to individuals and society, respect for privacy, and accountability. Even if most of these principles might seem overly abstract, a few of the frameworks listed above offer more detailed policy and design recommendations that can be embedded into actual AI systems. Additionally, the thorny issue of preventing innovative criminal actors from exploiting these principles of openness and fairness to escape legitimate surveillance and enforcement activities has not yet been discussed. The inevitable tension between the conflicting rights of offenders and their victims has not been addressed either. Maybe these worries are slightly premature, as a few researchers are starting to question the claimed effectiveness of AI to make reliable predictions in highly unstructured domains of application.

6.2. Effectiveness challenges

The advances of AI in general, and deep learning in particular, have been impressive over the past few years after a long hiatus of several decades. However, they have so far been limited to a few domains where data is plentiful and already fairly well structured and labelled, such as speech recognition and translation, image recognition, or game playing.[348] Gary Marcus, a psychology professor at NYU who also founded a machine learning company presented the most elaborate discussion of why DL approaches do not seem very well suited

[348] Marcus, supra note 22 at 1.

to unstable domains where generalizations have to be made from limited data. He lists 10 challenges that we already mentioned in chapter 1 but that we believe need to be detailed here to illustrate why the impressive results delivered by AI in certain fields might not transfer seamlessly to criminal justice applications:

- While humans can learn quickly from a few rules and examples, machine learning models must ingest vast amounts of data to produce reliable decisions. The quantity of useable data that criminal justice agencies can feed to AI models on rare forms of offending might not be sufficient to generate robust predictions;
- The learning process underlying many AI tools is also shallower or narrower than the Deep Learning terminology leads to believe, meaning that an impressive performance in one area (language translation) cannot easily translate into a different area (such as predicting the chances of recidivism);
- Deep Learning has no natural way to deal with hierarchical structure, which means that all the available variables are considered on the same level, as 'flat' or non-hierarchical. This presents a major hurdle when decisions carry a heavy moral or legal weight that must supersede other features;
- Deep Learning tools struggle with open-ended inferences that an investigator, a judge or a parole officer might pick up intuitively and effortlessly;
- The 'black box' nature of AI tools enables them to make

predictions based on thousands or even millions of variables whose interactions are impervious to human analysis. This extraordinary level of complexity also makes the reflexive process that led to those predictions very hard to explain. If this opacity might not be too controversial when labelling cat pictures or providing YouTube videos subtitles, it is a lot more disturbing when AI tools are used to assess the recidivism risk of a convicted offender or even to conduct pre-emptive patrols in minority neighbourhoods, with outcomes and a potential for mistakes that can affect the lives and freedoms of many;

- This is compounded by the fact that AI systems can hardly differentiate causation from correlation, which is problematic for institutions that need to remain highly accountable;
- Because of the 'flat' and 'black box' approaches mentioned above, Deep Learning resists integrating prior knowledge. It is highly revealing for example that the core PredPol algorithm has been borrowed from seismology rather than developed from the multiple theories of crime and place that are common in criminology.[349] This refusal to recognize prior knowledge seems deliberate, both epistemically due to the history of a research field that has favoured self-contained problems to solve, and technically because it would mean making AI tools less effective. So, in areas where knowledge has to be integrated across very diverse fields (such as in criminal justice), humans will remain

[349] Bilel Benbouzid, "Des crimes et des séismes: La police prédictive entre science, technique et divination", 6: 206 Réseaux 95 at 123.

much more effective than AI, even if researchers are exploring the potential of 'apprenticeship learning' to enable machines to learn from observing experts at work;[350]

- The technical features highlighted above imply that AI systems are most effective in stable environments where the interactions between underlying variables and outcomes remain constant over time and the growing availability of data can only enhance a system's performance. Unfortunately, criminal offenders are a very innovative bunch who relentlessly imagine new ways to manipulate their environment and evade social control mechanisms and enforcement strategies;
- Fragility remains a key feature of AI systems: they can outperform humans on very narrow tasks most of the time but can also fail spectacularly when seemingly minute details in the data they analyze interfere with their internal logic. In a highly publicized paper, Jiawei Su and his colleagues showed that a deep learning algorithm performing image recognition tasks could be fooled by changing a single pixel in an otherwise perfectly normal picture. As a result, it misidentified a horse as a frog, a deer as an airplane, or a cat as a dog.[351] One can imagine that criminal justice agencies require much more robust and reliable tools with very limited failure rates;

---

350 P. Abdeel, & A.Y. Ng, A., "Apprenticeship learning via inverse reinforcement learning", (Paper delivered at the 21st International Conference on Machine Learning, 4-8 July 2004), online: https://dl.acm.org/citation.cfm?id=1015430.

351 J. Su, D. Vasconcellos Vargas, & K. Sakurai, "One pixel attack for fooling deep neural networks", (2017) arXiv Working Paper, arXiv:1710.08864 [cs.LG], online at https://arxiv.org/abs/1710.08864v4.

- Finally, from an engineering perspective, it appears that even high performing AI systems are difficult to embed in legacy systems that may have been in operation for a few decades, particularly in the context of criminal justice agencies that have been slower than other organizations to adopt new technologies and operate therefore with legacy systems that create major frictions with contemporary technologies.[352]

Hence, one should remain very careful about the marketing hype that envelops AI systems and their real-life applications by bureaucratic organizations that do not always possess the skill-sets, infrastructures and cultures needed for such a paradigmatic shift. To paraphrase a concept popularized by the consultancy firm Gartner, AI may have reached the "peak of inflated expectations", but its "plateau of productivity" might still be years ahead.[353] An example of this disconnect between the promises and the reality of AI can be found in a recent investigative article published in *The Guardian*, which examined the case of "pseudo-AIs", where companies selling those systems hire humans in developing countries to manually do the work supposed to be automated by their technology.[354] This

---

352 C. Bellamy, & J. Taylor, "New information and communications technologies and institutional change: The case of the UK criminal justice system," (1996) 9:4 International Journal of Public Sector Management 51.

353 "Gartner Hype Cycle", Gartner (Website), online: https://www.gartner.com/en/research/methodologies/gartner-hype-cycle .

354 O. Solon. "The rise of 'pseudo-AI': how tech firms quietly use humans to do bots' work", The Guardian (6 July 2018), online: https://www.theguardian.com/technology/2018/jul/06/artificial-intelligence-ai-humans-bots-tech-companies.

"fake it until you make it" approach should serve as a warning to criminal justice agencies considering the purchase of an AI tool, an exercise fraught with challenges, as we'll see in the next section.

6.3. Procurement challenges

The ethical and technical considerations outlined above also reverberate through the acquisition processes of AI systems by criminal justice organizations, raising a number of procedural issues that can in turn create ethical and performance implications of their own if they are not handled properly. In other words, the competitive business practices of companies that design and market AI technologies, and in particular the confidentiality requirements that they attach to their products to protect their intellectual property, often collide with the need for public transparency and accountability that characterize the work of government agencies. One of the best examples of this tension is the refusal from Northpointe Inc. (now Equivant), the company that sells the COMPAS system discussed previously in this report, to let defendants and journalists review and challenge the software's secret algorithm.[355] A comprehensive analysis of the best practices government users should adopt when purchasing and implementing AI solutions, to better manage the ethical and performance risks associated with this complex technology, has been provided by Gretchen Greene.[356]

---

[355] Adam Liptak, "Sent to prison by a software program's secret algorithm", The New York Times (1 May 2017), online: https://www.nytimes.com/2017/05/01/us/politics/sent-to-prison-by-a-software-programs-secret-algorithms.html.

[356] K. Gretchen Greene, "Buying you first AI or 'never trust a used

She highlights six issues that should be discussed in great detail by government agencies with the AI companies selling them these new systems.

Despite resistance from the companies that develop AI solutions, a government agency acquiring this kind of product should be able to access its source code and to analyze the algorithms that power it. The practice of buying 'black box algorithms' is often justified by its proponents on the basis of maintaining a seller's technological edge (its 'secret sauce') in the face of relentless competition, but also to avoid the manipulation of neural networks by malicious actors, as we have seen in chapter 2.[357] While not all public organizations may have the maturity and resources to develop their own open source tools and algorithms, they should at least be able (some would add compelled) to inspect how the technology they plan to buy is built and how it makes the decisions that will impact their citizens. One of the key features of Deep Learning algorithms is that they may produce results that are not fully explainable because of the large number and complexity of variables that they are able to incorporate in their computations, but a robust understanding of their underlying code should nevertheless inform their deployment by criminal justice institutions, to reduce unforeseen instances of bias.

---

algorithm salesman'", Berkman Klein Center for Internet & Society — AI Ethics & Governance (7 November 2018), online: https://medium.com/berkman-klein-center/buying-your-first-ai-136cd2e6dd2.

357 L. Maffeo, "The case for open source classifiers in AI algorithms", opensource.com (18 October 2018), online: https://opensource.com/article/18/10/open-source-classifiers-ai-algorithms.

The minimum requirements for source code and algorithm transparency outlined above should also extend to the data that has been used to train the algorithms under consideration, or that will be used to make predictions. Machine learning models usually require vast amounts of data to reach optimal outcomes and make reliable predictions, but the nature of the data fed to these systems at the training stage determines the quality of the decisions made when they become operational. The use of biased data—such as data reflecting racial disparities stemming from discriminatory enforcement or sentencing practices—to train an AI model will generate an equally-biased outcome that will tend to reproduce an undesirable situation, only coated with a scientific varnish. It is therefore essential that any ready-to-use AI tool be examined not only for the quality of its algorithm, but also for the quality of the data used to train it. When AI tools are developed internally with local data, this assessment is much easier to make than when a police organization or a court system purchases an off-the-shelf AI that has been trained with data from an uncertain origin.

Finally, the independent variables that are used by algorithms to make predictions about particular outcomes should also be thoroughly scrutinised. These variables are the levers that algorithms pull to classify the data and make predictions. In criminal justice applications, some common variables traditionally used in statistical analyses are the age, gender, race, income, education, health, social network or prior convictions of a suspect. However, the analytical power of machine learning algorithms and the computer systems that run them means that they can process thousands of variables to make a decision. In the context of an AI used to assess eligibility for parole, the

algorithm could for example make use of seemingly unrelated features such as the color of one's eyes, musical tastes or downloaded apps, providing they can be extracted from the data. Some of those variables might be correlated with race or socio-economic status and be particularly prone to bias. Hence, it becomes essential to review what variables have the biggest effect and to make sure the causality is well understood and aligned with the principles of justice and fairness.

Some companies such as IBM are developing tools that help organizations translate those code, data and variable transparency principles into practice. Its *AI OpenScale* technology, launched in 2018, claims to be able to automate bias detection and mitigate it for a broad range of machine learning products, providing explanations on how decisions are being made and reinforcing the confidence in their outcome.[358] DARPA, the American defense research agency, has also launched an *Explainable AI* program that will seek to produce machine learning techniques enabling human users to understand more easily how predictions are made and how reliable they are.[359] These new applications will be particularly useful in the criminal justice context.

Beyond pure technical considerations, many defendants, victims and criminal justice professionals will be affected by the growing number of decisions made by AI systems. The odds

358 "AI OpenScale", IBM (Website), online: https://www.ibm.com/cloud/ai-openscale/

359 David Gunning, "Explainable Artificial Intelligence (XAI)", DARPA (Website), online: https://www.darpa.mil/program/explainable-artificial-intelligence.

for a defendant of being prosecuted, convicted, sentenced and released on parole might be significantly altered under this new regime. This radical transformation in the administration of criminal justice cannot be implemented without a proper understanding of how outcomes will differ from the current arrangements, where decisions are made exclusively by humans. The harms that can be caused by AI malfunctions (false positives or false negatives for example) should in particular be incorporated in the decision-making process. Meanwhile, the expertise required from police officers, prosecutors, defense lawyers, judges, correctional and parole officers will require considerable re-skilling efforts. AI tools might also be used to generate efficiency gains that will result in job losses. Therefore, comprehensive algorithmic impact assessments should be conducted when implementing a new AI system in order to assess the multiple organizational and service delivery implications of such a decision.[360]

Whether the criminal justice organization implementing a new AI tool decides to invest directly in the digital infrastructure required to deploy such technology, or on the contrary prefers to rely on a cloud provider to host the production backend, data security and privacy will need to be guaranteed. As we've indicated many times throughout this report, the predictive effectiveness of an AI model rests heavily on the quantity of data it can ingest and process, the more the better. However, large databases are exposed to the constant attacks of malicious

---

360 Greene, supra note 365; Stats NZ, "Algorithm assessment report" (Wellington: New Zealand Government, 2018) online: https://www.data.govt.nz/assets/Uploads/Algorithm-Assessment-Report-Oct-2018.pdf at 33.

hackers motivated by financial gain, ideology, revenge, or sponsored by government agencies. Since 2013, the data breach database maintained by Breach Level Index, a Gemalto initiative, has identified more than 9,700 breaches that have compromised more than 13 billion records.[361] Criminal justice agencies are not immune from this trend and many police services, court databases and even correctional computer systems have already been hacked. A security incident might involve a malicious actor accessing the vast troves of personal information centralized by an AI system to predict a criminal justice outcome, or trying to poison the AI system in order to change a prediction and thereby influence the outcome for which a prediction is sought. Such use cases should not be discarded as science fiction scenarios, and the purchase of any AI system by a criminal justice agency should not be completed before stringent security audits of the service providers competing for the contract, as well as their IT contractors, are conducted to ensure that their technology and the data that it will process benefit from high levels of protection against theft and tampering.

Finally, contractual terms should be studied carefully to ensure a full understanding of the licensing model that is being offered. It is particularly important to establish how IP rights will be allocated over time, especially for a technology that learns constantly from new data and adjusts its models accordingly. The costs incurred over the lifetime of an AI deployment also need to be clearly understood by all parties. Training strategies and infrastructure choices will have vastly different financial implications on the success or failure of such projects. A testing

361 "Data Breach Statistics", Breach Level Index (Website), online: https://breachlevelindex.com/

period is an option recommended by Gretchen Greene, who also advises to define clear performance criteria and goals that will be used to measure success.[362] Finally, a criminal justice agency buying this kind of complex product or service should not hesitate to ask what warranty comes with it, both in terms of effectiveness and liability against failures.

A recent assessment conducted by the New Zealand government across 14 agencies indicates that most of them (10) use a mixed-procurement model, by contrast with an internal development or a 'pure' external procurement model, to which most of the challenges discussed above apply. The mixed approach favoured in New Zealand involves contracting external expertise into an internal development process to mitigate the potential risks associate with the two other alternatives (lack of expertise or lack of control over external expertise).[363] Beside this first country-wide assessment, there is still very limited knowledge of the modalities through which AI is being introduced into government agencies.

Even when procurement challenges are deftly negotiated, the direct users of a technology also play a central role in its successful adoption, no matter how sophisticated and powerful this technology proves to be.

6.4. Appropriation challenges

We have assumed until now that AI systems will find their way

[362] Greene, supra note 365.
[363] Stats NZ, supra note 370.

into criminal justice organizations in a neutral environment, where professionals passively implement them as intended by their hierarchy and designers. This is of course a sociological fiction that ignores the powerful appropriation practices of frontline police officers, crime analysts, judges, parole officers and many other criminal justice professionals. The policing and security literature has established that if security technologies and devices have certainly become compulsory and shape the everyday practices of their human users, the latter always retain high levels of agency that can take different forms and range from domestication to resistance and even sabotage.[364] The concept of appropriation reflects the creativity of individual agents within complex organizations, who translate the technology they are entrusted with into practices that can either be routinized or innovative, meaning that they can absorb a technology into existing cultural values and disarm its reform potential, or on the contrary repurpose a technology to fit their operational needs in unexpected ways. Bluntly stated in a law enforcement context, "whatever technology increases the officer's sense of efficacy will be used and modified, and what is not useful will be destroyed, sabotaged, avoided, or used poorly".[365] Hence, AI is the latest technology in a long succession of criminal justice innovations that have sought to improve the delivery of justice and the effectiveness of its institutions, but that may end up being much less disruptive that anticipated.

---

364 R. Ericson, & K. Haggerty, Policing the risk society (Oxford: Clarendon Press, 1997); A. Amicelle, C. Aradau, & J. Jeandesboz, "Questioning security devices: Performativity, resistance, politics," (2015) 46:4 Security Dialogue 293.

365 P. K. Manning, The technology of policing: Crime mapping, information technology, and the rationality of crime control (New York: NY, New York University Press, 2008) at 250.

Thus, the final recommendation this report makes is to start planning the research efforts that will be needed to understand how the new assemblages of humans and AI-powered machines that will soon been pervasive in criminal justice institutions will operate, not in theory or in a dystopian configuration, but in day-to-day practice, and what sorts of intended and unintended consequences will emerge as a result. Ethnographic studies adopting a similar approach as Ericson and Haggerty's 'Policing the risk society' or Manning's 'The technology of policing' should be funded to capture how AI systems will be "'retro-fitted' to the [criminal justice] organizations' practices, structures, and routines".[366] Only then will we be able to move beyond the current fetishism of algorithms to assess the full scope of the promised AI revolution on the delivery of security and justice.

[366] Ibid at 276.

List of References

"'I'm Not A Robot': Google's Anti-Robot reCAPTCHA Trains Their Robots To See", AI Business, (25 October 2017), online: https://aibusiness.com/recaptcha-trains-google-robots/.

"8 Companies Using AI for Law Enforcement", Nanalyze (Website), online: https://www.nanalyze.com/2017/11/8-companies-ai-law-enforcement/.

"AI OpenScale", IBM (Website), online: https://www.ibm.com/cloud/ai-openscale/

"AI vs. Lawyers", LawGeex Blog (26 February 2018), online: https://blog.lawgeex.com/ai-more-accurate-than-lawyers/.

"Algorithms in the Criminal Justice System", Electronic Privacy Information Center (Website), online: https://epic.org/algorithmic-transparency/crim-justice/.

"Amazon Deep Learning AMIs", Amazon Web Service (Website) online: https://aws.amazon.com/machine-learning/amis/

"Analytics Desktop", Cellebrite (Website), online: https://www.cellebrite.com/en/products/analytics-desktop/.

"Analytics Enterprise", Cellebrite (Website), online: https://www.cellebrite.com/en/products/analytics-enterprise/.

"Artificial intelligence to enhance Australian judiciary system", Swineburn University of Technology (Blog) (29 January

2018), online: http://www.swinburne.edu.au/news/latest-news/2018/01/artificial-intelligence-to-enhance-australian-judiciary-system.php.

"arXiv.org e-Print archive", arXiv.org (Website), online: https://arxiv.org/.

"Astroturfing, Twitterbots, Amplification - Inside the Online Influence Industry", The Bureau of Investigative Journalism (7 December 2017), online: https://www.thebureauinvestigates.com/stories/2017-12-07/twitterbots.

"Axon AI and Policing Technology Ethics Board", Axon (Website), online: https://ca.axon.com/info/ai-ethics.

"Bail Schedule Law and Legal Definition", USLegal (Website), online: https://definitions.uslegal.com/b/bail-schedule/.

"Boomerang III: State-of-the-Art Shooter Detection", Raytheon (Website), online: https://www.raytheon.com/capabilities/products/boomerang.

"China uses facial recognition to arrest fugitives", NHK World – Japan (26 December 2018), online: https://www3.nhk.or.jp/nhkworld/en/news/20181227_10/.

"Cloud AI | Cloud AI", Google Cloud (Website), online: https://cloud.google.com/products/ai/; jonbeck7, "Azure Windows VM sizes - GPU", Microsoft (Website), online: https://docs.microsoft.com/en-us/azure/virtual-machines/windows/sizes-gpu.

"Cloud TPUs - ML accelerators for TensorFlow", Google Cloud (Website), online: https://cloud.google.com/tpu/.

"COMPAS Classification", Equivant (Website), online : http://www.equivant.com/solutions/inmate-classification.

"COMPAS Risk & Need Assessment System", Northpointe (Website), online: http://www.northpointeinc.com/files/downloads/FAQ_Document.pdf

"COMPAS", Northpointe (Website, via Internet Archive), online: https://web.archive.org/web/20160315175056/http://www.northpointeinc.com:80/risk-needs-assessment

"Data Breach Statistics", Breach Level Index (Website), online: https://breachlevelindex.com/

"DeepMind", DeepMind (website) online: https://deepmind.com.

"DNA Forensics: The application of genetic testing for legal purposes", GeneEd (Website), online: https://geneed.nlm.nih.gov/topic_subtopic.php?tid=37.

"Evidence-Based Practice Implementing the COMPAS Assessment System", Northpointe (Website, via Internet Archive), online: https://web.archive.org/web/20160506140944/http://www.northpointeinc.com/downloads/whitepapers/EVIDENCE-BASED_PRACTICE-implementing_COMPAS.pdf.

"Faception : Facial Personality Analytics", Faception (Website), online : https://www.faception.com/.

"Faception: Our technology", Faception Website, online: https://www.faception.com/our-technology.

"Faception", Faception (Website), online: https://www.faception.com/.

"Facial recognition the future of cashless payment in China", Asia Times (20 December 2018), online: http://www.atimes.com/article/facial-recognition-the-future-of-cashless-payment-in-china/.

"fast.ai", fast.ai (Website), online: https://www.fast.ai/; "Google Launches Free Course on Deep Learning: The Science of Teaching Computers How to Teach Themselves", Open Cult (Website), online: http://www.openculture.com/2017/07/google-launches-free-course-on-deep-learning.html.

"FGNet Results", MegaFace (Website), online:

"Gartner Hype Cycle", Gartner (Website), online: https://www.gartner.com/en/research/methodologies/gartner-hype-cycle.

"GET statuses/user_timeline", Twitter (Website) online: https://developer.twitter.com/en/docs/tweets/timelines/api-reference/get-statuses-user_timeline.html/

"Google Duplex: An AI System for Accomplishing Real-World Tasks Over the Phone", Google AI (Blog), online: http://ai.googleblog.com/2018/05/duplex-ai-system-for-natural-conversation.html.

"Google Unveils Neural Network with 'Superhuman' Ability to Determine the Location of Almost Any Image", MIT Technology Review (24 February 2016), online: https://www.technologyreview.com/s/600889/google-unveils-neural-network-with-superhuman-ability-to-determine-the-location-of-almost/.

"Have I Been Pwned: Check if your email has been compromised

in a data breach", Have I Been Pwned (Website) online: https://haveibeenpwned.com/.

"How does facial recognition work?", Norton Security Center, online: https://us.norton.com/internetsecurity-iot-how-facial-recognition-software-works.html

"ImageNet", Image-Net (Website) online: http://image-net.org/index.

"Intelligence", Palantir (Website), online: https://www.palantir.com/solutions/intelligence/.

"Introducing Magnet.AI: Putting Machine Learning to Work for Forensics", Magnet Forensics (Website), online: https://www.magnetforensics.com/blog/introducing-magnet-ai-putting-machine-learning-work-forensics/

"Introduction to Computer Vision", Algorithmia Blog (2 April 2018), online https://blog.algorithmia.com/introduction-to-computer-vision/ ; Golan Levin, "Image Processing and Computer Vision", OpenFrameworks, online: https://openframeworks.cc/ofBook/chapters/image_processing_computer_vision.html.

"Is facial recognition technology racist?", The Week UK (27 July 2018), online: https://www.theweek.co.uk/95383/is-facial-recognition-racist.

"L'Intelligence artificielle au service de la quantification du risque juridique",, Case Law Analytics (Website), online: http://caselawanalytics.com

"La justice prédictive (1/3) : l'enjeu de l'ouverture des données", Le Monde Internet Actu (Blog, 9 September 2017), online : http://internetactu.blog.lemonde.fr/2017/09/09/la-justice-

predictive-13-lenjeu-de-louverture-des-donnees/.

"Le moteur de recherche juridique", Doctrine (Website), online: https://www.doctrine.fr.

"Let's Go Vishing", (22 December 2014), online: Security Through Education. https://www.social-engineer.org/general-blog/lets-go-vishing>.

"Life Facial Recognition Trial", Metropolitan Police (Website), online: https://www.met.police.uk/live-facial-recognition-trial/.

"LSI-R: Level of Service Inventory-Revised", MH Assessments (Website), online: https://www.mhs.com/MHS-Publicsafety?prodname=lsi-r

"Lyrebird: Ultra-Realistic Voice Cloning and Text-to-Speech", Lyrebird.a (Website), online: https://lyrebird.ai/.

"Magnet Forensics Launches Magnet.AI to Fight Child Exploitation", StartUp Toronto (Website), online: https://startupheretoronto.com/sectors/technology/magnet-forensics-launches-magnet-ai-to-fight-child-exploitation/.

"Measuring the Filter Bubble: How Google is influencing what you click", DuckDuckGo Blog (4 December 2018), online: https://spreadprivacy.com/google-filter-bubble-study/.

"Meet the Safety Team", Facebook Safety (9 August 2011), online: https://www.facebook.com/notes/facebook-safety/meet-the-safety-team/248332788520844/

"Neural fuzzing: applying DNN to software security testing", Microsoft Research (13 November 2017), online: https://www.microsoft.com/en-us/research/blog/neural-fuzzing/; Mohit Rajpal, William Blum & Rishabh Singh, "Not

all bytes are equal: Neural byte sieve for fuzzing", (2017) arXiv Working Paper, arXiv:1711.04596 [cs.SE], online: https://arxiv.org/abs/1711.04596.

"New Phishing Techniques To Be Aware of: Vishing and Smishing", MakeUseOf (Website), online: https://www.makeuseof.com/tag/new-phishing-techniques-aware-vishing-smishing/.

"Northpointe Suite: Automated Decision Support", Northpointe (Website, via Internet Archive), online: https://web.archive.org/web/20160307002839/http://www.northpointeinc.com/.

"Official Launch Of The Montréal Declaration For Responsible Development Of Artificial Intelligence", Montréal Declaration for Responsible Development of Artificial Intelligence (Website) (4 December 2018), online: https://www.declarationmontreal-iaresponsable.com/blogue/d%C3%A9voilement-de-la-d%C3%A9claration-de-montr%C3%A9al-pour-un-d%C3%A9veloppement-responsable-de-l-ia.

"OpenFace: Free and open source face recognition with deep neural networks", OpenFace (Website), online: https://cmusatyalab.github.io/openface/

"Overview", PredPol (Website), online: https://www.predpol.com/about/.

"Partners", National Center for Missing & Exploited Kids (Webste), online: http://www.missingkids.org/supportus/partners.

"Penetration Testing Software, Pen Testing Security", Metasploit

(Website), online: https://www.metasploit.com/.

"Phishing", Know4Be (Website), online: https://www.knowbe4.com/phishing.

"Phishing", Security Through Education (Website), online: https://www.social-engineer.org/framework/attack-vectors/phishing-attacks-2/.

"PhotoDNA", Microsoft (Website), online: https://www.microsoft.com/en-us/photodna.

"Pittsburgh-based tech company debuts first facial recognition technology designed to halt global human trafficking", Marinus Analytics (Website), online:

"Practitioner's Guide to COMPAS Software", Northpointe (Website, via Internet Archive), online: https://web.archive.org/web/20160507022911/http://www.northpointeinc.com/downloads/compas/Practitioners-Guide-COMPAS-Core-_031915.pdf.

"Predictive Policing Research", National Institute of Justice (Website), online: https://www.nij.gov/topics/law-enforcement/strategies/predictive-policing/Pages/research.aspx.

"Pretrial Release Recommendation Decision Making Framework (DMF)", New Jersey Courts (March 2018), online: https://www.njcourts.gov/courts/assets/criminal/decmakframwork.pdf?cacheID=JOvH2H8.

"Principles for Accountable Algorithms and a Social Impact Statement for Algorithms", FAT/ML (Website), online: http://www.fatml.org/resources/principles-for-accountable-algorithms.

"Public Safety Assessment New Jersey Risk Factor Definitions-March 2018", New Jersey Courts, online: https://www.njcourts.gov/courts/assets/criminal/psariskfactor.pdf?cacheID=IDYJVkr.

"Public Safety Assessment: A risk tool that promotes safety, equity, and justice", Arnold Foundation (Blog), online: http://www.arnoldfoundation.org/public-safety-assessment-risk-tool-promotes-safety-equity-justice/.

"PyTorch", PyTorch (Website), online: https://www.pytorch.org; "TensorFlow",TensorFlow (Website) online: https://www.tensorflow.org/.

"Quandl", Quandl (Website), online: https://www.quandl.com.

"RFP: National Provider of Training & Technical Assistance", Arnold Foundation (Website), online: https://www.arnoldfoundation.org/wp-content/uploads/PSA-National-Provider-RFP.pdf.

"Risk Assessment", Pretrial Justice Center for Courts (Website), online: http://www.ncsc.org/Microsites/PJCC/Home/Topics/Risk-Assessment.aspx.

"Sample-COMPAS-Risk-Assessment-COMPAS-'CORE'", DocumentCloud (Hosting service), online: https://www.documentcloud.org/documents/2702103-Sample-Risk-Assessment-COMPAS-CORE.html

"Self-Exclusion Program", Ontario Lottery and Gaming Corporation (Website), online: https://about.olg.ca/self-exclusion/facial-recognition/.

"SenseTime: Our Company", SenseTime (Website), online:

https://www.sensetime.com/ourCompany.

"Siri", Apple (Website), online: https://www.apple.com/siri/.

"Social Engineering Defined", Security Education (Website), online: https://www.social-engineer.org/framework/general-discussion/social-engineering-defined/.

"Spear Phishing", Know4Be (Website) online: https://www.knowbe4.com/spear-phishing/.

"StopLift", Stoplift (Website), online: https://www.stoplift.com/.

"TASER International's (TASR) CEO Rick Smith on Q4 2016 Results - Earnings Call Transcript", Seeking Alpha (28 February 2017), online: https://seekingalpha.com/article/4050796-taser-internationals-tasr-ceo-rick-smith-q4-2016-results-earnings-call-transcript?page=3.

"Tattoo Recognition", FBI.gov, (25 June 2015), online: https://www.fbi.gov/audio-repository/news-podcasts-thisweek-tattoo-recognition.mp3/view.

"The Future of Firearm Forensics is 3D", Cadre Forensics (Website), online: https://www.cadreforensics.com/.

"The Malicious Use of Artificial Intelligence", The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation (Website) online: https://maliciousaireport.com/.

"The top frauds of 2017", Consumer Information, (1 March 2018), online: https://www.consumer.ftc.gov/blog/2018/03/top-frauds-2017.

"This PSA About Fake News From Barack Obama Is Not What

It Appears", BuzzFeed News (17 April 2018), online: https://www.buzzfeednews.com/article/davidmack/obama-fake-news-jordan-peele-psa-video-buzzfeed.

"UFED Ultimate", Cellebrite (Website), online: https://www.cellebrite.com/en/products/ufed-ultimate/.

"VALCRI", VALCRI (Website), online: http://www.valcri.org/.

"Veritone® Announces New AI-Powered Law Enforcement Application Suite to Collectively Expedite Investigations and Evidence Disclosure", Business Wire (20 September 2018), online: https://www.businesswire.com/news/home/20180920005160/en/Veritone%C2%AE-Announces-New-AI-Powered-Law-Enforcement-Application.

"Vishing", Security Through Education (Website), online: https://www.social-engineer.org/framework/attack-vectors/vishing/.

"Ways to Build with Amazon Alexa", Amazon (Website), online: https://developer.amazon.com/alexa.

"What is AGI?", (11 August 2013), online: Machine Intelligence Research Institute https://intelligence.org/2013/08/11/what-is-agi/.

"What is Bail? Understanding What Bail is & Different Types of Bal Bonds", Bail USA (Website), online: http://www.bailusa.net/what-is-bail.php.

Abdeel, P. & A.Y. Ng, A., "Apprenticeship learning via inverse reinforcement learning", (Paper delivered at the 21st International Conference on Machine Learning, 4-8 July 2004), online: https://dl.acm.org/citation.cfm?id=1015430.

Aggarwal, Alok, "The Current Hype Cycle in Artificial Intelligence", Scry Analytics (20 January 2018) online: https://scryanalytics.ai/the-current-hype-cycle-in-artificial-intelligence.

Alabama Rules of Criminal Procedure, Rule 7.2(b), online: http://judicial.alabama.gov/docs/library/rules/cr7_2.pdf.

AlMarhoos, Rasha, "Phishing for the answer: Recent developments in combating phishing", (2007) 3:3 I/S: A Journal of Law and Policy for the Information Society 595.

Anderson, Hyrum S et al, "Learning to Evade Static PE Machine Learning Malware Models via Reinforcement Learning" (2018) arXiv Working Paper, arXiv:180108917 [cs], online: http://arxiv.org/abs/1801.08917.

Angwin, Julia, Jeff Larson, Surya Mattu & Lauren Kirchner, "Machine Bias", ProPublica (23 May 2016), online: https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing.

Arthur, Charles, "Twitter to introduce PhotoDNA system to block child abuse images", The Guardian (22 July 2013), online: https://www.theguardian.com/technology/2013/jul/22/twitter-photodna-child-abuse

Artificial intelligence-based system warns when a gun appears in a video", PhysOrg (Website) (7 July 2017), online: https://phys.org/news/2017-07-artificial-intelligence-based-gun-video.html

Asher, Jeff & Rob Arthur, "Inside the Algorithm That Tries to Predict Gun Violence in Chicago", The New York Times (13 June 2017), online:

https://www.nytimes.com/2017/06/13/upshot/what-an-algorithm-reveals-about-life-on-chicagos-high-risk-list.html

Assheuer, Thomas, "Die Big-Data-Diktatur", Die Zeit (29 November 2017), online: https://www.zeit.de/2017/49/china-datenspeicherung-gesichtserkennung-big-data-ueberwachung.

Atkinson, Robert, "'It's Going to Kill Us!' and Other Myths About the Future of Artificial Intelligence" (2016) Information Technology 50.

Babuta, Alexander, Marion Oswald, & Christine Rinik, "Machine learning algorithms and police decision-making: Legal, ethical and regulatory challenges" (2018) Whitehall Reports (21 September), at 5, online: https://rusi.org/publication/whitehall-reports/machine-learning-algorithms-and-police-decision-making-legal-ethical.

Banks, Alec, "What Are Deepfakes & Why the Future of Porn is Terrifying", Highsnobiety (20 December 2018), online: https://www.highsnobiety.com/p/what-are-deepfakes-ai-porn.

Barker, Colin, "How the GPU became the heart of AI and machine learning", ZDNet (13 August 2018), online: https://www.zdnet.com/article/how-the-gpu-became-the-heart-of-ai-and-machine-learning/; Bernard Fraenkel, "For Machine Learning, It's All About GPUs", Forbes (1 December 2017), online: https://www.forbes.com/sites/forbestechcouncil/2017/12/01/for-machine-learning-its-all-about-gpus/.

Barrett, David, "One surveillance camera for every 11 people in Britain, says CCTV survey", The Telegraph (10 July

2013), online:
https://www.telegraph.co.uk/technology/10172298/One-surveillance-camera-for-every-11-people-in-Britain-says-CCTV-survey.html

Bellamy, C. & J. Taylor, "New information and communications technologies and institutional change: The case of the UK criminal justice system," (1996) 9:4 International Journal of Public Sector Management 51.

Benbouzid, Bilel, "À qui profite le crime? Le marché de la prediction du crime aux États-Unis" (2016) La Vie des Idées, online at https://laviedesidees.fr/A-qui-profite-le-crime.html.

Benbouzid, Bilel, "Des crimes et des séismes: La police prédictive entre science, technique et divination", 6: 206 Réseaux 95.

Benedikt, Carl Frey & Michael A Osborne, "The future of employment: How susceptible are jobs to computerisation?" (2017) 114 Technological Forecasting and Social Change 254.

Berg, Nate, 'Predicting crime, LAPD-style', The Guardian (25 June 2014), online:
https://www.theguardian.com/cities/2014/jun/25/predicting-crime-lapd-los-angeles-police-data-analysis-algorithm-minority-report.

Best, Jo, "IBM Watson: The inside story of how the Jeopardy-winning supercomputer was born, and what it wants to do next", TechRepublic (9 September 2013), online:
https://www.techrepublic.com/article/ibm-watson-the-inside-story-of-how-the-jeopardy-winning-supercomputer-was-born-and-what-it-wants-to-do-next.

Bhushan, Kul, "Meet Staqu, a startup helping Indian law

enforcement agencies with advanced AI", Live Mint (26 June 2018), online: https://www.livemint.com/AI/DIh6fmR6croUJps6x7JW5K/Meet-Staqu-a-startup-helping-Indian-law-enforcement-agencie.html.

Biggio, Battista, Blaine Nelson & Pavel Laskov, "Poisoning Attacks against Support Vector Machines" (2012) arXiv Working Paper, arXiv:12066389 [cs, stat], online: http://arxiv.org/abs/1206.6389.

Bolukbasi, Tolga, et al, "Man is to Computer Programmer as Woman is to Homemaker? Debiasing Word Embeddings" (2016) arXiv Working Paper, arXiv:160706520 [cs, stat], online: http://arxiv.org/abs/1607.06520.

Borak, Masha, "China's public toilets now have facial recognition, thanks to Xi Jinping", Tech in Asia (21 December 2018), online: https://www.techinasia.com/chinas-public-toilets-facial-recognition-xi-jinping.

Bostrom, Nick, "Ethical Issues in Advanced Artificial Intelligence" (2003) Science Fiction and Philosophy: From Time Travel to Superintelligence.

Bragg, Lucia, "Federal Criminal Justice Reform in 2018" (2018) 26:10 LegisBrief, online: http://www.ncsl.org/research/civil-and-criminal-justice/federal-criminal-justice-reform-in-2018.aspx.

Breslin, Susannah, "Meet The Terrifying New Robot Cop That's Patrolling Dubai" Forbes (3 June 2017), online: https://www.forbes.com/sites/susannahbreslin/2017/06/03/

robot-cop-dubai/#287b11c96872.

Brewster, Thomas, "Apple Face ID 'Fooled Again' -- This Time By $200 Evil Twin Mask", Forbes (27 Novembere 2017), online: https://www.forbes.com/sites/thomasbrewster/2017/11/27/apple-face-id-artificial-intelligence-twin-mask-attacks-iphone-x/#7df1a8052775.

Buchanan, Bruce, "A (Very) Brief History of Artificial Intelligence" 26 AI Magazine (2005).

Bughin, Jacques et al., "Artificial Intelligence: The Next Digital Frontier?", McKinsey Global Institute (June 2017) online: https://www.mckinsey.com/~/media/McKinsey/Industries/Advanced%20Electronics/Our%20Insights/How%20artificial%20intelligence%20can%20deliver%20real%20value%20to%20companies/MGI-Artificial-Intelligence-Discussion-paper.ashx.

Burgess, Matt, "AI is invading UK policing, but there is little proof it's useful", Wired (21 September 2018), online at https://www.wired.co.uk/article/police-artificial-intelligence-rusi-report.

Burgess, Matt, "UK police are using AI to inform custodial decisions – but it could be discriminating against the poor", WIRED (1 March 2018), online: https://www.wired.co.uk/article/police-ai-uk-durham-hart-checkpoint-algorithm-edit.

Cadwalladr, Carole, "'I made Steve Bannon's psychological warfare tool': meet the data war whistleblower", The Guardian (18 March 2018), online: http://www.theguardian.com/news/2018/mar/17/data-war-

whistleblower-christopher-wylie-faceook-nix-bannon-trump.

Camille Polloni, "Police prédictive : la tentation de « dire quel sera le crime de demain", L'Obs (27 May 2015), online : https://www.nouvelobs.com/rue89/rue89-police-justice/20150527.RUE9213/police-predictive-la-tentation-de-dire-quel-sera-le-crime-de-demain.html.

Carpenter, Julia, "Google's algorithm shows prestigious job ads to men, but not to women. Here's why that should worry you.", Washington Post (6 July 2015), online: https://www.washingtonpost.com/news/the-intersect/wp/2015/07/06/googles-algorithm-shows-prestigious-job-ads-to-men-but-not-to-women-heres-why-that-should-worry-you/.

Chakraborty, R., "Sample size requirements for addressing the population genetic issues of forensic use of DNA typing" (1992) 64:2 Human Biology 141

Chen, Stephen, "China to build giant facial recognition database to identify any citizen within seconds", South China Morning Post (12 October 2017), online: https://www.scmp.com/news/china/society/article/2115094/china-build-giant-facial-recognition-database-identify-any.

Chesney, Robert & Danielle Keats Citron, "Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security", (2019) 107 California Law Review (forthcoming).

Chessen, Matt, "The Madcom Future: How Artificial Intelligence Will Enhance Computational Propaganda, Reprogram Human Culture, and Threaten Democracy...and What Can Be Done About It", The Atlantic Council (1 September 2017), online: https://www.scribd.com/document/359972969/The-MADCOM

-Future.

Cho, Charles et al, "Astroturfing Global Warming: It Isn't Always Greener on the Other Side of the Fence" (2011) 104:4 J Bus Ethics 571.

Chui, Michael, James Manyika & Mehdi Miremadi, "Where machines could replace humans--and where they can't (yet)", McKinsey Quarterly (July 2016), online: https://www.mckinsey.com/business-functions/digital-mckinsey/our-insights/where-machines-could-replace-humans-and-where-they-cant-yet.

Citron, Danielle Keats, "Technological Due Process" (2008), 85:6 Washington University Law Review 1249, online: https://openscholarship.wustl.edu/cgi/viewcontent.cgi?article=1166&context=law_lawreview.

Cloonan, John, "Advanced Malware Detection - Signatures vs. Behavior Analysis", Infosecurity Magazine (11 April 2017), online: https://www.infosecurity-magazine.com:443/opinions/malware-detection-signatures/.

Coldewey, Devin, "Taser rebrands as Axon and offers free body cameras to any police department", Tech Crunch (5 April 2017), online: https://techcrunch.com/2017/04/05/taser-rebrands-as-axon-and-offers-free-body-cameras-to-any-police-department/.

Cole, Samantha & Emanuel Maiberg, "People Are Using AI to Create Fake Porn of Their Friends and Classmates", Motherboard (26 January 2018), online: https://motherboard.vice.com/en_us/article/ev5eba/ai-fake-

porn-of-friends-deepfakes; Ruiz, Rebecca, "Deepfakes are about to make revenge porn so much worse" Mashable (24 June 2018), online: https://mashable.com/article/deepfakes-revenge-porn-domestic-violence/.

Copel, Michael, "The Difference Between AI, Machine Learning, and Deep Learning?", The Official NVIDIA Blog (29 July 2016), online: https://blogs.nvidia.com/blog/2016/07/29/whats-difference-artificial-intelligence-machine-learning-deep-learning-ai/.

Cristiani, Francesca, "How Lyrebird Uses AI to Find Its (Artificial) Voice", Wired (15 October 2018), online: https://www.wired.com/brandlab/2018/10/lyrebird-uses-ai-find-artificial-voice/

Curran, Dylan, "Are you ready? This is all the data Facebook and Google have on you", The Guardian (30 March 2018), online: http://www.theguardian.com/commentisfree/2018/mar/28/all-the-data-facebook-google-has-on-you-privacy.

Dastin, Jeffrey, "Amazon scraps secret AI recruiting tool that showed bias against women", Reuters (10 October 2018), online: https://www.reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKCN1MK08G.

David Gunning, "Explainable Artificial Intelligence (XAI)", DARPA (Website), online: https://www.darpa.mil/program/explainable-artificial-intelligence.

Daws, Ryan, "Chinese facial recognition flags bus ad woman for jaywalking", IoT News (28 November 2018), online: https://www.iottechnews.com/news/2018/nov/28/chinese-facial-recognition-ad-jaywalking/.

Dettmers, Tim, "Which GPU(s) to Get for Deep Learning", Tim Dettmers (5 November 2018), online: http://timdettmers.com/2018/11/05/which-gpu-for-deep-learning/.

Dietrich, William, Christina Mendoza & Tim Brennan "COMPAS Risk Scales: Demonstrating Accuracy Equity and Predictive Parity", Volaris Groupe (Website), online: http://go.volarisgroup.com/rs/430-MBX-989/images/ProPublica_Commentary_Final_070616.pdf.

Diop, Mouhamadou-Lamine, "Explainable AI: The data scientists' new challenge", Towards Data Science (14 June 2018), online: https://towardsdatascience.com/explainable-ai-the-data-scientists-new-challenge-f7cac935a5b4

Domingos, Pedro, "A few useful things to know about machine learning" (2012) 55:10 Communications of the ACM 78.

Douglas, T., J. Pugh, I. Singh, J. Savulescu, and S. Fazelb, "Risk assessment tools in criminal justice and forensic psychiatry: The need for better data" (2017) 42 Eur Psychiatry 134, online: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5408162/.

Dowd, Maureen, "Elon Musk's Billion-Dollar Crusade to Stop the A.I. Apocalypse", Hive - Vanity Fair (26 March 2017), online: https://www.vanityfair.com/news/2017/03/elon-musk-billion-dollar-crusade-to-stop-ai-space-x.

Drange, Matt, "We're Spending Millions On This High-Tech System Designed To Reduce Gun Violence. Is It Making A Difference?", Forbes (17 November 2016), online: https://www.forbes.com/sites/mattdrange/2016/11/17/shotspotter-struggles-to-prove-impact-as-silicon-valley-answer-to-gun-violence/#11ee763731cb.

Dressel, Julia & Hany Farid, "The accuracy, fairness, and limits of predicting recidivism" Science Advances (17 January 2018), online: http://advances.sciencemag.org/content/4/1/eaao5580.full.

Dumiak, Michael, "Interpol's New Software Will Recognize Criminals by Their Voices", IEEE Spectrum (16 May 2018), online: https://spectrum.ieee.org/tech-talk/consumer-electronics/audiovideo/interpols-new-automated-platform-will-recognize-criminals-by-their-voice.

Eisenstein, Paul A., "Millions of jobs are on the line when autonomous cars take over", NBC News (5 November 2017), online: https://www.nbcnews.com/business/autos/millions-professional-drivers-will-be-replaced-self-driving-vehicles-n817356.

Ericson, R., & K. Haggerty, Policing the risk society (Oxford: Clarendon Press, 1997); A. Amicelle, C. Aradau, & J. Jeandesboz, "Questioning security devices: Performativity, resistance, politics," (2015) 46:4 Security Dialogue 293.

Eubanks, Virginia, Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor (New York: NY, St Martin's Press, 2017).

Eykholt, Kevin et al, "Robust Physical-World Attacks on Deep Learning Models" (2017) arXiv Working Paper, arXiv:1707.08945 [cs.CR] online: http://arxiv.org/abs/1707.08945.

Faggella, Daniel, "The AI Advantage of the Tech Giants: Amazon, Facebook, and Google", TechEmergence (24 November 2018), online: https://www.techemergence.com/the-ai-advantage-of-the-tech-giants-amazon-facebook-and-google-etc/.

FortiGuard SE Team, "Predictions: AI Fuzzing and Machine Learning Poisoning", Fortinet Blog (15 November 2018), online: https://www.fortinet.com/blog/industry-trends/predictions-ai-fuzzing-and-machine-learning-poisoning-.html.

Fry, Hannah, Hello World: Being Human in the Age of the Machine (New York, NY: W.W. Norton, 2018).

Fuller, Thomas, "California Is the First State to Scrap Cash Bail", The New York Times (28 August 2018), online: https://www.nytimes.com/2018/08/28/us/california-cash-bail.html; Rebecca Ibarra, "New Jersey's Bail Reform Law Gets Court Victory", WNYC (9 July 2018), online: https://www.wnyc.org/story/new-jerseys-bail-reform-law-gets-court-victory/.

García-Teodoro, P. et al, "Anomaly-based network intrusion detection: Techniques, systems and challenges" (2009) 28:1 –2 Computers & Security 18

Garland, David, The culture of control: Crime and social order in contemporary society (Oxford: Oxford University Press, Oxford, 2001).

Gatys, Leon A, Alexander S Ecker & Matthias Bethge, "A Neural Algorithm of Artistic Style" (2015) arXiv Working Paper, arXiv:150806576 [cs, q-bio], online: http://arxiv.org/abs/1508.06576; "Deep Dream Generator", Deep Dream Generator (Website), online: https://deepdreamgenerator.com/.

Ghoshal, Abhimanyu, "I trained an AI to copy my voice and it scared me silly", The Next Web (22 January 2018), online: https://thenextweb.com/insights/2018/01/22/i-trained-an-ai-to-copy-my-voice-and-scared-myself-silly/.

Gibbs, Samuel "AlphaZero AI beats champion chess program after teaching itself in four hours", The Guardian (7 December 2017), online: https://www.theguardian.com/technology/2017/dec/07/alphazero-google-deepmind-ai-beats-champion-program-teaching-itself-to-play-four-hours.

Goertzel, Ben, Matt Iklé & Jared Wigmore, "The Architecture of Human-Like General Intelligence" in Pei Wang & Ben Goertzel, eds, Theoretical Foundations of Artificial General Intelligence (Paris: Atlantis Press, 2012).

Grauer, Yael & Emanuel Maiberg, "What Are 'Data Brokers,' and Why Are They Scooping Up Information About You?", VICE Motherboard (27 March 2018), online: https://motherboard.vice.com/en_us/article/bjpx3w/what-are-data-brokers-and-how-to-stop-my-private-data-collection.

Greene, Dan & Genevieve Patterson, "The Trouble With Trusting AI to Interpret Police Body-Cam Video", IEEE Spectrum (21 November 2018), online: https://spectrum.ieee.org/computing/software/the-troubl

e-with-trusting-ai-to-interpret-police-bodycam-video

Gretchen Greene, K., "Buying you first AI or 'never trust a used algorithm salesman'", Berkman Klein Center for Internet & Society — AI Ethics & Governance (7 November 2018), online: https://medium.com/berkman-klein-center/buying-your-first-ai-136cd2e6dd2.

Grieco, Gustavo & Artem Dinaburg, "Toward Smarter Vulnerability Discovery Using Machine Learning". (Paper delivered at the Proceedings of the 11th ACM Workshop on Artificial Intelligence and Security, Toronto, Canada, 2018)

Grosse, Kathrin et al, "Adversarial Perturbations Against Deep Neural Networks for Malware Classification" (2016) arXiv Working Paper, arXiv:160604435 [cs], online: http://arxiv.org/abs/1606.04435.

Gunning, David, "Explainable Artificial Intelligence (XAI): Technical Report", (2016) Defense Advanced Research Projects Agency DARPA-BAA-16-53; Sandra Wachter, Mittelstadt, Brent & Chris Russell, "Counterfactual Explanations without Opening the Black Box: Automated Decisions and the GDPR", (2017) arXiv Working Paper, arXiv:1711.00399 [cs.AI], online: https://arxiv.org/abs/1711.00399.

Guzmán, Arelis, "Top 10 Pretrained Models to get you Started with Deep Learning (Part 1 - Computer Vision)", Analytics Vidhya (27 July 2018), online: https://www.analyticsvidhya.com/blog/2018/07/top-10-pretrained-models-get-started-deep-learning-part-1-computer-vision/.

Hannah-Moffat, Kelly, “Actuarial Sentencing: An ‘Unsettled’ Proposition” (2013) 30:2 Justice Quarterly 270-296, DOI: 10.1080/07418825.2012.682603; Mark H. Bergstrom & Kern, Richard P., “A View from the Field: Practitioners' Response to Actuarial Sentencing: An ‘Unsettled’ Proposition” (2013) 25:3 Federal Sentencing Reporter 185.

Harp, Steven et al, “Automated Vulnerability Analysis Using AI Planning” (Paper delivered at the 2005 AAAI Spring Symposium, Stanford, CA, 2018), online: https://www.researchgate.net/publication/221250445_Automated_Vulnerability_Analysis_Using_AI_Planning.

Hawkins, Andrew J. ≤ “Waymo is first to put fully self-driving cars on US roads without a safety driver”, The Verge (7 November 2017), online: https://www.theverge.com/2017/11/7/16615290/waymo-self-driving-safety-driver-chandler-autonomous.

Hern, Alex, “Apple: don't use Face ID on an iPhone X if you're under 13 or have a twin”, The Guardian (27 September 2017), online: https://www.theguardian.com/technology/2017/sep/27/apple-face-id-iphone-x-under-13-twin-facial-recognition-system-more-secure-touch-id.

Hill, Kashmir, “How Target Figured Out A Teen Girl Was Pregnant Before Her Father Did”, Forbes (16 February 2012), online: https://www.forbes.com/sites/kashmirhill/2012/02/16/how-target-figured-out-a-teen-girl-was-pregnant-before-her-father-did/.

Hitaj, Briland et al, “PassGAN: A Deep Learning Approach for

Password Guessing" (2017) arXiv Working Paper, arXiv:170900440 [cs, stat], online: http://arxiv.org/abs/1709.00440.

Holt, Tom, "Exploring the social organisation and structure of stolen data markets", (2013) 14:2-3 Global Crime 155; Alice Hutchings and Tom Holt, "A crime script analysis of the online stolen data market", (2015) 55:3 The British Journal of Criminology 596; "McAfee Labs 2017 Threats Predictions Report", McAfee (Website), online: https://www.mcafee.com/enterprise/en-us/assets/reports/rp-threats-predictions-2017.pdf.
http://www.marinusanalytics.com/articles/2017/6/27/face-search-debut.
https://megaface.cs.washington.edu/results/fgnetresults.html.
https://www.shotspotter.com/press-releases/chicago-signs-23-million-multi-year-agreement-with-shotspotter-to-extend-gunshot-detection-coverage-into-next-decade/.

Information & Communications Technology Law 223, online : https://www.tandfonline.com/doi/pdf/10.1080/13600834.2018.1458455.

Israni, Ellora, "Algorithmic Due Process: Mistaken Accountability and Attribution in State v. Loomis" Jolt Digest (31 August 2017), online: https://jolt.law.harvard.edu/digest/algorithmic-due-process-mistaken-accountability-and-attribution-in-state-v-loomis-1, "Taking Algorithms To Court Current Strategies for Litigating Government Use of Algorithmic Decision-Making", AI Now Institute (24 September 2018), online: https://medium.com/@AINowInstitute/taking-algorithms-to-court-7b90f82ffcc9

Jing, Meng, "Chinese home sharing site Xiaozhu to roll out facial recognition-enabled smart locks in Chengdu pilot scheme", South China Morning Post (26 December 2018), online: https://www.scmp.com/tech/start-ups/article/2179495/chinese-home-sharing-site-xiaozhu-roll-out-facial-recognition-enabled.

Johnson, Alistair, et al, "MIMIC-III, a freely accessible critical care database" (2016) 3 Scientific Data.

Kao, Jeff, "More than a Million Pro-Repeal Net Neutrality Comments Were Likely Faked", Hacker Noon (23 November 2017), online: https://hackernoon.com/more-than-a-million-pro-repeal-net-neutrality-comments-were-likely-faked-e9f0e3ed36a6.

Karras, Tero, et al, "Progressive Growing of Gans for Improved Quality, Stability, and Variation" (2018) arXiv Working Paper, arXiv:1710.10196 [cs.NE], online: https://arxiv.org/abs/1710.10196.

Khurana, Nitika, Sudip Mittal & Anupam Joshi, "Preventing Poisoning Attacks on AI based Threat Intelligence Systems" (2018), arXiv Working Paper, arXiv:1807.07418 [cs.SI], online: https://arxiv.org/abs/1807.07418v1

Kofman, Ava, "Interpol Rolls Out International Voice Identification Database Using Samples From 192 Law Enforcement Agencies", The Intercept (25 June 2018), online: https://theintercept.com/2018/06/25/interpol-voice-identification-database/

Korolov, Maria, "Hackers get around AI with flooding, poisoning and social engineering", CSO Online (16 December 2016), online:

https://www.csoonline.com/article/3150745/security/hackers-get-around-ai-with-flooding-poisoning-and-social-engineering.html.

Kosinski, Michal, David Stillwell & Thore Graepel, "Private traits and attributes are predictable from digital records of human behavior" (2013) 110:15 Proceedings of the National Academy of Sciences 5802.

Kraska, Peter B., "Militarization and policing: Its relevance to 21st Century police", (2007) 1:4 Policing: A Journal of Policy and Practice 501.

Krebs, Brian, "Buying Battles in the War on Twitter Spam", Krebs on Security (Website) online: https://krebsonsecurity.com/2013/08/buying-battles-in-the-war-on-twitter-spam/.

Krebs, Brian, "Voice Phishing Scams Are Getting More Clever", Krebs on Security (Website), online: https://krebsonsecurity.com/2018/10/voice-phishing-scams-are-getting-more-clever/.

Langston, Jennifer, "How PhotoDNA for Video is being used to fight online child exploitation", Microsoft On the Issues (12 September 2018), online: https://news.microsoft.com/on-the-issues/2018/09/12/how-photodna-for-video-is-being-used-to-fight-online-child-exploitation/.

Larson, Jeff, Surya Mattu, Lauren Kirchner & Julia Angwin, "How We Analyzed the COMPAS Recidivism Algorithm", ProPublica (23 May 2016), online: https://www.propublica.org/article/how-we-analyzed-the-

compas-recidivism-algorithm.

Latonero, Mark, "Governing Artificial Intelligence: Upholding Human Rights & Dignity", Data & Society Research Institute (10 October 2018), online: https://datasociety.net/output/governing-artificial-intelligence/.

LeCun, Yann, Yoshua Bengio & Geoffrey Hinton, "Deep learning" (2015) 521:7553 Nature 436

Lee, Dave, "Why Big Tech pays poor Kenyans to programme self-driving cars", BBC (3 November 2018), online: https://www.bbc.com/news/technology-46055595.

Lee, Dave, "Why Big Tech pays poor Kenyans to programme self-driving cars", BBC (3 November 2018), online: https://www.bbc.com/news/technology-46055595.

Lee, Kai-Fu, ed, AI Superpowers: China, Silicon Valley, and the New World Order, (New York, NY: Houghton Mifflin Harcourt, 2018).

Leplâtre, Simon, "En Chine, la reconnaissance faciale envahit le quotidian", Le Monde (9 December 2017), online: https://www.lemonde.fr/economie/article/2017/12/09/en-chine-la-reconnaissance-faciale-envahit-le-quotidien_5227160_3234.html.

Liptak, Adam, "Sent to prison by a software program's secret algorithm", The New York Times (1 May 2017), online: https://www.nytimes.com/2017/05/01/us/politics/sent-to-prison-by-a-software-programs-secret-algorithms.html.

Liu, B. et al, "Software Vulnerability Discovery Techniques: A Survey" (Paper delivered at the Fourth International Conference on Multimedia Information Networking and

Security, 2012), online:
https://ieeexplore.ieee.org/document/6405650.

Liu, Joyce, "In Your Face: China's all-seeing state", BBC News (10 December 2017), online:
https://www.bbc.com/news/av/world-asia-china-42248056/in-your-face-china-s-all-seeing-state.

Lubin, Gus, "'Facial-profiling' could be dangerously inaccurate and biased, experts warn", Business Insider (12 October 2016), online:
https://www.businessinsider.com/does-faception-work-2016-10

Lunden, Ingrid, "Element AI, a platform for companies to build AI solutions, raises $102M", TechCrunch (November 2016), online:
http://social.techcrunch.com/2017/06/14/element-ai-a-platform-for-companies-to-build-ai-solutions-raises-102m.

Lyon, Thomas P & John W Maxwell, "Astroturf: Interest Group Lobbying and Corporate Strategy" (2004) 13:4 J Econ Manag Strategy 561; Kevin Grandia, "Bonner & Associates: The Long and Undemocratic History of Astroturfing", Huffington Post (26 August 2009), online:
https://www.huffingtonpost.com/kevin-grandia/bonner-associates-the-lon_b_269976.html.

Maas, Dave, "FBI Wish List: An App That Can Recognize the Meaning of Your Tattoos", EFF Deep Links (16 July 2018), online:
https://www.eff.org/deeplinks/2018/07/fbi-wants-app-can-recognize-meaning-your-tattoos

Maffeo, L., "The case for open source classifiers in AI algorithms", opensource.com (18 October 2018), online: https://opensource.com/article/18/10/open-source-classifiers-ai-algorithms.

Mahapatra, Sambit, "Why Deep Learning over Traditional Machine Learning?", Towards Data Science (21 March 2018), online: https://towardsdatascience.com/why-deep-learning-is-needed-over-traditional-machine-learning-1b6a99177063.

Mann, Ian, Hacking the human: Social engineering techniques and security countermeasures, (London: Routledge, 2008).

Manning, P. K., The technology of policing: Crime mapping, information technology, and the rationality of crime control (New York: NY, New York University Press, 2008).

Marcus, Gary, "Deep Learning: A Critical Appraisal" (2018) arXiv Working Paper, arXiv:1801.00631 [cs.AI], online: https://arxiv.org/abs/1801.00631.

Marr, Bernard, "The Fascinating Ways Facial Recognition AIs Are Used In China", Forbes (17 December 2018), online: https://www.forbes.com/sites/bernardmarr/2018/12/17/the-amazing-ways-facial-recognition-ais-are-used-in-china/#5842e21c5fa5.

McCormick, Rich, "Google scans everyone's email for child porn, and it just got a man arrested", The Verge (5 August 2014), online: https://www.theverge.com/2014/8/5/5970141/how-google-scans-your-gmail-for-child-porn.

Mennell, Julie, "Technology Supporting Crime Detection: An

Introduction" (2012) 45:12 Measurement + Control 304.

Mercer, Christina & Thomas Macaulay, "How tech giants are investing in artificial intelligence", Techworld (27 November 2018), online: https://www.techworld.com/picture-gallery/data/tech-giants-investing-in-artificial-intelligence-3629737.

Meyer, Robinson, "My Facebook Was Breached by Cambridge Analytica. Was Yours?", The Atlantic (10 April 2018), online: https://www.theatlantic.com/technology/archive/2018/04/facebook-cambridge-analytica-victims/557648/.

Mitchell, Tom, Machine Learning, (New York: McGraw-Hill Education, 1997).

Moon, Louise, "Pay attention at the back: Chinese school installs facial recognition cameras to keep an eye on pupils "South China Morning Post (16 March 2018), online: https://www.scmp.com/news/china/society/article/2146387/pay-attention-back-chinese-school-installs-facial-recognition.

Mozur, Paul, "Inside China's Dystopian Dreams: A.I., Shame and Lots of Cameras", The New York Times (8 July 2018), online: https://www.nytimes.com/2018/07/08/business/china-surveillance-technology.html.

MSV, Janakiram, "Why Do Developers Find It Hard To Learn Machine Learning?", Forbes (1 January 2018), online: https://www.forbes.com/sites/janakirammsv/2018/01/01/why-do-developers-find-it-hard-to-learn-machine-learning/.

Müller, Vincent, ed, Risks of Artificial Intelligence (Florida: Chapman and Hall/CRC Press, 2015).

Murphy, Finn, "Truck drivers like me will soon be replaced by automation. You're next", The Guardian (17 November 2017), online: https://www.theguardian.com/commentisfree/2017/nov/17/truck-drivers-automation-tesla-elon-musk.

Newton, Casey, "Microsoft sounds an alarm over facial recognition technology", The Verge (7 December 2018), online: https://www.theverge.com/2018/12/7/18129858/microsoft-facial-recognition-ai-now-google.

Nilsson, Nils J., The Quest for Artificial Intelligence (Cambridge, UK: Cambridge University Press, 2013).

Novikov, Ivan, "How AI Can Be Applied To Cyberattacks", Forbes (22 March 2018), online: https://www.forbes.com/sites/forbestechcouncil/2018/03/22/how-ai-can-be-applied-to-cyberattacks/.

Oberoi, Gaurav, "Exploring DeepFakes", Hacker Noon (5 March 2018), online: https://hackernoon.com/exploring-deepfakes-20c9947c22d9.

Ocbazghi, Emmanuel, "We put the iPhone X's Face ID to the ultimate test with identical twins — and the results surprised us" Business Insider (31 October 2017), online: https://www.businessinsider.com/can-iphone-x-tell-difference-between-twins-face-id-recognition-apple-2017-10.

Olafenwa, Moses "Object Detection with 10 lines of code", Towards Data Science (16 June 2018), online: https://towardsdatascience.com/object-detection-with-10-lines-of-code-d6cb4d86f606.

Olsen, Dana, "2017 Year in Review: The top VC rounds &

investors in AI", PitchBook News & Analysis (20 December 2017), online: https://pitchbook.com/news/articles/2017-year-in-review-the-top-vc-rounds-investors-in-ai.

Oswald, Marion, et al, "Algorithmic risk assessment policing models: lessons from the Durham HART model and 'Experimental' proportionality" (2018) 27:2

Palin, Megan, "Big Brother: China's chilling dictatorship moves to introduce scorecards to control everyone", news.com.au (19 September 2018), online: https://www.news.com.au/technology/online/big-brother-chinas-chilling-dictatorship-moves-to-introduce-scorecards-to-control-everyone/news-story/6c821cbf15378ab0d3eeb3ec3dc98abf.

Papernot, Nicolas et al, "Practical Black-Box Attacks against Machine Learning" (2016) arXiv Working Paper, arXiv: 160202697 [cs], online: http://arxiv.org/abs/1602.02697.

Paquet-Clouston, Masarah, Bernhard Haslhofer & Benoît Dupont, "Ransomware payments in the bitcoin ecosystem", (Paper delivered at the 17th Annual Workshop on the Economics of Information Security (WEIS), 2018) online: https://arxiv.org/abs/1804.04080.

Pasquale, Frank, "Secret Algorithms Threaten the Rule of Law", MIT Technology Review (1 June 2017), online: https://www.technologyreview.com/s/608011/secret-algorithms-threaten-the-rule-of-law/.

Pasternack, Alex, "Body camera maker will let cops live-stream their encounters", Fast Company (10 August 2018), online:

https://www.fastcompany.com/90247228/axon-new-body-cameras-will-live-stream-police-encounters.

Patton, Jessica, "What is 'ShotSpotter'? Controversial gunshot detector technology approved by Toronto police", Global News (20 July 2018), online: https://globalnews.ca/news/4344093/controversial-gunshot-detector-shotspotter-toronto-police/; "Chicago Signs $23 Million Multi-Year Agreement With Shotspotter to Extend Gunshot Detection Coverage Into Next Decade", ShotSpotter (Website) (5 September 2018), online:

Pearson, Jordan, "Toronto Approves Gunshot-Detecting Surveillance Tech Days After Mass Shooting", VICE Motherboard (25 July 2018), online: https://motherboard.vice.com/en_us/article/7xqk44/toronto-approves-shotspotter-gunshot-detecting-surveillance-tech-danforth-shooting.

Penney, Jon et al., "Advancing Human-Rights-By-Design In The Dual-Use Technology Industry", Columbia Journal of International Affairs (August 2018), online: https://jia.sipa.columbia.edu/advancing-human-rights-design-dual-use-technology-industry.

Perry, Nancy, "How Axon is accelerating tech advances in policing", Police One (Blog) (22 June 2018), online: https://www.policeone.com/police-products/body-cameras/articles/476840006-How-Axon-is-accelerating-tech-advances-in-policing/

Piekniewski, Filip, "AI winter is well on its way", Piekniewski's Blog (28 May 2018), online: https://blog.piekniewski.info/2018/05/28/ai-winter-is-well-

on-its-way.

Ranum, Marcus, "It's Worse Than You Think: Robo-Profiling", Free Thought Blogs (16 March 2017), online: https://freethoughtblogs.com/stderr/2017/03/16/its-worse-than-you-think-robo-profiling/.

Reedy, Christiana, "Kurzweil Claims That the Singularity Will Happen by 2045", Futurism (5 October 2017), online: https://futurism.com/kurzweil-claims-that-the-singularity-will-happen-by-2045.

Regalado, Antonio, "Investigators searched a million people's DNA to find Golden State serial killer", MIT Technology Review (27 April 2018), online: https://www.technologyreview.com/s/611038/investigators-searched-a-million-peoples-dna-to-find-golden-state-serial-killer/.

Revell, Timothy, "Computer vision algorithms pick out petty crime in CCTV footage", NewScientist (4 January 2017), online: https://www.newscientist.com/article/2116970-computer-vision-algorithms-pick-out-petty-crime-in-cctv-footage/.

Rieland, Randy, "Artificial Intelligence Is Now Used to Predict Crime. But Is It Biased?", Smithsonian Magazine (5 March 2018), online: https://www.smithsonianmag.com/innovation/artificial-intelligence-is-now-used-predict-crime-is-it-biased-180968337/.

Rinaldi, Eva, "Reese Witherspoon", Flickr (Website), online: https://goo.gl/a2sCdc.

Rinaldi, Eva, "Russell Crowe", Flickr (Website), online: https://goo.gl/AO7QYu.

Roose, Kevin, “Here Come the Fake Videos, Too”, The NY Times (8 June 2018), online: https://www.nytimes.com/2018/03/04/technology/fake-videos-deepfakes.html.

Rosenfeld, Amir, Richard Zemel & John K Tsotsos, “The Elephant in the Room” (2018) arXiv Working Paper, arXiv:1808.03305 [cs.CV] online: http://arxiv.org/abs/1808.03305.

Rubinstein, Benjamin IP et al, “ANTIDOTE: understanding and defending against poisoning of anomaly detectors” (Paper delivered at the 9th ACM SIGCOMM Conference on Internet Measurement, 2009), online: https://people.eecs.berkeley.edu/~tygar/papers/SML/IMC.2009.pdf

Rushe, Dominic, “End of the road: will automation put an end to the American trucker?”, The Guardian (10 October 2017), online: https://www.theguardian.com/technology/2017/oct/10/american-trucker-automation-jobs.

Russell, Jon, “China’s CCTV surveillance network took just 7 minutes to capture BBC reporter”, Tech Crunch (13 December 2017), online: https://techcrunch.com/2017/12/13/china-cctv-bbc-reporter/.

Ryan, Julie J.C.H., “How do computer hackers ‘get inside’ a computer?”, Scientific American, online: https://www.scientificamerican.com/article/how-do-computer-hackers-g/.

Schroff, Florian, Dmitry Kalenichenko & James Philbin, “FaceNet: A Unified Embedding for Face Recognition and Clustering”,

arXiv Working Paper, arXiv:1503.03832v3 [cs.CV], online: https://arxiv.org/pdf/1503.03832.pdf

Schwartz, Oscar, "You thought fake news was bad? Deep fakes are where truth goes to die", The Guardian (12 November 2018), online: https://www.theguardian.com/technology/2018/nov/12/deep-fakes-fake-news-truth.

Sejnowski, Terrence, The Deep Learning Revolution (Cambridge, Massachusetts: The MIT Press, 2018).

Seymour, John & Philip Tully, "Weaponizing data science for social engineering: Automated E2E spear phishing on Twitter" (Paper delivered at Black Hat USA 2016, DEF CON 24, 2016), online: 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.

Sharif, Mahmood, Sruti Bhagavatula, Lujo Bauer & Michael K. Reiter, "Accessorize to a Crime: Real and Stealthy Attacks on State-of-the-Art Face Recognition", Conference Paper (October 2016), online: https://www.cs.cmu.edu/~sbhagava/papers/face-rec-ccs16.pdf

Shead, Sam, "Google's Mysterious AI Ethics Board Should Be Transparent Like Axon's", Forbes (27 April 2018), online: https://www.forbes.com/sites/samshead/2018/04/27/googles-mysterious-ai-ethics-board-should-be-as-transparent-as-axons/#12e80d0019d1.

Shenfield, Alex, David Day & Aladdin Ayesh, "Intelligent intrusion

detection systems using artificial neural networks" (2018) 4:2 ICT Express 95.

Shoham, Yoav, Raymond Perrault, Erik Brynjolfsso & Jack Clark, "Artificial Intelligence Index: 2017 Annual Report", AI Index (November 2017) online: http://cdn.aiindex.org/2017-report.pdf.

Silver, David & Demis Hassabis, Cade Metz, "In Two Moves, AlphaGo and Lee Sedol Redefined the Future", Wired (16 March 2016), online: https://www.wired.com/2016/03/two-moves-alphago-lee-sedol-redefined-future/.

Simonite, Tom, "AI and 'Enormous Data' Could Make Tech Giants Like Google Harder to Topple", Wired (13 July 2017), online: https://www.wired.com/story/ai-and-enormous-data-could-make-tech-giants-harder-to-topple/.

Solon, O.. "The rise of 'pseudo-AI': how tech firms quietly use humans to do bots' work", The Guardian (6 July 2018), online: https://www.theguardian.com/technology/2018/jul/06/artificial-intelligence-ai-humans-bots-tech-companies.

Stanford, Stacy, "The 50 Best Public Datasets for Machine Learning", Data Driven Investor (2 October 2018), online: https://medium.com/datadriveninvestor/the-50-best-public-datasets-for-machine-learning-d80e9f030279.

Stanley, Jay, "Secret Service Announces Test of Face Recognition System Around White House", ACLU Free Future (4 December 2018), online:

https://www.aclu.org/blog/privacy-technology/surveillance-technologies/secret-service-announces-test-face-recognition.

Stats NZ, "Algorithm assessment report" (Wellington: New Zealand Government, 2018) online: https://www.data.govt.nz/assets/Uploads/Algorithm-Assessment-Report-Oct-2018.pdf.

Streitfeld, David, "Book Reviewers for Hire Meet a Demand for Online Raves", The New York Times (25 August 2012), online: https://www.nytimes.com/2012/08/26/business/book-reviewers-for-hire-meet-a-demand-for-online-raves.html.

Su, J. D. Vasconcellos Vargas, & K. Sakurai, "One pixel attack for fooling deep neural networks", (2017) arXiv Working Paper, arXiv:1710.08864 [cs.LG], online at https://arxiv.org/abs/1710.08864v4.

Subramanian, Sankar, "The effects of sample size on population genomic analyses – implications for the tests of neutrality" (2016) 17:123 BMC Genomics.

Swaine, Jon, "Russian propagandists targeted African Americans to influence 2016 US election", The Guardian (17 December 2018), online: https://www.theguardian.com/us-news/2018/dec/17/russian-propagandists-targeted-african-americans-2016-election.

Szegedy, Christian, et al., "Intriguing properties of neural networks" (2013) arXiv Working Paper, arXiv:1312.6199 [cs.CV], online: http://arxiv.org/abs/1312.6199.

The Privacy Expert's Guide to Artificial Intelligence and Machine Learning (Future of Privacy forum, 2018).

Usersub, "Nick Cage DeepFakes Movie Compilation", online: https://www.youtube.com/watch?time_continue=25&v=BU9YAHigNx8.

Uz, Fidan Boylu, "GPUs vs CPUs for deployment of deep learning models", Microsoft Azure (11 September 2018), online: https://azure.microsoft.com/en-us/blog/gpus-vs-cpus-for-deployment-of-deep-learning-models/.

VanNostrand, Marie & Kenneth Rose, "Pretrial Risk Assessment In Virginia", Virginia Department of Criminal Justice (Website), online: https://www.dcjs.virginia.gov/sites/dcjs.virginia.gov/files/publications/corrections/virginia-pretrial-risk-assessment-report.pdf.

Vincent, James & Russell Brandom, "Axon launches AI ethics board to study the dangers of facial recognition", The Verge (26 April 2018), online: https://www.theverge.com/2018/4/26/17285034/axon-ai-ethics-board-facial-recognition-racial-bias.

Vincent, James, "This is when AI's top researchers think artificial general intelligence will be achieved", The Verge (27 November 2018), online: https://www.theverge.com/2018/11/27/18114362/ai-artificial-general-intelligence-when-achieved-martin-ford-book.

Vincent, James, "Twitter taught Microsoft's friendly AI chatbot to be a racist asshole in less than a day", The Verge (24 March 2016), online: https://www.theverge.com/2016/3/24/11297050/tay-microsoft-chatbot-racist.

Vincent, James, "Watch Jordan Peele use AI to make Barack Obama deliver a PSA about fake news", The Verge, (17 April 2018), online: https://www.theverge.com/tldr/2018/4/17/17247334/ai-fake-news-video-barack-obama-jordan-peele-buzzfeed.

Voss, Peter, "From Narrow to General AI", Intuition Machine 3 October 2017), online: https://medium.com/intuitionmachine/from-narrow-to-general-ai-e21b568155b9

Votipka, Daniel et al, "Hackers vs. Testers: A Comparison of Software Vulnerability Discovery Processes" (Paper delivered at the 2018 IEEE Symposium on Security and Privacy, San Francisco, CA, 2018), online: https://ieeexplore.ieee.org/document/8418614.

Wagner, David & Paolo Soto, "Mimicry Attacks on Host-Based Intrusion Detection Systems" (Paper delivered at the 9th ACM conference on Computer and communications security, Washington DC, 2002), online: https://dl.acm.org/citation.cfm?id=586145 at 10.

Washan, Nitin & Sandeep Sharma, "Speech Recognition System: A Review" 115:18 International Journal of Computer Applications 7, online: https://pdfs.semanticscholar.org/8f2c/b3f70bb75b6235514b192b83e413a0e23dd8.pdf.

Weller, Chris, "There's a secret technology in 90 US cities that listens for gunfire 24/7", Business Insider (27 June 2017), online: https://www.businessinsider.com/how-shotspotter-works-microphones-detecting-gunshots-2017-6.

Whittaker , Meredith et al., “AI Now Report”, AI Now Institute (December 2018), online: https://ainowinstitute.org/AI_Now_2018_Report.pdf; AI Now Institute, “After a Year of Tech Scandals, Our 10 Recommendations for AI”, AI Now Institute (6 December 2018), online: https://medium.com/@AINowInstitute/after-a-year-of-tech-scandals-our-10-recommendations-for-ai-95b3b2c5e5.

Winick, Erin, “Lawyer-bots are shaking up jobs”, MIT Technology Review (12 December 2017), online: https://www.technologyreview.com/s/609556/lawyer-bots-are-shaking-up-jobs/.

Winston, Ali & Ingrid Burrington, “A pioneer in predictive policing is starting a troubling new project”, The Verge (26 April 2018), online: https://www.theverge.com/2018/4/26/17285058/predictive-policing-predpol-pentagon-ai-racial-bias.

Wolff, Honorable Michael A, “Evidence-Based Judicial Discretion: Promoting Public Safety Through State Sentencing Reform” (2008) 83:5 New York University Law Review 1389, online: https://www.nyulawreview.org/wp-content/uploads/2018/08/NYULawReview-83-5-Wolff.pdf.

Yuan, Xiaoyong et al, “Adversarial Examples: Attacks and Defenses for Deep Learning” (2017) arXiv Working Paper, arXiv:171207107 [cs, stat], online: http://arxiv.org/abs/1712.07107.

Zenaida Kotala, “Orlando Crime Scene Video Analysis Goes High-Tech With $1.3 Million Grant to UCF” Space Coast Daily,19 April 2016), online)

http://spacecoastdaily.com/2016/04/orlando-crime-scene-video-analysis-goes-high-tech-with-1-3-million-grant-to-ucf/.

Zubairi, Amira, "Magnet Forensics launches Magnet.AI to fight child exploitation", Betakit (Website) (16 May 2017), online: https://betakit.com/magnet-forensics-launches-magnet-ai-to-fight-child-exploitation/.

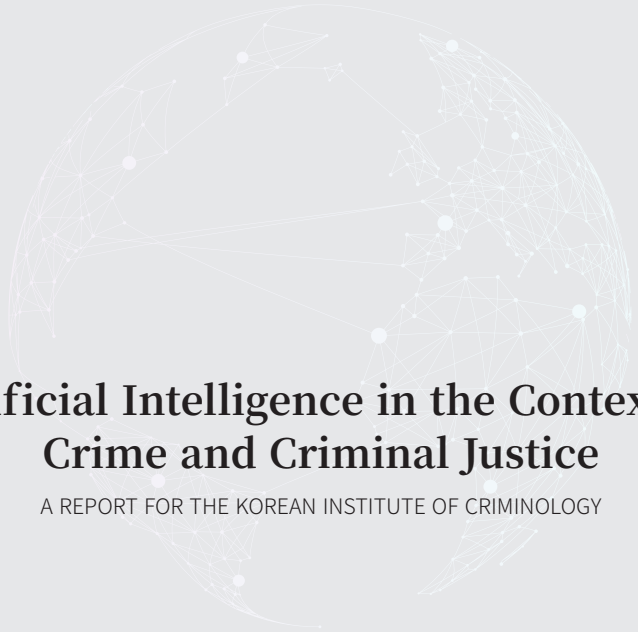
Zucconi, Alan, "Understanding the Technology Behind DeepFakes", Alan Zucconi (14 March 2018), online: https://www.alanzucconi.com/2018/03/14/understanding-the-technology-behind-deepfakes/.

Artificial Intelligence in the Context of Crime and Criminal Justice

First Published July 19, 2019

Printed in Seoul, Korea
by Korean Institute of Criminology
Registered in March 20, 1990 21-143
+82-2-575-5282
https://eng.kic.re.kr/
KRW 10,000

ISBN 979-11-89908-25-6



Artificial Intelligence in the Context of Crime and Criminal Justice

A REPORT FOR THE KOREAN INSTITUTE OF CRIMINOLOGY

KIC | Korean Institute of Criminology