The Dark Side of Artificial Intelligence: Challenges for the Legal System

Willem Gravett
https://orcid.org/0000-0001-7400-0036
Associate Professor, Faculty of Law, University of Pretoria
willem.gravett@up.ac.za

Abstract

The development of artificial intelligence has the potential to transform lives and work practices, raise efficiency, savings and safety levels, and provide enhanced levels of services. However, the current trend towards developing smart and autonomous machines with the capacity to be trained and make decisions independently holds not only economic advantages, but also a variety of concerns regarding their direct and indirect effects on society as a whole. This article examines some of these concerns, specifically in the areas of privacy and autonomy, state surveillance, and bias and algorithmic transparency. It concludes with an analysis of the challenges that the legal system faces in regulating the burgeoning field of artificial intelligence.

Keywords: artificial intelligence; privacy; state surveillance; bias; algorithmic transparency; algorithmic opacity; regulation

UNISA university of south africa PRESS

Southern African Public Law
https://upjournals.co.za/index.php/SAPL
Volume 35 | Number 1 | 2020 | #6979 | 30 pages

https://doi.org/10.25159/2522-6800/6979
ISSN 2522-6800 (Online), 2219-6412 (Print)

Introduction

There can be no doubt that development of robotics and artificial intelligence (AI) has the potential to transform lives and work practices, raise efficiency, savings and safety levels, and provide enhanced levels of services in the short to medium term. Robotics and AI promise to bring benefits of efficiency and savings not only in production and commerce, but also in areas such as transportation, medical care, rescue, education and farming. At the same time, they make it possible to avoid exposing humans to dangerous conditions, such as those faced when cleaning up toxically polluted sites.[1]

In the long term, however, the current trend towards developing smart and autonomous machines with the capacity to be trained and make decisions independently not only holds many economic advantages, but also raises a variety of concerns regarding their direct and indirect effects on society as a whole.[2] It is the thesis of this article that AI is creating a tangled web of legal issues that legal systems the world over will have to deal with and resolve. The purpose of this article is not to offer comprehensive solutions, but to raise awareness among legal scholars and practitioners of the most pressing legal challenges presented by the increased application of AI. In particular, there are the challenges to ensure privacy, the dissemination of factually accurate information, non-discrimination, due process, transparency and accountability in decision-making processes.[3]

This article starts with a description of key terms; it then examines some of these challenges highlighted above, specifically in the areas of privacy, the spread of disinformation, state surveillance, and bias and algorithmic transparency. It concludes with an analysis of the challenges that the legal system faces in regulating the burgeoning field of AI.

Description of Key Terms

One of the main issues that must be faced when discussing the legal underpinnings of technological innovations arises from the vocabulary used by those developing and marketing these tools. Information technology (IT) professionals, like lawyers, have developed a 'somewhat dense and opaque lexicon' that the uninitiated find too complex to master.[4] It is important, therefore, to offer a general outline of the principal terms

1 European Parliament Committee on Legal Affairs, *Report with Recommendations to the Commission on Civil Law Rules on Robotics* (PE582.443v03-00) (2017) 3.

2 ibid 4.

3 ibid 5.

4 Iria Giuffrida, Fredric Lederer and Nicolas Vermerys, 'A Legal Perspective on the Trials and Tribulations of AI: How Artificial Intelligence, the Internet of Things, Smart Contracts, and Other Technologies Will Affect the Law' (2018) 68 Case Western Reserve LR 751.

appearing in the media, and that will undoubtedly make their appearance in the boardroom and the courtroom sooner rather than later.

Although AI is talked about in the media almost every day, there is still no generally accepted definition of the term. The term 'artificial intelligence' may have been coined by John McCarthy and others in a paper first published in 1955. The authors explained that:

> An attempt will be made to find how to make machines use language, from abstractions and concept, solve kinds of problems now reserved for humans, and improve themselves … For the present purpose the artificial intelligence problem is taken to be that of making a machine behave in ways that would be called intelligent if a human were so behaving.[5]

Therefore, in the sense intended in this article, AI refers to a computer's ability to imitate human intelligent behaviour, especially human cognitive functions, such as the ability to reason, discover meaning, generalise and learn from past experience.[6] AI is generally thought to refer to:

> [M]achines that respond to stimulation consistent with traditional responses from humans, given the human capacity for contemplation, judgment and intention.[7]

Accordingly, they operate in an intentional, intelligent and adaptive manner.[8]

---

5 John McCarthy, Marvin L Minsky, Nathaniel Rochester, and Claude E Shannon, 'A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence' (2006) 27(4) AI Magazine 12. In short, AI is a device or a system which is able to perform tasks that usually require human intelligence. It has the ability to mimic 'cognitive' functions that humans associate with other human minds, such as learning and problem-solving. See David R Richie II and Jarina D Duffy 'Artificial Intelligence in the Legal Field', *Association of Corporate Counsel Greater Philadelphia In-House Counsel Conference* (25 April 2018) 1.

6 Alan Turing defined artificial intelligence as the 'science and engineering of making intelligent machines, especially intelligent computer programs': Alan M Turing, 'Mind' (1950) 59(236) Computing Machinery and Intelligence, 433.

7 According to researchers Shubendu and Vijay, these software systems 'make decisions which normally require [a] human level of expertise' and help some people anticipate problems or deal with issues as they come up—as quoted in Darrell M West and John R Allen 'How Artificial Intelligence is Transforming the World' (2018) *Brookings Report*, 24 April <https://www.brookings.edu/research/how-artificial-intelligence-is-transforming-the-world/> accessed 14 February 2019.

8 West and Allen (n 7) 74.

Max Tegmark,[9] a cosmologist at the Massachusetts Institute of Technology and co-founder of the Future of Life Institute, described the domain of AI in the following way:

> Artificial Intelligence today is properly known as *narrow AI (or weak AI)*, in that it is designed to perform a narrow task (eg, only facial recognition or only internet searches or only driving a car). However, the long-term goal of many researchers is to create *general AI (AGI or strong AI)*. While narrow AI may outperform humans at whatever the specific task is, like playing chess or solving an equation, AGI would outperform humans at *nearly every cognitive task*.

While AI has many attributes useful for its varied applications, at present two are most important in the legal domain.

First, 'machine learning' (ML), the leading innovative force in AI, has proven enormously efficient, performing in mere minutes, tasks that would otherwise take a team of lawyers tens of hours.[10] ML refers to the capability of AI systems to teach themselves and learn from experience. This means, in essence, that AI can do much more than blindly adhere to what it has initially been programmed to do; it can learn from experience and data to improve its capabilities constantly.[11]

Initially, ML frameworks were used to understand vast amounts of data—so-called 'big data'—that collectively are almost unimaginably vast in the human context, and far beyond what highly skilled, experienced workers could reasonably construct on their own.[12] However, modern ML has evolved beyond understanding the types of

---

9 Max Tegmark, 'Benefits and Risks of Artificial Intelligence' (2016) Future of Life <http://www.futureoflife.org/background/benefits-risks-of-artificial-intelligence/> accessed 2 March 2019 (emphasis original).

10 See, generally, Kathryn D Betts and Kyle R Jaep, 'The Dawn of Fully Automated Contract Drafting: Machine Learning Breathes New Life into a Decades-old Promise' (2017) 15 Duke L & Tech Rev 216.

11 Gary E Merchant, 'Artificial Intelligence and the Future of Legal Practice' (2017) 14(1) Scitech Lawyer 21. In more technical terms, ML enables computers to learn to optimise certain tasks without the benefit of explicit rules-based programming. Therefore, ML can be described as the ability of a computer to modify its programming to account for new data and modify its operations accordingly. Put differently, ML is the practice of using algorithms to parse data, learn from them, and then determine or predict something. Therefore, rather than hand-coding software routines with a specific set of instructions (algorithms) to accomplish a particular task, the machine is 'trained' using large amounts of data and algorithms that give it the ability to learn how to perform the task: Anonymous, Artificial Intelligence Primer (*Victorian All-Party Parliamentary Group on Artificial Intelligence*, February 2018) 2 <https://www.parliament.vic.gov.au/images/stories/AI-Primer_Feb2018.pdf> accessed 14 March 2019.

12 Benjamin Alarie, Anthony Niblett and Albert H Yoon, 'How Artificial Intelligence Will Affect the Practice of Law' (2018) 68(1) University of Toronto LJ 116.

information that can be categorised in big data to provide insights about how that information could be relevant to a particular set of facts.[13]

ML should be understood as a spectrum that ranges from relatively simple algorithms to complex self-teaching systems that could eventually mirror the human brain in their complexity, if not their structure. Such self-teaching systems are termed 'deep learning'.[14] Deep learning relies on what are referred to as 'neural networks', an interconnected group of nodes designed to mimic the activity of neurons in the human brain in order to recognise complex patterns in data sets.[15]

Secondly, 'natural-language processing' (NLP) is the capability of algorithms and software to interpret, understand and generate spoken and written human language[16] and then to apply and integrate that understanding in order to perform human-like analysis.[17] Search engines, speech-to-speech translation and artificially intelligent assistants, such as iPhone's *Siri,* are built with NLP technology for the user's benefit.[18]

The Threat to Privacy

Privacy is a fundamental right that is essential to human dignity. The right to privacy also reinforces other rights, such as the rights to freedom of expression and association.[19] AI systems are often trained through their being given access to and being able to analyse big data sets. However, the collection of data impedes the right to privacy in

---

13 Mark Burdon, 'Interview with Mark Burdon: Artificial Intelligence and the Law' (*Justice and The Law Society, The University of Queensland*, 8 March 2018) <http://www.jatl.org/blog/2018/3/8/interview-with-mark-burdon-artificial-intelligence-and-the-law> accessed 28 February 2019.

14 'Deep learning' is 'a sub-field of ML, where models inspired by how our brain works are expressed mathematically, and the parameters defining the mathematical models, which can be in the order of few thousands to 100+ million, are learned automatically from the data': Jonathan Sanito, Sayan Pathak, and Roland Fernandez, 'Deep Learning Explained' (2018) edX 4 March <https://www.edx.org/course/deep-learning-explained-microsoft-dat236x-1> accessed 14 February 2019. See also Giuffrida (n 4) 751. While an ML model needs to be told how it should make accurate predictions (by feeding it more data), a deep-learning model is able to learn through its own computing 'brain'. It is similar to the way in which a human being would perceive something, think about it and then draw a conclusion. To achieve this, deep learning uses a layered structure of algorithms referred to as an 'artificial neural network', the design of which is inspired by the biological neural network of the human brain: Anonymous, Artificial Intelligence Primer (*Victorian All-Party Parliamentary Group on Artificial Intelligence* February 2018) 2 <https://www.parliament.vic.gov.au/images/stories/AI-Primer_Feb2018.pdf> accessed 14 March 2019.

15 Richie and Duffy (n 5) 1; Giuffrida and others (n 4) 755.

16 Richie and Duffy (n 5) 1.

17 Merchant (n 11) 21.

18 Alarie and others (n 12) 116.

19 See, for example, s 14 of the Constitution of the Republic of South Africa, 1996; Article 12 of the Universal Declaration of Human Rights; and Article 17 of the International Covenant on Civil and Political Rights.

that the analysis of data by AI systems might reveal private information about individuals.[20] But data collection has become ubiquitous.

For example, many people are already willing to wear or carry devices that provide great detail about their circumstances to databases.[21] Our cellphones are capable of providing real-time spatial location data[22] and of retaining a secret record of every location that we visit.[23] Similarly, Fitbit will soon add glucose monitoring to its products, which currently track, among other things, steps, sleeplessness, heart rate and distance.[24] In late 2018, Apple rolled out a feature on the Apple Watch that enables it to take the wearer's electrocardiogram (ECG) through specially designed sensors.[25] And we have already embraced highly contextualised and automated directives in the travel context: we eagerly (and sometimes blindly) accept directions from Google Maps.[26] The capability of machines to invade human privacy will only increase.[27]

The major issue here is that the more convenient an agent is, the more it needs to know about a person. This creates a trade-off: more help requires more intrusion. The record to date is that convenience overwhelms privacy. This will probably continue[28]—privacy and independence will increasingly be sacrificed to convenience.[29]

Researchers are also working on a variety of technologies aimed at what can loosely be referred to as 'mind reading'.[30] For example, based on measurements of brain activity,

---

20 Lindsey Andersen and others, 'Human Rights In the Age of Artificial Intelligence' (2018) Access Now November <https://www.accessnow.org/cms/assets/uploads/2018/11/AI-and-Human-Rights.pdf> accessed 2 March 2020.

21 Brian Sheppard, 'Warming Up to Inscrutability: How Technology Could Change Our Concept of Law' (2018) 68 University of Toronto LJ 41.

22 Yu-Che Chen and Michael J Ahn, *Routledge Handbook of Information Technology in Government* (Routledge 2017) 109.

23 Charles Arthur, 'iPhone Keeps Record of Everywhere You Go' *The Guardian*, 20 April 2011 <https://www.theguardian.com/technology/2011/apr/20/iphone-tracking-prompts-privacy-fears> accessed 4 June 2019.

24 Greg von Portz and Satish Misra, 'Medtronic & Fitbit Partner to Connect Activity Data with Continuous Glucose Monitoring' (*iMedical Apps*, 24 January 2017) <https://www.imedicalapps.com/2017/01/medtronic-fitbit-partner-connect-activity-data-continuous-glucose-monitoring/> accessed 12 February 2019.

25 Lauren Goode, 'A Guide to Using Apple Watch's Heart Rate Features, Including ECG' (*Wired*, 6 December 2018) <https://www.wired.com/story/how-to-take-an-ecg-reading-on-apple-watch/> accessed 4 June 2019.

26 Sheppard (n 21) 41.

27 Anthony J Casey and Anthony Niblett, 'Self-driving Laws' (2016) 66 University of Toronto LJ 438.

28 Michale M Roberts, as quoted in Janna Anderson, Lee Raine and Alex Luchsinger, 'Artificial Intelligence and Future of Humans' (*Pew Research Center Internet and Technology*, 10 December 2018) <https://www.pewinternet.org/2018/12/10/artificial-intelligence-and-the-future-of-humans/> accessed 12 April 2019.

29 Kostas Alexandridis, as quoted in Anderson and others (n 28).

30 Adam J Kolber, 'Will There Be a Neurolaw Revolution?' (2014) 89 Indiana LJ 835.

researchers can make remarkably accurate predictions about what images are shown to subjects in a brain scanner, be it a still image[31] or even, to a more limited extent, a video.[32] One study demonstrated that subjects under functional magnetic-resonance imaging (fMRI) can be taught to spell words mentally in a manner that can be decoded in real time by the experimenters.[33] In 2012, Jack Gallant, Professor of Psychology at the University of California at Berkeley, predicted that '[w]ithin a few years, we will be able to determine someone's natural language thoughts using fMRI-based technology.'[34]

These new brain-imaging techniques point to a future in which our thoughts will not be as private as they are now.[35] People could be scanned for one purpose—for example, to see how advertising campaigns affect their brains—while they inadvertently generate information that bears on their racial biases, sexual orientation or other sexual preferences.

The Spread of Disinformation

AI can also facilitate the creation of so-called 'deep fakes', which are AI-enhanced photorealistic pictures and videos.[36] These AI-enhanced deep fakes leverage on

> machine-learning algorithms to insert faces and voices into video and audio recordings of actual people that enables the creation of realistic impersonations out of digital whole cloth. [37]

---

31 See, for example, Kendrick N Kay, Thomas Naselaris, Ryan J Prenger and Jack L Gallant, 'Identifying Natural Images from Human Brain Activity' (2008) 452 Nature 352.

32 See, for example, Shinji Nishimoto, An T Vu, Thomas Naselaris, Yuval Benjamini, Bin Yu and Jack L Gallant, 'Reconstructing Visual Experiences from Brain Activity Evoked by Natural Movies' (2011) 21 Current Biology 1641.

33 Bettina Sorger, Joel Reithler, Brigitte Dahmen and Rainer Goebe, 'A Real-Time fMRI-based Spelling Device Immediately Enabling Robust Motor-independent Communication' (2012) 22 Current Biology 1333.

34 As quoted in Kolber (n 30) 835.

35 ibid 836.

36 The first use of deep-fake technology was to paste people's faces onto target videos, often in order to create non-consensual pornography. James Vincent, 'ThisPersonDoesNotExist.com Uses AI to Generate Endless Fake Faces' (*The Verge*, 15 February 2019) <https://www.theverge.com/tldr/2019/2/15/18226005/ai-generated-fake-people-portraits-thispersondoesnotexist-stylegan> accessed 4 June 2019.

37 Bobby Chesney and Danielle Citron, 'Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security' (forthcoming 2019) 107 California LR <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3213954> accessed 6 June 2019.

'Deep fakes' are becoming the exemplification of the power of AI to generate misinformation and fake news.[38] The current imperfect deep-fake technology makes funny videos that caricature celebrities and political figures. But as technology improves—as it does every single day—the human eye will soon find it impossible to distinguish between real and fake. The concern is that this technology might be used to create fake photos, videos and news stories for malicious ends—sowing chaos, instigating conflict and furthering a 'crisis of truth'.[39]

In February 2019, the creators of a revolutionary AI system that can write news stories and works of fiction—nicknamed 'deep fakes for text'—took the unusual step of not releasing their research publicly, for fear of potential misuse. *OpenAI*, a non-profit research company backed by, among others, Elon Musk, stated that its new AI model, called GPT2, is so good and the risk of malicious use so high, that it is deviating from its normal practice of releasing the full research to the public in order to allow itself more time to discuss the ramifications of the technological breakthrough.[40]

It functions like this: GPT2 is fed text—anything from a few words to a whole page—and is then asked to write the next few sentences based on its predictions of what should come next. GPT2 is capable of writing plausible passages that match what it is given in both style and subject. For example, when fed the opening line of George Orwell's *Nineteen Eighty-Four* —'It was a bright cold day in April, and the clocks were striking thirteen'—the system recognised the vaguely futuristic tone and the novelistic style, and continued with:

> I was in my car on my way to a new job in Seattle. I put the gas in, put the key in, and then I let it run. I just imagined what the day would be like. A hundred years from now. In 2045, I was a teacher in some school in a poor part of rural China. I started with Chinese history and history of science.[41]

Having previously created an AI that could generate realistic-looking facial images, the scientists at *DataGrid*, a startup company based at the Kyoto University in Japan, have

---

38 Researchers have developed tools that lets one perform face swaps in real time; Adobe is creating a 'Photoshop for audio' that lets users edit dialogue as easily as a photo; and a Canadian start-up company Lyrebird, offers a service that lets users fake someone else's voice with just a few minutes of audio. James Vincent, 'Watch Jordan Peele Use AI to Make Barack Obama Deliver a PSA About Fake News' (*The Verge*, 17 April 2018) <https://www.theverge.com/tldr/2018/4/17/17247334/ai-fake-news-video-barack-obama-jordan-peele-buzzfeed> accessed 4 June 2019.

39 Andersen and others (n 28).

40 Alex Hern, 'New AI Fake Text Generator May Be Too Dangerous to Release, Say Creators' (*The Guardian*, 14 February 2019) <https://www.theguardian.com/technology/2019/feb/14/elon-musk-backed-ai-writes-convincing-news-fiction> accessed 4 June 2019.

41 ibid.

now developed an AI system that is the first ever to fabricate images of full human beings, including their clothes, hairstyles and even the very way in which they pose.[42]

The ability to manipulate and generate realistic imagery at scale is going to have a substantial effect on the way in which modern societies think about evidence and trust. Such software could also be extremely useful for creating political propaganda and influencing election campaigns, as evidenced by the data-mining and psychological influencing during the 2016 US presidential election and the Brexit referendum in the United Kingdom.[43] Technologist Aviv Ovadya summed up the fears created by this technology:

> What happens when anyone can make it appear as if anything has happened, regardless of whether or not it did?[44]

This technology implies, of course, that anyone with a vendetta can create a 'deep fake' that depicts someone doing something unsavoury or illegal. The courtroom is not immune to misleading evidence; this fake evidence will inevitably leak into the courtroom, and it could dupe factfinders into believing that an innocent person committed a crime.

Bobby Chesney and Danielle Citron predict a development stemming from deep-fake evidence: 'immutable life logs' as an alibi service.[45] Because deep-fake technology will be able to portray people saying and doing things that they actually never said or did, alibis will become essential for digitally ensnared accused to prove their innocence in the courtroom. Hence, deep fakes will create a heightened demand for proof of where a person was and what they were doing at all times. Therefore, companies—and perhaps even the government—will enlist the life-logging business, through which wearable technology (such as an Apple Watch, for example) could track its user around the clock.[46] Life-logging's potent solution to deep-fake evidence, however, might very well destroy privacy.

---

42 Ian Randall, 'First "Deepfake" AI That Can Replicate People Moving Creates Footage of Crowds of Imaginary Humans That Are Indistinguishable from the Real Thing' (*Mail Online*, 7 May 2019) <https://www.dailymail.co.uk/sciencetech/article-7001293/Deepfake-AI-replicate-bodies-motion-creates-footage-crowds-imaginary-people.html> accessed 4 June 2019.

43 Vincent (n 36). See also, generally, Christopher Wylie, *Mindf*ck: Cambridge Analytica and the Plot to Break America* (Random House 2019).

44 Vincent (n 36).

45 Chesney and Citron (n 37).

46 Daniel Rankin, 'How Artificial Intelligence Could Change the Law in Three Major Ways' (2018) October The Journal of Law and Technology at Texas <http://jolttx.com/2018/10/14/how-artificial-intelligence-could-change-the-law-in-three-major-ways/> accessed 7 March 2019.

The AI Surveillance State

One possible outcome of these new technological developments under increased state regulation is what the French social theorist, Gilles Deleuze, called a 'society of control', that is, a world in which human actions are increasingly managed and monitored by machines.[47]

In the United States, both federal and state governments have outsourced many regulatory and legal decisions to computation. Tax returns are too voluminous for Internal Revenue Service personnel to examine manually; 'audit flags' are programmed to determine which returns should receive greater scrutiny or be rejected outright. Homeland Security officials are using big data and algorithms to determine which travellers pose a security risk and who can pass without any scrutiny to their flights. So-called 'predictive policing' deploys law-enforcement resources before crimes are committed. And once perpetrators are convicted, 'evidence-based sentencing' may quantify punishment by using data and algorithms to adjust the length of prison sentences based on myriad factors.[48]

Facial Recognition and Behaviour Prediction

Privacy proponents will recoil upon learning that AI is also increasing the effectiveness of state surveillance techniques.[49] Before AI, cameras were useful only to the extent that someone either observed a live feed or reviewed recorded footage. That time has passed. With the assistance of AI, cameras can now navigate three dimensions and make sense of what they 'see'—all without any human intervention or assistance. Moreover, facial-recognition cameras are beginning to operate beyond ordinary human capability: they can identify millions of faces and predict human behaviour.[50]

Facial-recognition technology is nothing new. We see it, for example, on the iPhone X with its face-scanning technology.[51] But, thus far, China is the world leader in using facial-recognition technology as a surveillance tool. The advent of China's social credit system (SCS) is a sign of what is likely to come: our rights and affordances as individuals will be determined by the SCS. This is the Orwellian nightmare realised.[52]

---

47 Lawrence B Solum, 'Artificial Meaning' (2014) 89 Washington University LR 69.

48 See Frank Pasquale and Glynn Cashwell, 'Four Futures of Legal Automation' (2015) 63 UCLA LR 37 and the sources cited there.

49 Rankin (n 46).

50 ibid.

51 Conceptually, the way in which it works is simple: the camera looks at a face, extracts distinguishing facial features (such as the size and width of the nose, for example) and then compares those features against a database of pictures (sometimes taken from driver's licence photos): Rankin (n 46).

52 Simon Biggs, as quoted in Andersen and others (n 20).

Under its *Sharp Eyes* programme,[53] China's goal is to recognise all Chinese citizens within seconds of their faces appearing on a camera.[54] To this end, China has scattered cameras across the country.[55] Its comprehensive face database has already led to one Chinese city capturing 375 suspects and 39 fugitives since its inception.[56] By the end of 2020, China aims to have a file on every Chinese citizen that includes all the data collected on their behaviour.[57] China has become the world's leading AI-powered surveillance state.[58]

The Chinese government has policies in place to monitor individuals and punish bad behaviour. A citizen's social ranking in the government's eyes might be lowered if they evade taxes, swindle other people or create fake advertisements. The SCS is also supposed to help prevent annoying behaviour on public transportation, such as one well-publicised case in which a passenger who took up another person's reserved seat refused to get up.

In May 2018, the government of China introduced a travel ban on people with poor 'social credit'. According to a report from China's National Public Credit Information Centre, during the last week of February 2019 people have been blocked 17,5 million times from purchasing airplane tickets, and 5,5 million times from buying high-speed train tickets. These people had become 'discredited' for unspecified 'behavioural crimes'.

---

53 New technologies make it possible to match images and voices with other types of information, and to use AI on these combined data sets to improve law enforcement and national security. Through its *Sharp Eyes* program, Chinese law enforcement is matching video images, social media activity, online purchases, travel records and personal identity into a 'police cloud'. This integrated database enables authorities to keep track of criminals, potential law-breakers and terrorists. See Biggs, as quoted in Andersen (n 20).

54 See, generally, Simon Denyer, 'In China, Facial Recognition is Sharp End of a Drive for Total Surveillance' (*Washington Post,* 7 January 2018) <https://www.washingtonpost.com/news/world/wp/2018/01/07/feature/in-china-facial-recognition-is-sharp-end-of-a-drive-for-total-surveillance/?utm_term=.0b9cac9ecab4> accessed 7 March 2019.

55 'There are more than 360 000 surveillance cameras installed in Linyi City, out of 2 930 000 surveillance cameras in all Shandong province.' Oiwan Lam, 'With "Sharp Eyes", Smart Phones and TV Sets Are Watching Chinese Citizens' (*ADVOX,* 3 April 2018) <https://advx.globalvoices.org/2018/04/03/with-sharp-eyes-smart-phones-and-tv-sets-are-watching-chinses-citizens/> accessed 7 March 2019.

56 Rosalie Chan, 'One Chinese City is Using Facial-recognition That Can Help Police Detect and Arrest Criminals in as Little as 2 Minutes' (*Business Insider,* 19 March 2018) <https://www.businessinsider.com/china-guiyang-using-facial-recognition-to-arrest-criminals-2018-3> accessed 7 March 2019.

57 Shannon Liao, 'China Banned Millions of People with Poor Social Credit from Transportation in 2018' (*The Verge,* 1 March 2019) <https://www.theverge.com/2019/3/1/18246297/china-transportation-people-banned-poor-social-credit-planes-trains-2018> accessed 4 June 2019.

58 West and Allen (n 7).

In the world of technology, facial recognition has become a known commodity. 'Behaviour prediction', on the other hand, is the latest trend.[59] In addition to recognising who you are, AI-augmented cameras will be 'intelligent' enough to predict your behaviour. This technology already exists and it is improving by the day.[60] One company claims that it has created a machine that can predict an individual's sexual orientation. The machine has already proven its ability to determine sexual orientation by using algorithms based on facial features and expressions, to an accuracy level of ninety-one per cent.[61]

AI Harnessed to Predict and Apprehend Criminals

Another company, *Faception*, in Tel Aviv created a program that purports to determine whether someone is a criminal—only by looking at a face. The camera does not simply run the photo of a person against a criminal database: based on the premise that facial features reveal personality traits (called 'physiognomy'), the program reads a face and assigns the probability of criminal intent. In one demonstration, the program achieved ninety per cent accuracy.[62]

New AI software is being used in Japan to monitor the body language of shoppers for signs that they are planning to steal. This software, developed by Japanese company *Vaak*, differs from similar products that match faces to criminal records. Instead, *VaakEye* uses algorithms to analyse footage from security cameras to spot fidgeting, restlessness and other body-language cues that could be suspicious, and then alerts shop employees about potential thieves via an app.[63]

Using AI to apprehend thieves raises ethical questions that have led the human rights NGO, *Liberty*, to advocate banning facial-recognition technology in the United

---

59 Rankin (n 46).

60 Rankin (n 46).

61 Sam Levin, 'New AI Can Guess Whether You're Gay or Straight from a Photograph' (*The Guardian*, 7 September 2017) <https://www.theguardian.com/technology/2017/sep/07/new-artificial-intelligence-can-tell-whether-youre-gay-or-straight-from-a-photograph> accessed 7 March 2019.

62 Gus Lubin, 'Facial-profiling Could Be Dangerously Inaccurate and Biased, Experts Warn' (*Business Insider*, 12 October 2016) <https://www.businessinsider.com/does-faception-work-2016-10> accessed 7 March 2019.

63 The company fed the algorithm 100 000 hours of surveillance data to train it to monitor everything from the facial expressions of shoppers to their movement and clothing. *VaakEye* was launched in 50 shops in Japan during March 2019, and the company plans to expand to 100 000 shops in Japan within three years. Proponents of systems such as this claim that they could help reduce global retail costs from shoplifting, which reached $USD34 billion in 2017. Nell Lewis, 'Should AI Be Used to Catch Shoplifters?' (*CNN*, 18 April 2019) <https://edition.cnn.com/2019/04/18/business/ai-vaak-shoplifting/index.html> accessed 29 April 2019; Christopher Carbone, 'Creepy AI Reportedly Spot Shoplifters Before They Steal' (*Fox News*, 18 April 2019) <https://foxnews.com/tech/creepy-ai-will-reportedly-spot-shoplifters-before-they-steal> accessed 29 April 2019.

Kingdom. The NGO is particularly concerned that a retail environment—a private sphere—is starting to perform something akin to a police function.[64]

To exacerbate these concerns, there is also the potential of AI being used to fuel discrimination. A 2018 study by researchers from the Massachusetts Institute of Technology and Stanford University found that various commercial facial analysis programs demonstrate skin-type and gender biases, depending on the types of data that is used.[65] Amazon has also recently incurred the ire of legislators and privacy advocates over bias in its AI-based facial recognition systems.[66] Technologies that rely on algorithms—particularly with regard to human behaviour—have the potential to engage in discrimination. After all, it is human beings who have to train the algorithms about what or whom to treat suspiciously.

One way in which the police arrest suspects is through arrest warrants, which, in most common-law jurisdictions at least, is based on a 'reasonable grounds' standard. If an AI-equipped camera identifies someone as a likely criminal, will that be enough to meet the reasonable grounds standard? If so—and assuming the technology assigns a percentage of criminality to an individual—how much will satisfy reasonable grounds: ninety per cent, seventy per cent or fifty per cent? This, of course, also raises the legal–ethical question of whether it is even legal or ethical to arrest a person *before* they commit a crime.[67]

AI Technology in the Courtroom?

What about the role of this technology as evidence in the courtroom? Would it be too prejudicial to show the fact-finder that AI software determined that an accused is a criminal? What if, instead, prosecutors used the technology during trial to buttress their arguments? In closing address, for example, the prosecutor might argue: 'Based on all the eye-witness testimony, along with the determination that the accused, considering his facial features, has an 80% likelihood of having committed the crime charged, you should find the accused guilty.'

These types of arguments could be commonplace in the future, yet there currently is no regulatory framework in place to regulate these technologies in the circumstances discussed above. Clarity is needed from lawmakers and regulators regarding who will ultimately decide the circumstances in which the use of this technology will be appropriate or desirable as a matter of public policy.[68]

64 Lewis (n 63).
65 ibid; Carbone (n 63).
66 Carbone (n 63).
67 Rankin (n 46).
68 Lewis (n 63).

Bias and Algorithmic Transparency

Developments in technology raise important policy and ethical issues.[69] For example, how should we promote data access? How do we guard against biased or unfair data that are used in algorithms? What types of ethical principle are introduced through software programming, and how transparent should designers be about their choices?

It must be remembered that technology is not necessarily neutral and objective. A software design may expressly, through its programming, reflect a preference for certain values over others. eBay's online dispute-resolution mechanism offers an example: eBay has been accused of favouring buyers over sellers through its explicit adoption of a 'buyer-is-always-right' policy.[70]

AI systems can also be inadvertently programmed to have bias because of the biases of the programmers or, in the case of ML algorithms, actually learn to be biased based on the data set from which AI is learning.[71]

Algorithms—the set of instructions according to which computers carry out tasks—have become an integral part of everyday life and they have immersed themselves in the law.[72] In the United States, for instance judges in certain states use algorithms as part of the sentencing process to assess recidivism risk. Many law-enforcement agencies use algorithms to predict when and where crimes are likely to occur (so-called 'predictive policing').[73]

Most algorithms are created with good intentions, but questions have started surfacing over algorithmic bias on employment search websites, in credit reporting bureaus, on social media websites and even the in criminal justice system, where sentencing and parole decisions seem to be biased against African-Americans.[74] These issues are likely to become exacerbated as ML and predictive analytics become more sophisticated, particularly because with deep learning (which learns autonomously) algorithms can

69 West and Allen (n 7).

70 Justice MJ Beazley, 'Law in the Age of Algorithm' (27 September 2017) State of the Profession Address, New South Wales Young Lawyers, Sydney 9.

71 Erwin Loh, 'Medicine and the Rise of the Robots: A Qualitative Review of Recent Advances of Artificial Intelligence in Health' (2018) 2 BMJ Leader 61.

72 Luis Millán, 'Artificial Intelligence' (*Canadian Lawyer Magazine,* 3 April 2017) <https://www.canadianlawyermag.com/article/artificial-intelligence-3585> accessed 6 March 2019.

73 See, for example, Rashida Richardson, Jason Schultz and Kate Crawford, 'Dirty Data, Bad Predictions: How Civil Rights Violations Impact Police Data, Predictive Policing Systems and Justice' (2019) 94 New York University Law Review 192–233.

74 Millán (n 72).

quickly reach a point where human beings can often no longer explain or understand them. Nicolas Vermeys, of the *Cyberjustice Laboratory* in Montreal, stated that:[75]

> We have no idea how [algorithms] arrive at their decisions and, therefore, cannot evaluate whether the decisions have value or not … There is a risk to relying completely on machines without necessarily understanding its reasoning.

No human being is completely objective,[76] and so it is with algorithms, which, after all, have been programmed by human programmers. Programmers operate on certain premises and assumptions, which are not tested by anyone else, and this leads to results based on those premises and assumptions, which, in turn, give rise to bias.[77]

Moreover, it is very difficult to challenge a computer's decisions, because whoever owns the algorithms owns the trade secrets associated with them, and is neither going to reveal the source code nor likely be willing to even discuss the secret source and how it makes the algorithm functions.[78] What justifies the algorithm from an economic viability perspective is its success or perceived success, which is an entirely different question of whether or not it operates in biased ways.[79]

Racial issues also come up in facial recognition software. Most of these systems operate by comparing a person's face to a range of faces in a database. As pointed out by Joy Buolamwini, a researcher at the MIT Media Lab: 'If your facial recognition data contains mostly Caucasian faces, that is what your program will learn to recognize.'[80]

Unless the databases have access to diverse data, these programs perform poorly when attempting to identify African-American or Asian-American features. Many historical data sets reflect traditional values, which may or may not represent the desired preferences in a current system. As Buolamwini notes, such an approach risks repeating inequities of the past:[81]

> The rise of automation and the increased reliance on algorithms for high-stakes decisions—such as whether someone can get insurance or not, your likelihood to default on a loan or somebody's risk of recidivism—means that this is something that needs to be resolved. Even admission decisions are increasingly being automated—what school

---

75 As quoted in ibid.

76 See, generally, Willem H Gravett, 'The Myth of Rationality: Cognitive Biases and Heuristics in Judicial Decision-Making' (2017) 134 South African Law Journal 53–79.

77 Millán (n 72).

78 ibid.

79 ibid.

80 Editorial, 'Joy Buolamwini' (*Bloomberg Businessweek*, 3 July 2017) 80.

81 ibid.

> our children go to and what opportunity they have. We don't have to bring the structural inequalities of the past into the future we create.

As algorithms have become an established part of high-stakes projects, concerns have arisen that they are not adequately transparent to allow for accountability, especially if they are used as the basis for harmful or coercive decisions.[82] In 2012, the principal researcher at Microsoft Research New England, Tarleton Gillespie, stated: 'There may be something in the end impenetrable about algorithms.' Others are not quite as fatalistic, but there is growing consensus among computer scientists that it would take aggressive research to cut through algorithmic opacity, particularly in ML, where opacity is at its densest.[83]

One of the major problems is that classic values of administrative procedure, such as due process, are not easily coded into software language. In the United States, many automated implementations of social-welfare programmes, ranging from state emergency assistance to Affordable Care Act (Obamacare) exchanges, have resulted in erroneous denials of benefits, lengthy delays and troubling outcomes.[84] Financial engineers may quantify risks in ever more precise ways for compliance purposes, but through ML their models have also led to financial instability and even financial crisis. As the recession of 2008 has shown, even when structured securities, parsed by proprietary software, proved good for the investment banks' bottom lines, they did not contribute to overall economic productivity—in fact, quite the opposite.

The most fully automated part of the financial sector—high-frequency trading—has generated considerable controversy.[85] Consider, for instance, the flash crash of 6 May 2010, when the stock market lost hundreds of points and close to USD 1 trillion in market value in a matter of minutes. Traders had programmed split-second algorithmic strategies to gain a competitive edge, but soon found themselves in the position of sorcerer's apprentice, unable to control the technology they had developed. Although prices returned to normal later that same day, there is no guarantee that in future the markets would be so lucky.[86]

Depending on how AI systems are set up, they can help people to discriminate against individuals they do not like or help screen or build lists of individuals based on unfair criteria. The types of consideration that go into programming decisions matter a lot in terms of how the systems operate and how they affect customers.

---

82 Sheppard (n 21) 47.

83 ibid.

84 Pasquale and Cashwell (n 48) 38.

85 ibid.

86 Note also the disastrous USD 440 million loss of Knight Capital in August 2012 that was traced to IT and software issues at the firm that took nearly an hour to fix: ibid 39.

For these reasons, the European Union (EU) has implemented the General Data Protection Regulation (GDPR) in May 2019. The rules specify that people have 'the right to … contest "legal or similarly significant" decisions made by algorithms and appeal for human intervention' in the form of an explanation of how the algorithm generated a particular outcome. Each guideline is designed to ensure the protection of personal data and provide individuals with information on the way the 'black box' operates.[87]

ML is the ability of a computer to modify its programming to account for new data and modify its operations accordingly. It uses computers to run predictive models that learn from existing data to forecast future behaviours, outcomes and trends.[88] ML, therefore, is dependent on data. The more data it can access, the better it can learn. However, the quality of the data, the way the data are input into the system and how the system is 'trained' to analyse the data can all have dire effects on the validity, accuracy and usefulness of the information generated by the algorithm.

In short, not only can an otherwise perfect algorithm fail to accomplish its set goals, but it may also prove affirmatively harmful.[89] For example, the algorithm employed by Google to answer user questions erroneously declared that Barack Obama, a Christian, was a Muslim.[90] The algorithm simply did what it was 'trained' to do—it gathered information from the internet, 'feeding' on websites that propagated false information. Its data pool was polluted, and the algorithm could not discern between 'good' and 'bad' data. This was also brought to light, for example, by the Microsoft chatbot, 'Tay', which learned to interact with human beings via Twitter.[91] Within twenty-four hours, the chatbot became racist, because internet trolls had bombarded it with mostly offensive and erroneous data in the form of inflammatory tweets, from which the chatbot had 'learned'.[92]

---

87 Cliff Kuang, 'Can AI Be Taught to Explain Itself' (*New York Times Magazine*, 21 November 2017) <https://www.nytimes.com/2017/11/21/magazine/can-ai-be-taught-to-explain-itself.html> accessed 5 June 2019.

88 Giuffrida and others (n 4) 753.

89 ibid 754.

90 Jack Nicas, 'Google Has Picked an Answer for You – Too Bad It's Often Wrong' (*Wall Street Journal*, 16 November 2017) <https://www.wsj.com/articles/googles-featured-answers-aim-to-distill-thruthbut-often-get-it-wrong-1510847867> accessed 17 March 2019.

91 Tay was able to perform a number of tasks, such as telling jokes to users and commenting on pictures that users sent it. Nisith Desai Associates (*The Future Is Here: Artificial Intelligence and Robotics*, May 2018) 12 <http://www.nishithdesai.com/fileadmin/user_upload/pdfs/Research_Papers/Artificial_Intelligence_and_Robotics.pdf> accessed 4 June 2019.

92 Daniel Victor, 'Microsoft Created a Twitter Bot to Learn from Users. It Quickly Became a Racist Jerk' (*New York Times*, 24 March 2016) <https://www.nytimes.com/2016/03/25/technology/microsoft-

Even if the data were accurate, the person 'training' the AI could infuse their own biases into the system. This may have been a factor in the crime-predicting software that has led to the arrest of an unjustifiably high number of African-Americans and other minorities in the United States,[93] as well as sentencing tools that predict higher rates of recidivism for these same individuals.[94]

Accordingly, the effective accuracy of an algorithm is dependent on both the programming and the data. This dictates a further, legally troubling conclusion. If there are doubts about the results of an algorithm, one can at least theoretically inspect and analyse the programming that constitutes the algorithm. However, given the sheer volume of data available on the internet, it may be impossible to adequately determine and inspect the data used by the algorithm.[95]

Consider, for example, that a computer performing trades on a stock exchange monitors and responds to internet-derived data relating to financial transactions occurring all over the world.[96] Needless to say, given the immense number of devices and the vast amount of data available on the internet, a computer that relies on internet-derived data can yield unpredictable results. As stated, one of the most difficult issues inherent in AI is how to ensure that the data relied on by the computer are in fact accurate. Not only is information that originates on the internet often inaccurate, such as information on social media, but the internet also contains data that are intentionally false, and that are often spread extensively by 'bots' and similar technologies that run automated tasks—such as spreading deliberately false and inflammatory content—at a rate much higher than is humanly possible.[97]

Because AI-enabled devices frequently use data from the internet or implement their algorithms via the internet, AI functions are especially vulnerable to cybersecurity

created-a-twitter-bot-to-learn-from-users-it-quickly-became-a-racist-jerk.html> accessed 19 March 2019.

93 Cathy O'Neil, *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy* (Crown Books 2016) 85–87.

94 Julia Angwin, Jeff Larson, Surya Mattu and Lauren Kirchner, 'Machine Bias' (*ProPublica,* 23 May 2016) <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing> accessed 17 March 2019.

95 Giuffrida (n 4) 755.

96 ibid 758.

97 Consider the allegations that the United States and other national elections have been intentionally influenced by false data, such as computer-produced or 'bot' social media communications. See, for example, Scott Shane, 'The Fake Americans Russia Created to Influence the Election' (*New York Times,* 7 September 2017) <https://www.nytimes.com/2017/09/07/us/politics/russia-facebook-twitter-election.html> accessed 17 March 2019; Kai Kupferschmidt, 'Social Media "Bots" Tried to Influence the US Election. Germany May Be Next' (*Science,* 13 September 2017) <http://www.sciencemag.org/news/2017/09/social-media-bots-tried-influence-us-election-germany-may-be-next> accessed 17 March 2019.

threats.[98] In July 2017, for example, *Forbes* reported that criminals hacked a fish tank to steal data from a casino.[99] The fish tank was connected to the internet to permit remote monitoring of water conditions, and the thieves used that connection as a route into the casino's computers.[100]

Bias and discrimination are serious issues facing AI. There already have been a number of cases of unfair treatment linked to historical data, and steps need to be taken to make sure that does not become prevalent in AI. Existing statutes governing discrimination in the physical economy need to be extended to digital platforms. This will help protect consumers and build confidence in these systems as a whole.[101]

Some individuals argue that there needs to be avenues for human beings to exercise oversight and control over AI systems. For example, Oren Etzioni, the CEO of *Allen Institute for Artificial Intelligence*, posits that there should be rules for regulating these systems. First, AI must be governed by all the laws that have already been developed for human behaviour, including regulations concerning 'cyberbullying, stock manipulation or terrorist threats' and 'entrap[ping] people into committing crimes'.[102] Second, he believes that these systems should disclose that they are automated systems and not human beings. Third, he states that an AI system 'cannot retain or disclose confidential information without [the] explicit approval of the source of that information.'[103] The *rationale* he provides is that these tools store so much data that people have to be cognisant of the privacy risks posed by AI.[104]

The Challenge of Regulating AI

The first question that arises is whether we have indeed reached the point at which we need to devise a legislative instrument on robotics and AI.[105] The classic line of thinking is that legislation becomes necessary once a societal or technological change calls for an adequate legal framework.[106] Once every home and business is equipped with an

98 Giuffrida and others (n 4) 776.

99 Lee Mathews, 'Criminals Hacked a Fish Tank to Steal Data from a Casino' (*Forbes*, 27 July 2017) <https://www.forbes.com/sites/leemathews/2017/07/27/criminals-hacked-a-fish-tank-to-steal-data-from-a-casino/#1547e65c32b9> accessed 20 March 2019.

100 ibid.

101 West and Allen (n 7).

102 Oren Etzioni, 'How to Regulate Artificial Intelligence' (*New York Times*, 1 September 2017) <https://www.nytimes.com/2017/09/01/opinion/artificial-intelligence-regulations-rules.html> accessed 19 March 2019

103 ibid.

104 ibid.

105 The Legal Affairs Committee of the European Parliament has called for the immediate creation of a legislative instrument governing robotics and AI. Nathalie Nevejans, 'European Civil Law Rules in Robotics' *(Study Commissioned by European Parliament's Legal Affairs Committee)* (2016) PE 571.379 6.

106 ibid.

autonomous robot, society will change dramatically. People will work, collaborate, interact, live and perhaps even fall in love with highly sophisticated machines.[107] We will need to consider humanity's place in the face of these technologies.[108]

In considering how to legislate in the face of the staggering rate of technological advancement, the Committee on Legal Affairs of the European Parliament takes a pragmatic approach. It proposes to adopt a legislative instrument for a period of 10–15 years, because any document that concerns a field that advances at the pace of robotics and AI would soon become obsolete. Therefore, the legislative instrument will take account only of foreseeable—and not unforeseeable—progress. It would then, of course, be imperative to review the legislative instrument once technological changes overtake current forecasts.[109]

Innovations such as the internet and networked AI have enormous short-term benefits, along with long-term negative effects that could take decades to become recognisable. AI will drive a vast range of efficiency optimisations, but also enable the hidden discrimination and arbitrary penalisation of individuals in areas such as insurance, job-seeking and performance assessment.[110] Without significant changes in our political economy and governance regimes, AI is likely to create greater economic inequalities, more surveillance and more programmed and non-human-centric interactions.[111] As to liberty, there are clear risks. AI affects agency by creating entities with meaningful intellectual capabilities for monitoring, enforcing and even punishing individuals. Those who know how to use it will have immense potential power over those who do not or cannot.[112]

Governments around the world are already mobilising. In 2015 the Japanese government announced a 'New Robot Strategy', which has strengthened collaboration between industry, government and academia. In late 2016, the government of the United Kingdom created a parliamentary group—the All Party Parliamentary Group on Artificial Intelligence—to explore the impact and implementation of AI, including ML.[113] Also in late 2016, the Obama administration released the reports, *Artificial*

107 ibid.

108 ibid

109 ibid 7.

110 Andrew McLaughlin, as quoted in Lindsey Andersen and others (n 20).

111 Marina Gorbis, as quoted in ibid.

112 Greg Shannon, as quoted in ibid.

113 Related to this, the House of Commons Science and Technology Committee has stated: 'While it is too soon to set down sector-wide regulations for this nascent field, it is vital that careful scrutiny of the ethical, legal and societal dimensions of artificially intelligent systems begins now.' See also House of Lords Select Committee on Artificial Intelligence, as referred to in Mark Deem, 'Law Vital to the Future of Artificial Intelligence' (*ICAEW Economia,* 18 October 2017) <https://economia.icaew.com/opinion/october-2017/law-vital-to-the-future-of-artificial-intelligence> accessed 18 April 2019.

Intelligence, Automation, and the Economy and *Preparing for the Future of Artificial Intelligence*. These reports consider the challenge for policy-makers in updating, strengthening and adapting policies to respond to the economic effects of AI.[114] In February 2017, the European Parliament approved a report of its Legal Affairs Committee calling for the review of draft legislation to clarify liability issues, especially for driverless cars. It also called for the creation of a specific legal status for robots (so-called 'electronic persons') to be considered in order to establish who would be liable if they cause damage.[115]

There are, broadly speaking, two schools of thought on the issue of the regulation of AI.[116] The first is based on the premise that regulation is bad for innovation.[117] Entrepreneurs in this camp do not want the field of AI to be defined too soon, and certainly not by non-technical people. Among their concerns are that bad policy created bad technology, regulation stifles innovation and regulation is premature because we do not yet have any clear sense of what we would be regulating.[118]

The other school of thought seeks to protect against potentially harmful creations that poison the well for other AI entrepreneurs.[119] Subscribers to this school believe that national governments should act expeditiously to promote existing standards and guidelines or, where necessary, create new guidelines, to ensure a basic respect for the principle of 'first, do no harm'.[120]

Rapid innovations in technology far exceed the ability of the world's domestic and international legal systems to keep pace.[121] The law is often criticised for trailing technology by decades. Given the pace of technological innovation and its potential implications, we cannot afford to be in the same boat this time.[122]

---

114 Carole Piovesan, 'Speaker's Corner: Lawyers Need to Keep Up With AI' (*Law Times*, 5 June 2017) <https://www.lawtimesnews.com/author/na/speakers-corner-lawyers-need-to-keep-up-with-ai-13408/> accessed 12 April 2019.

115 ibid.

116 Joshua New, 'How (and How Not) to Fix AI' (*TechCrunch* 26 July 2018) <https://techcrunch.com/2018/07/26/how-and-how-not-to-fix-ai/> accessed on 3 July 2020.

117 Piovesan (n 114).

118 ibid.

119 Martin Wright, 'First, Do No Harm: Regulators and Tech Industry Scramble to Tame the AI Tiger' (Reuters, 16 January 2018) <https://www.ethicalcorp.com/first-do-no-harm-regulators-and-tech-industry-scramble-tame-ai-tiger> accessed 3 July 2020.

120 ibid.

121 Clayton Rice, 'Artificial Intelligence' (6 January 2016) <https://www.claytonrice.com/artificial-intelligence> accessed 23 April 2019.

122 Piovesan (n 114).

The key for humanity in general and lawyers specifically will be to develop the positive aspects of the technology while managing its risks and challenges.[123] AI regulation will be a necessity, particularly in the areas of safety and errors, liability laws and social impact.[124] Policy-makers will have to embrace the benefits that AI can bring, but at the same time they must be sensitive in order to pre-empt the dramatic and potentially devastating effects of misusing AI.[125]

Conclusion

AI may well be a revolution in human affairs and become the single most influential innovation in history.[126] There already are significant deployments of AI and data analytics in finance, national security, healthcare, criminal justice, transportation and smart cities, that have altered decision-making, business models, risk mitigation and system performance.[127] These developments are generating substantial economic and social benefits.

By the same token, however, the manner in which AI systems unfold has major implications for society as a whole. It matters how policy issues are dealt with, ethical conflicts reconciled, legal realities resolved, and how much transparency is required in AI and data-analytic solutions.[128] Human choices about software development affect the way in which decisions are made and the manner in which they are integrated into organisational routines. Exactly how these processes are executed needs to be better understood, because they will have a substantial impact on the general public soon, and for the foreseeable future.[129]

In the regulation of AI, legal systems should avoid imposing a rigorous regulatory regime that bans outright the production of certain AI systems, but, by the same token, it should also provide strong incentives for AI developers to incorporate adequate safeguards.[130] Some of these safeguards might include:

- when a government seeks to acquire an AI system, procurement should be done in an open and transparent manner, including publication of the purpose of the

123 Anthon P Botha, 'Artificial Intelligence II: The Future of Artificial Intelligence' (*Foresight for Development*, undated) <http://www.foresightfordevelopment.org/featured/artificial-intelligence-ii> accessed 9 April 2019.

124 ibid.

125 ibid.

126 ibid.

127 ibid.

128 ibid.

129 ibid.

130 See Matthew Scherer, 'Regulating Artificial Intelligence Systems: Risks, Challenges, Competencies, and Strategies' (2016) 29 Harvard J L & T 255–400 at 398.

system, and its goals, parameters and other information to enable public understanding;

- governments must thoroughly investigate AI systems in order to identify risks to society before developing or acquiring such systems, and on an ongoing basis throughout the lifecycle of the systems;
- maximum possible transparency is necessary for any AI system, including its purpose, how it is used and how it works, which must continue throughout the lifecycle of the system; and
- the fact that an AI system performs a task previously performed by a human being does not remove the requirements for accountability in government decision-making processes—there should preferably always be a human being in the loop, and for high-risk areas, such as criminal justice, significant human oversight will be necessary.[131]

Because there is no way to accurately predict either the pace of AI development or the capabilities of AI systems in the long term, the pragmatic approach of the Committee on Legal Affairs of the European Parliament seems sensible. Regulation should take account only of foreseeable—and not unforeseeable—progress. However, such an approach would necessitate periodic review of the regulatory instrument whenever technological changes overtake current forecasts.

Moreover, it is clear that AI—whether in the form of autonomous vehicle systems, lethal autonomous systems, automated surveillance techniques or powerful data-mining applications—transcends national borders. There is a danger that differing domestic approaches might conflict, raising significant difficulties for those affected by more than one regime.[132] National efforts to develop AI regulatory policies should be coordinated and supported by an international regulatory framework to avoid the risks that stem from the imperfect interaction of fragmented domestic regulatory approaches.[133]

These are tentative and general proposals that are meant to start a conversation rather than to be the final word. There is an ancient Chinese saying, ‘May you live in interesting times.’ We can say without doubt that we do. We would do well, however,

131 Lindsey Andersen and others (n 20). ‘Human Rights in the Age of Artificial Intelligence’ (*Access Now*, November 2018) <https://www.accessnow.org/cms/assets/uploads/2018/11/AI-and-Human-Rights.pdf > accessed 2 March 2020.

132 See, for example, Olivia Erdélyi and Judy Goldsmith, ‘Regulating Artificial Intelligence: Proposal for a Global Solution’ Conference on Artificial Intelligence, Ethics and Society, 2–3 February 2018, New Orleans, Louisiana, USA 95–101.

133 ibid.

to recognise that the saying is usually uttered as a curse. Let us work proactively to ensure that, legally at least, AI might prove to be a blessing and not a curse.[134]

[134] Giuffrida and others (n 4) 781.

References

Alarie B, Niblett A and Yoon HY, 'How Artificial Intelligence Will Affect the Practice of Law' (2018) 68(1) University of Toronto LJ 116 <https://doi.org/10.3138/utlj.2017-0052>

Andersen L, 'Human Rights in the Age of Artificial Intelligence' (*Access Now*, November 2018) <https://www.accessnow.org/cms/assets/uploads/2018/11/AI-and-Human-Rights.pdf >.

Anderson J, Raine L and Luchsinger A, 'Artificial Intelligence and Future of Humans' (*Pew Research Center Internet and Technology*, 10 December 2018) <https://www.pewinternet.org/2018/12/10/artificial-intelligence-and-the-future-of-humans/>.

Angwin J, Larson J, Mattu S and Kirchner L, 'Machine Bias' (*ProPublica*, 23 May 2016) <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>.

Arthur C, 'iPhone Keeps Record of Everywhere You Go' (*The Guardian*, 20 April 2011) <https://www.theguardian.com/technology/2011/apr/20/iphone-tracking-prompts-privacy-fears>.

Artificial Intelligence Primer (Victorian All-Party Parliamentary Group on Artificial Intelligence, February 2018) 2 <https://www.parliament.vic.gov.au/images/stories/AI-Primer_Feb2018.pdf>.

Beazley Justice MJ, 'Law in the Age of Algorithm' (27 September 2017) *State of the Profession Address, New South Wales Young Lawyers, Sydney* 1.

Betts KD and Kyle RJ, 'The Dawn of Fully Automated Contract Drafting: Machine Learning Breathes New Life into a Decades-Old Promise' (2017) 15 Duke L & Tech Rev 216.

Botha AP, 'Artificial Intelligence II: The Future of Artificial Intelligence' (*Foresight For Development*, undated) <http://www.foresightfordevelopment.org/featured/artificial-intelligence-ii>.

Burdon M, 'Interview with Mark Burdon: Artificial Intelligence and the Law' (*Justice and The Law Society, The University of Queensland*, 8 March 2018) <http://www.jatl.org/blog/2018/3/8/interview-with-mark-burdon-artificial-intelligence-and-the-law>.

Carbone C, 'Creepy AI Reportedly Spot Shoplifters Before They Steal' (*Fox News*, 18 April 2019) <https://foxnews.com/tech/creepy-ai-will-reportedly-spot-shoplifters-before-they-steal>.

Casey AJ and Niblett A, 'Self-driving Laws' (2016) 66 University of Toronto LJ 429 <https://doi.org/10.3138/UTLJ.4006>

Chan R, 'One Chinese City is Using Facial-Recognition That Can Help Police Detect and Arrest Criminals in as Little as 2 Minutes' (*Business Insider*, 19 March 2018) <https://www.businessinsider.com/china-guiyang-using-facial-recognition-to-arrest-criminals-2018-3>.

Chen Y and Ahn MJ, Routledge Handbook of Information Technology in Government (Routledge 2017) <https://doi.org/10.4324/9781315683645>

Chesney B and Citron D, 'Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security' (forthcoming 2019) 107 California LR <https://doi.org/10.2139/ssrn.3213954>

Deem M, 'Law Vital to the Future of Artificial Intelligence' (*ICAEW Economia*, 18 October 2017) <https://economia.icaew.com/opinion/october-2017/law-vital-to-the-future-of-artificial-intelligence>.

Denyer S, 'In China, Facial Recognition is Sharp End of a Drive for Total Surveillance' (*Washington Post*, 7 January 2018) <https://www.washingtonpost.com/news/world/wp/2018/01/07/feature/in-china-facial-recognition-is-sharp-end-of-a-drive-for-total surveillance/?utm_term=.0b9cac9ecab4>.

Editorial, 'Joy Buolamwini' (*Bloomberg Businessweek*, 3 July 2017).

Erdélyi O and Goldsmith J, 'Regulating Artificial Intelligence: Proposal for a Global Solution' *Conference on Artificial Intelligence, Ethics and Society*, 2–3 February 2018, New Orleans, Louisiana, United States, 95 <https://doi.org/10.1145/3278721.3278731>

Etzioni O, 'How to Regulate Artificial Intelligence' (*New York Times*, 1 September 2017) <https://www.nytimes.com/2017/09/01/opinion/artificial-intelligence-regulations-rules.html>.

European Parliament Committee on Legal Affairs, *Report with Recommendations to the Commission on Civil Law Rules on Robotics* (PE582.443v03-00) (2017).

Executive Office of the President, *Preparing for the Future of Artificial Intelligence* (October 2016).

Giuffrida I, Lederer F and Vermerys N, 'A Legal Perspective on the Trials and Tribulations of AI: How Artificial Intelligence, the Internet of Things, Smart Contracts, and Other Technologies Will Affect the Law' (2018) 68 Case Western Reserve LR.

Glusac E, 'As Airbnb Grows, So Do Claims of Discrimination' (*New York Times*, 21 June 2016) <https://www.nytimes.com/2016/06/26/travel/airbnb-discrimination-lawsuit.html>.

Goode L, 'A Guide to Using Apple Watch's Heart Rate Features, Including ECG' (*Wired,* 6 December 2018) <https://www.wired.com/story/how-to-take-an-ecg-reading-on-apple-watch/>.

Gravett WH, 'The Myth of Rationality: Cognitive Biases and Heuristics in Judicial Decision-Making' (2017) 134 South African Law Journal.

Hern A, 'New AI Fake Text Generator May Be Too Dangerous to Release, Say Creators' (*The Guardian,* 14 February 2019) <https://www.theguardian.com/technology/2019/feb/14/elon-musk-backed-ai-writes-convincing-news-fiction>.

Kay KN, Naselaris T, Prenger RJ and Gallant JL, 'Identifying Natural Images from Human Brain Activity' (2008) 452 Nature <https://doi.org/10.1038/nature06713>

Kolber AJ, 'Will There Be a Neurolaw Revolution?' (2014) 89 Indiana LJ.

Kuang C, 'Can AI Be Taught to Explain Itself' (*New York Times Magazine,* 21 November 2017) <https://www.nytimes.com/2017/11/21/magazine/can-ai-be-taught-to-explain-itself.html>.

Kupferschmidt K, 'Social Media "Bots" Tried to Influence the US Election. Germany May Be Next' Science (13 September 2017) <http://www.sciencemag.org/news/2017/09/social-media-bots-tried-influence-us-election-germany-may-be-next> <https://doi.org/10.1126/science.aap9514>

Lam O, 'With "Sharp Eyes", Smart Phones and TV Sets Are Watching Chinese Citizens' (*ADVOX,* 3 April 2018) <https://advx.globalvoices.org/2018/04/03/with-sharp-eyes-smart-phones-and-tv-sets-are-watching-chinses-citizens/>.

Levin S, 'New AI Can Guess Whether You're Gay or Straight from a Photograph' (*The Guardian,* 7 September 2017) <https://www.theguardian.com/technology/2017/sep/07/new-artificial-intelligence-can-tell-whether-youre-gay-or-straight-from-a-photograph>.

Lewis N, 'Should AI Be Used to Catch Shoplifters?' (*CNN,* 18 April 2019) <https://edition.cnn.com/2019/04/18/business/ai-vaak-shoplifting/index.html>.

Liao S, 'China Banned Millions of People with Poor Social Credit from Transportation in 2018' (*The Verge,* 1 March 2019) <https://www.theverge.com/2019/3/1/18246297/china-transportation-people-banned-poor-social-credit-planes-trains-2018>.

Loh E, 'Medicine and the Rise of the Robots: A Qualitative Review of Recent Advances of Artificial Intelligence in Health' (2018) 2 BMJ Leader <https://doi.org/10.1136/leader-2018-000071>

Lubin G, ‘Facial-profiling Could Be Dangerously Inaccurate and Biased, Experts Warn’ (*Business Insider,* 12 October 2016) <https://www.businessinsider.com/does-faception-work-2016-10>.

Mathews L, ‘Criminals Hacked a Fish Tank to Steal Data from a Casino’ (*Forbes,* 27 July 2017) <https://www.forbes.com/sites/leemathews/2017/07/27/criminals-hacked-a-fish-tank-to-steal-data-from-a-casino/#1547e65c32b9>.

McCarthy J, Minsky ML, Rochester N and Shannon CE, ‘A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence, 31 August 1955’ (2006) 27(4) AI Magazine.

Merchant GE, ‘Artificial Intelligence and the Future of Legal Practice’ (2017) 14(1) Scitech Lawyer.

Millán L, ‘Artificial Intelligence’ (*Canadian Lawyer Magazine,* 3 April 2017) <https://www.canadianlawyermag.com/article/artificial-intelligence-3585>.

Nevejans N, ‘European Civil Law Rules in Robotics’ (*Study Commissioned by European Parliament’s Legal Affairs Committee*) (2016) PE 571.379.

New J, ‘How (and How Not) to Fix AI’ (*TechCrunch* 26 July 2018) <https://techcrunch.com/2018/07/26/how-and-how-not-to-fix-ai/>.

Nicas J, ‘Google Has Picked an Answer for You – Too Bad It’s Often Wrong’ (*Wall Street Journal,* 16 November 2017) <https://www.wsj.com/articles/googles-featured-answers-aim-to-distill-thruthbut-often-get-it-wrong-1510847867>.

Nishimoto S, Vu AT, Naselaris T, Benjamini Y, Yu Band Gallant JL, ‘Reconstructing Visual Experiences from Brain Activity Evoked by Natural Movies’ (2011) 21 Current Biology 1641 <https://doi.org/10.1016/j.cub.2011.08.031>

Nisith Desai Associates, *The Future Is Here: Artificial Intelligence and Robotics* (May 2018) <http://www.nishithdesai.com/fileadmin/user_upload/pdfs/Research_Papers/Artificial_Intelligence_and_Robotics.pdf>.

O’Neil C, Weapons of Math Destruction (Crown 2017).

Pasquale F and Cashwell G, ‘Four Futures of Legal Automation’ (2015) 63 UCLA LR.

Piovesan C, ‘Speaker’s Corner: Lawyer Need to Keep Up With AI’ (*Law Times,* 5 June 2017) <https://www.lawtimesnews.com/author/na/speakers-corner-lawyers-need-to-keep-up-with-ai 13408/>.

Randall I, 'First "Deepfake" AI That Can Replicate People Moving Creates Footage of Crowds of Imaginary Humans That Are Indistinguishable From The Real Thing' (*Mail Online,* 7 May 2019) <https://www.dailymail.co.uk/sciencetech/article-7001293/Deepfake-AI-replicate-bodies-motion-creates-footage-crowds-imaginary-people.html>.

Rankin D, 'How Artificial Intelligence Could Change the Law in Three Major Ways' (*The Journal of Law and Technology at Texas,* 14 October 2018) <http://jolttx.com/2018/10/14/how-artificial-intelligence-could-change-the-law-in-three-major-ways/>.

Richardson R, Schultz JM and Crawford K, 'Dirty Data, Bad Predictions: How Civil Rights Violations Impact Police Data, Predictive Policing Systems and Justice' (2019) 94 New York University Law Review.

Rice C, 'Artificial Intelligence' (6 January 2016) <https://www.claytonrice.com/artificial-intelligence>.

Richie DR II and Duffy JD, 'Artificial Intelligence in the Legal Field' (25 April 2018) *Association of Corporate Counsel Greater Philadelphia In-House Counsel Conference.*

Sanito J, Pathak S and Fernandez R, 'Deep Learning Explained' (*edX,* 4 March 2018) <https://www.edx.org/course/deep-learning-explained-microsoft-dat236x-1>.

Scherer M, 'Regulating Artificial Intelligence Systems: Risks, Challenges, Competencies, and Strategies' (2016) 29 Harvard J L & T 255 <https://doi.org/10.2139/ssrn.2609777>

Shane S, 'The Fake Americans Russia Created to Influence the Election' (*New York Times,* 7 September 2017) <https://www.nytimes.com/2017/09/07/us/politics/russia-facebook-twitter-election.html>.

Sheppard B, 'Warming Up to Inscrutability: How Technology Could Change Our Concept of Law' (2018) 68 University of Toronto LJ 36 <https://doi.org/10.3138/utlj.2017-0053>

Solum LB, 'Artificial Meaning' (2014) 89 Washington University LR.

Sorger B, Reithler J, Dahmen Band Goebel R, 'A Real-Time fMRI-Based Spelling Device Immediately Enabling Robust Motor-Independent Communication' (2012) 22 Current Biology 1333 <https://doi.org/10.1016/j.cub.2012.05.022>

Tegmark M, 'Benefits and Risks of Artificial Intelligence' (*Future of Life,* 2016) <http://www.futureoflife.org/background/benefits-risks-of-artificial-intelligence/>.

Turing AM, 'Mind' (1950) 59(236) Computing Machinery and Intelligence <https://doi.org/10.1093/mind/LIX.236.433>

Victor D, 'Microsoft Created a Twitter Bot to Learn from Users. It Quickly Became a Racist Jerk' (*New York Times*, 24 March 2016) <https://www.nytimes.com/2016/03/25/technology/microsoft-created-a-twitter-bot-to-learn-from-users-it-quickly-became-a-racist-jerk.html>.

Vincent J, 'ThisPersonDoesNotExist.com Uses AI to Generate Endless Fake Faces' (*The Verge*, 15 February 2019) <https://www.theverge.com/tldr/2019/2/15/18226005/ai-generated-fake-people-portraits-thispersondoesnotexist-stylegan>.

Vincent J, 'Watch Jordan Peele Use AI to Make Barack Obama Deliver a PSA About Fake News' (*The Verge*, 17 April 2018) <https://www.theverge.com/tldr/2018/4/17/17247334/ai-fake-news-video-barack-obama-jordan-peele-buzzfeed>.

Von Portz G and Misra S, 'Medtronic & Fitbit Partner to Connect Activity Data with Continuous Glucose Monitoring' (*iMedical Apps*, 24 January 2017) <https://www.imedicalapps.com/2017/01/medtronic-fitbit-partner-connect-activity-data-continuous-glucose-monitoring/>.

West DM and Allen JR, 'How Artificial Intelligence is Transforming the World' (*Brookings Report*, 24 April 2018) <https://www.brookings.edu/research/how-artificial-intelligence-is-transforming-the-world/>.

Wylie C, *Mindf*ck: Cambridge Analytica and the Plot to Break America* (Random House 2019).