Office of the Victorian Information Commissioner

Artificial intelligence and privacy

Issues paper

June 2018

Artificial intelligence and privacy

Issues paper

Artificial intelligence and privacy

Issues paper

1. Introduction

Artificial Intelligence (**AI**) at its most simple, is a sub-field of computer science with the goal of creating programs that can perform tasks generally performed by humans. These tasks can be considered intelligent, and include visual and audio perception, learning and adapting, reasoning, pattern recognition and decision-making. 'AI' is often used as an umbrella term to describe a collection of related techniques and technologies including *machine learning, predictive analytics, natural language processing* and *robotics.*

While the philosophy of Artificial Intelligence has been argued since at least Leibnitz in the early 18th Century, the concept of AI as we use it has existed since the early 1940s and made famous with the development of the "Turing test" in 1950. More recently, we are experiencing a period of rapid development in the field of AI as a result of three factors: improved algorithms, increased networked computing power, and increased ability to capture and store an unprecedented amount of data.[1] As well as technological advancements, the very way of thinking about intelligent machines has shifted significantly since the 1960s, which has enabled many of the developments we are seeing today.

Real-life applications of AI technologies are already established in our everyday lives, although many people are not conscious of this. One of the characteristics of AI is that once the technology works, it stops being referred to as AI and transforms into mainstream computing.[2] For example, being greeted by an automated voice on the other end of the phone, or being suggested a movie based on your preferences, are examples of mainstream AI technology. Now that these systems are an established element in our lives, the fact that AI techniques – including speech recognition, natural language processing and predictive analytics – are at work is often forgotten.

The ways that AI can enrich our lives are immense. Increased efficiency and lower costs, huge improvements in healthcare and research, increased safety of vehicles, and general convenience, are just some of the promises of AI. But, as with any new technology, the opportunities of AI come with an array of challenges for society and the law.[3]

1.1. Purpose of this paper

This issues paper is an introduction to a wider conversation regarding information privacy and AI. It is written for a non-technical audience and does not endeavour to solve questions posed, nor provide legal guidance. It should be noted that there are many other ethical, technical and legal issues associated with AI that are beyond the scope of this document. The final page of the paper contains a list of suggested further readings, some of which delve into these other important issues.

[1] Alex Campolo, Madelyn Sanfilippo, Meredith Whittaker & Kate Crawford, 'AI Now 2017 Report', *AI Now*, 2017, available at: https://ainowinstitute.org/AI_Now_2017_Report.pdf, p 3.
[2] Toby Walsh, *It's Alive! Artificial Intelligence from the logic piano to killer robots*, Latrobe University Press, 2017, p 60.
[3] For example, Samuel Warren and Louis Brandeis wrote on the impact of the portable camera on the right to be let alone in the 19th century. See Samuel D. Warren and Louis D. Brandeis, 'The Right to Privacy', *Harvard Law Review*, Vol. IV, No. 6, 15 December 1890.

The purpose of this document is to:

- provide a high-level understanding of AI and its uses in the public sector, and
- highlight some of the challenges and opportunities that AI presents in relation to information privacy.

For the purpose of this document, the discussion is generally limited to information privacy, which is a subset of the more broad, abstract concept of privacy. In Victoria and more widely, the legislative approach to protecting privacy is focused on information privacy, rather than other types, such as physical privacy.[4] Information privacy relates to the level of control that one has over their personal information in determining when, how and for what purposes, it is used.

2. Terminology

There is a significant amount of terminology and technical jargon surrounding AI that is often used interchangeably and can cause confusion, especially for those without a technical background. Below is a simple explanation of key terms designed to assist the everyday reader understand some of the terminology surrounding AI, and the discussion within this document. This list is neither exhaustive, nor intended to be technologically in-depth.

2.1. Narrow, general and super artificial intelligence

Most AI that we experience today is considered to be 'narrow'. This means that it has been deliberately programmed to be competent in one specific area. It is sometimes also referred to as augmented intelligence to highlight its ability to enhance (but not necessarily replace) human intelligence. For example, a computer developed by IBM in the 1980s called Deep Blue can play chess at a level superior to human beings; a feat of huge importance in the timeline of AI development. However, while Deep Blue exhibits an above-human ability in chess, its intelligence ends there.

Conversely, the concept of artificial general intelligence (**AGI**) refers to a level of intelligence across multiple fields. The distinction between narrow and general intelligence is already apparent in the natural world: for instance, bees know how to build beehives, and ants know how to build a nest – both of which are examples of intelligence in a narrow sense. However, this intelligence is specific to a certain domain; bees can't build a nest and ants cannot build a hive. Humans, on the other hand, have the capacity to be intelligent across a range of areas, and can learn intelligence in new fields through experience and observation.

Building upon the idea of AGI, artificial superintelligence is generally regarded as AI that is both general and exceeds human levels of intelligence. A notable writer on this subject, Nick Bostrom, defines superintelligence as "an intellect that is much smarter than the best human brains in practically every field, including scientific creativity, general wisdom and social skills."[5] Many pop culture depictions of AI, such as in films *Ex Machina* and *Her*, display AI in the form of superintelligence. This kind of portrayal can contribute to the hype and/or fear surrounding AI, and while it is a popular idea in science fiction, there is significant debate regarding the likelihood, imminence and consequences of ever developing such technology.

[4] For more information on information privacy law in Australia, see *Privacy Background Paper*, 2015, available at http://www.ovic.vic.gov.au/.
[5] Nick Bostrom, 'How long before superintelligence?', *Linguistic and Philosophical Investigations*, Vol. 5, No.1, 2006, pp 11-30.

For the purpose of this issues paper, the scope of the discussion is limited to narrow AI, which will simply be referred to as AI hereafter.

2.2. Big data

The relationship between AI and big data goes two ways. While big data analytics processes already exist, much of big data's true value is only able to be realised using AI techniques. In the other direction, big data offers AI an immense and rich source of input data to develop and learn from. In this sense, AI and big data are strongly intertwined.

There is no one established definition of big data, however it is generally used to describe massive amounts of data produced and collected in a variety of forms.[6] The types and scale of information included under the term 'big data' cannot be understated; almost everything individuals do generates data – searching online; sharing and transmitting day to day information with government, companies and social media; even just walking around with a smartphone – all (intentionally or unintentionally) create vast amounts of information about individuals. As the Internet of Things (**IoT**) pushes the network further into our physical environment and personal spaces, the scope of data created, collected and fed into AI systems stands to delve further into our personal lives.

The Information Commissioner's Office of the United Kingdom sums up the connection between AI and big data quite eloquently:

> *Big data can be seen as an asset that is difficult to exploit. AI can be seen as a key to unlocking the value of big data; and machine learning is one of the technical mechanisms that underpins and facilitates AI.*[7]

2.3. Machine learning

Machine learning is a computer science technique that allows computers to 'learn' on their own. It is often characterised as AI, but that is only one element of it. The characteristic that separates machine learning from other forms of AI is its dynamic ability to modify itself when exposed to more data.[8] Through ingesting data, the machine is training itself by developing its own logic according to the data it has analysed.

There are two main types of machine learning: *supervised* and *unsupervised.* Supervised learning requires a human to provide both the data and the solution, letting the machine determine the connection between the two. Unsupervised learning allows the machine to learn more freely by ingesting a large amount of data (often big data) and iterating over it to find patterns and insights.

For example, you might be interested in predicting the price of a house. To do this you could tell the machine to look at a variety of features such as the number of rooms, if there is a garden etc. Using a supervised learning technique, you would also provide the historical prices of comparable houses, so the algorithm can build a model to understand the relationship between certain features and price, and therefore be able to reasonably predict a house price based on those features. In an unsupervised learning context, the machine would not be provided with the historical house prices, nor told which features are important to consider – rather, it would determine the patterns on its own.

These techniques are used in different contexts and for varying purposes. Neither requires explicit programming on what to look for, which gives a level of autonomy to the system to generate its own logic,

---

[6] A thorough explanation of big data can be found in the Report of the Special Rapporteur on the right to privacy, prepared for the Human Rights Council, A/72/43103, October 2017.
[7] The UK Information Commissioner's Office (ICO), *Big Data, artificial intelligence, machine learning and data protection*, 2017, p 8.
[8] See for example: *What's the difference between Artificial Intelligence, Machine Learning and Deep Learning?,* available at https://deeplearning4j.org/ai-machinelearning-deeplearning, last accessed 17 April 2018.

identifying trends that may otherwise have been missed by humans.[9] Machine learning algorithms are already widely used in modern life. Some examples include producing web search results, suggestive services such as Netflix and Pandora, and predicting the monetary value of a product given the existing market. The extent to which machine learning is useful is determined by the input data provided. Because of this, big data has played a pivotal role in the success of machine learning.

2.4. Deep learning

Deep learning is a subset of machine learning, most commonly used to refer to deep neural networks.[10] In generalist terms, a neural network processes data through a layered approach, where each successive layer takes its input from the output of the layer before it. The term deep refers to the number of layers in the neural network.

As the output of each layer becomes the input of the next, it can become increasingly difficult to understand the decisions and inferences made at each level. The process of going through each layer can create what is referred to as the 'black box' effect, making it challenging to truly understand and describe the steps that lead to a particular outcome.[11] The human brain is often used as an analogy to explain neural networks, however this is not particularly helpful as it implies machines understand information in a similar manner to human thinking, which is not the case.

Deep learning is an extremely powerful tool, and many credit it for the recent explosion of AI. It has given computers the ability to recognise spoken words almost as well as a human, transformed computer vision and dramatically improved machine translation – abilities that are far too complex to code into machines by hand. The nature of this process presents challenges for transparency of decisions, as the logic can become increasingly obscure to the human eye with each layer of processing. Further, neural networks are not immune to bias. For example, a recurrent neural network (**RNN**) will take data it has previously been exposed to into consideration.[12] Some describe RNNs as having a memory, which similarly to human beings, affects its output. For example, in 2016 Microsoft trained an AI bot using a RNN on Twitter data, which demonstrated the potential for unintended consequences of this way of learning.[13]

3. Artificial intelligence in the public sector

While development of AI technology is being driven mainly by industry and academic research, AI applications and development are also relevant to the public sector. Government already uses AI in many areas, but it stands to benefit from further adoption of these technologies. Further, government has a significant role to play in shaping how AI technologies impact citizens' lives through regulation, policy, and demonstrating best practices. It is important that government is not left behind as the private sector steams ahead – this means taking a proactive, dynamic and informed approach to the technology and its interaction with law and society.

The current and future use cases of AI in government remain bounded by resources, technical capability and public trust. Some of the most immediate beneficial opportunities for the public sector are those where AI can reduce administrative burdens and help resolve resource allocation problems. In the short-term, AI applications have the potential to be immensely useful in increasing efficiency of established government process such as answering questions, filling out and searching documents, routing requests,

[9] Will Knight, 'The Dark Secret at the Heart of AI', *MIT Technology Review,* 11 April 2017, available at https://www.technologyreview.com/s/604087/the-dark-secret-at-the-heart-of-ai/.
[10] It is also less frequently used to refer to deep reinforcement learning. For an explanation of both, see *Introduction to Deep Neural Networks*, available at https://deeplearning4j.org/neuralnet-overview.
[11] ICO, *Big Data, artificial intelligence, machine learning and data protection*, 2017, p 11.
[12] For more information on recurrent neural networks, see https://deeplearning4j.org/lstm.html.
[13] Elle Hunt, 'Tay, Microsoft's AI chatbot, gets a crash course in racism from Twitter', *The Guardian*, March 2016, available at: https://www.theguardian.com/technology/2016/mar/24/tay-microsofts-ai-chatbot-gets-a-crash-course-in-racism-from-twitter.

translation and drafting documents.[14] As an example, the use of chat bots to provide customer service and advice to individuals already occurs in some of the larger Australian government organisations.

In the longer term, AI has the potential to go beyond merely enhancing established processes and alter government operations altogether. It is likely to require organisations to adapt to evolving citizen needs and expectations, and to alter the regulatory and legislative landscape to make way for new uses of technology.

While AI promises many opportunities for the public sector, it cannot be seen as a panacea to all existing challenges of government. The use and regulation of AI technologies need to be implemented strategically and thoughtfully, with particular care given to information management including privacy, protective data security, and ethics more broadly.[15]

4. Privacy considerations

This section explores some of the key questions prompted by AI in relation to information privacy. This is not an exhaustive exploration of all issues; rather, it is designed to provide an overview and act as a launch pad for further discussion regarding some of the more prominent information privacy considerations.

In Victoria and more broadly, information privacy law is generally based on the 1980 *OECD Guidelines on the Protection of Privacy and Transborder Flows of Personal Data*. These guidelines contain eight key principles that continue to be enshrined in privacy law around the world, including the *Privacy and Data Protection Act 2014* (**PDP Act**). One of the benefits of having principle-based legislation is that it recognises the complicated and nuanced nature of privacy, and allows a degree of flexibility in how privacy can be protected in varying contexts and alongside evolving technologies and societal norms. While the OECD Guidelines have been remarkably successful in promoting information privacy legislation around the world, AI presents challenges to the underlying principles upon which the Guidelines are based.

While traditional notions of privacy may be challenged by AI, it is not a given that AI must undermine privacy by default; it is possible to envisage a future in which AI can help enable privacy. For instance, it is likely to mean that less people will actually need access to raw data in order to work with it, which could in turn minimise the risk of privacy breaches due to human error. It could also empower more meaningful consent, in which individuals receive personalised services dependent on privacy preferences that have been learnt over time.

The increased use of AI may require the status quo of privacy protection to be revisited, however it does not mean privacy will cease to exist or become irrelevant.

One important factor of information privacy is that it provides an important framework for making ethical choices about how we use new technologies. Considering the ethics of technology and solving the privacy challenges will be essential to the long-term success of AI. A balance between technological innovation and privacy considerations will promote the development of socially responsible AI that can assist in the creation of public value in the long term.

4.1. Why is AI different?

Emerging technology almost always brings with it important privacy considerations, yet the scale and application of AI creates a unique and unprecedented environment of challenges to information privacy. In some ways, the implications of AI can be seen as an extension of those created by big data, yet AI

[14] Hila Mehr, 'Artificial Intelligence for Citizen Services and Government', *Harvard Ash Center for Democratic Governance and Innovation*, August 2017, available at: https://ash.harvard.edu/files/ash/files/artificial_intelligence_for_citizen_services.pdf.
[15] Ibid., p 10.

technology brings with it not only the ability to process huge amounts of data, but also to use it to learn, develop adaptive models and make actionable predictions – much of this without transparent, explainable processes.

The development of AI technology brings with it a significant risk of the assumptions and biases of the individuals and companies that create it influencing the outcome of the AI. Unintended consequences caused by biases and opaque results from using neural networks pose challenges for government organisations wishing to use this technology for decision making purposes. The possibility for discrimination and how this interacts with privacy is discussed further below.

Another key point of differentiation between AI and existing analytics technologies is the potential to automate all of these areas. Where humans have historically been able to exercise a high degree of control over data processing, the increased use of AI means this may no longer be the case. Further, the application of AI to existing technologies stands to profoundly alter their current use and privacy considerations. For example, the use of CCTV cameras in public spaces for surveillance is a relatively widespread practice and not considered to be unreasonably intrusive in modern society. However, combined with the use of facial recognition software, a network of cameras could be transformed into a tool that is much more privacy invasive.

AI also has the potential to change the way that humans interact with machines. For instance, a lot of AI already embodies human characteristics. The use of anthropomorphic interfaces, such as human sounding voices used in assistants such as Alexa and Siri, may raise novel privacy concerns. Social science research indicates people are inclined to interact with technology as if it were human.[16] This means people may be more likely to develop trusting relationships with AI designed to replicate human characteristics, and consequently be more inclined to share increasingly personal information as compared with other forms of technology that collect information in a traditional manner.

Much of information privacy discourse around AI has not accounted for the growing power asymmetries between institutions that accumulate data, and the individuals who generate it.[17] Current models generally treat data as a good that can be traded, which does not fully acknowledge the difficulty for people to make decisions about their data when dealing with systems they do not understand – particularly when the system understands them well and has learnt, by way of ingesting their data, how to manipulate their preferences. Further, many adaptive algorithms used in AI change constantly, to the extent that often those who create them cannot fully explain the results they generate.

Established notions of information privacy are based on the idea that humans are the primary handlers of information and were not designed to contend with the computational ability of AI that does not conform to traditional ideas of data collection and handling.[18] The way we currently think about concepts such as informed consent, notice, and what it means to access or control personal information have never before been so fundamentally challenged as they are by AI. As highlighted above, incorporating privacy considerations as part of an ethical framework could assist in the creation of AI that does not undermine information privacy as these concepts evolve.

[16] Stanford University, 'Artificial Intelligence and Life in 2030', *One Hundred Year Study on Artificial Intelligence: Report of the 2015-2016 Study Panel*, Section III: Prospects and Recommendations for Public Policy, September 2016, available at: http://ai100.stanford.edu/2016-report; Kate Darling, *Extending legal protection to social robots: The effects of anthropomorphism, empathy, and violent behavior towards robotic objects*, 2012.
[17] Alex Campolo, Madelyn Sanfilippo, Meredith Whittaker & Kate Crawford, 'AI Now 2017 Report', *AI Now*, 2017, available at: https://ainowinstitute.org/AI_Now_2017_Report.pdf, p 28.
[18] Ibid.

4.2. Personal information

The PDP Act and many other pieces of information privacy law only protect personal information. In this sense, the definition of what constitutes 'personal information' acts as a gatekeeper to the legal protections offered to individuals. The definition of personal information can vary between jurisdictions and evolves alongside legal and societal norms. New technologies can also change the scope of personal information as new forms of information are created. For example, fitness trackers create information about individuals that did not previously exist, but which could now be considered to be personal information.

In general, the concept of personal information relies on the idea of identifiability – whether or not a person's identity can be reasonably ascertained from that information. However, the distinction between what is and is not considered to be 'personal' is being challenged by the increasing ability to link and match data to individuals, even where previously thought to be 'de-identified' or non-identifying to begin with. In this sense, a combination of seemingly non-personal information can become personal information when analysed or correlated. As the amount of available data increases, and technologies for processing and combining it improve, it becomes increasingly difficult to assess whether a given piece of data is 'identifiable'; considering a piece of data in isolation is not compatible with AI technology, and is no longer a true reflection of whether it can be deemed 'personal information'.

Much of the value of AI is its ability to identify patterns unseen to the human eye, learn, and make predictions about individuals and groups. In this sense, AI can create information that is otherwise difficult to collect or does not already exist. This means information being collected and used may extend beyond what was originally knowingly disclosed by an individual. Part of the promise of predictive technologies is that deductions can be made from other (seemingly unrelated and innocuous) pieces of data. For example, an AI system designed to make the recruitment process more efficient may be able to infer an applicant's political persuasion from other information they have supplied, and then incorporate it into the decision-making process.

Inferring information in this way not only challenges what is considered personal information, but also raises questions about whether it is acceptable to infer personal information about an individual who has chosen not to disclose it. Other questions such as who owns that information, and if it is subject to information privacy principles – including the requirement to inform that individual that information has been collected about them by means of inference – are also raised.

The current binary notion of personal information is already being challenged by mainstream technologies, yet AI blurs the distinction to the point where what is and is not 'personal information' is becoming considerably more difficult to define. The increased emergence of AI is likely to lead to an environment in which all information that is generated by or related to an individual is identifiable. In this situation, determining what is or is not protected by privacy law according to the definition of personal information is not likely to be technically or legally practical, nor particularly helpful as an effective way to protect the privacy of individuals. Many argue that there is a need to shift focus away from the binary understanding of personal information in order for privacy law to continue to protect the information privacy of individuals in an AI environment.

4.3. Collection, purpose and use

Three longstanding pillars of information privacy stemming from the OECD Guidelines are:

- **Collection limitation:** collection of personal information should be limited to only what is necessary; personal information should only be collected by lawful and fair means; and where appropriate, should be collected with the knowledge and/or consent of the individual.

- **Purpose specification:** the purpose of collecting personal information should be specified to the individual at the time of collection.
- **Use limitation:** personal information should only be used or disclosed for the purpose for which it was collected, unless there is consent or legal authority to do otherwise.

The underlying goal of these intertwined principles is to minimise the amount of information any one organisation holds about an individual, and to ensure that the way the information is handled is consistent with the expectations of that individual. AI fundamentally challenges all three of these principles.

4.4. Collection limitation

The very nature of many AI techniques, particularly machine learning, rely on ingesting massive amounts of data in order to train and test algorithms. Collecting such large amounts of data can assist the development of AI, but it can also directly oppose the collection limitation principle. Technological developments in IoT devices, smartphones and web tracking means that the data being fed into AI systems is often not collected in a traditional transaction whereby people consciously provide their personal information to someone who is asking for it.[19] In fact, many individuals are often not fully aware of the amount of information being collected about them from their devices and subsequently being used as input data for AI systems. This creates a level of conflict as limiting the collection of personal information is incompatible with the functionality of AI technologies and the devices that collect data to support it, but collecting such vast amounts of information creates inherent privacy risks.

4.5. Purpose specification

Providing an explanation of the purpose of collection (generally through a collection notice) is how most organisations adhere to the purpose specification principle. The ability of AI to extract meaning from data beyond what it was initially collected for presents a significant challenge to this principle. In some cases, organisations may not necessarily know ahead of time how the information will be used by AI in the future. There is a risk of excessive data collection beyond what is necessary 'just in case', using overly broad collection notices and privacy policies in an attempt to 'catch-all'. This kind of practice allows organisations to claim technical compliance with their privacy obligations, but it is disingenuous and inconsistent with the underlying goal of the collection limitation principle. Further, it undermines the ability of individuals to exercise meaningful control over their personal information.

Conversely, AI could be leveraged to enhance the ability of individuals to specify their preferences for how their personal information is used. For instance, it is not unreasonable to imagine services that are able to learn their users' privacy preferences and apply different conditions to the data that is collected about different individuals. In this way AI could be pivotal in the establishment of individualised, preference-based models that have the potential to meet the transparency, consent and reasonable expectations objectives of information privacy law, even more effectively than the current model of notice and consent.

4.6. Use limitation

Once collected, the use limitation principle endeavours to ensure personal information is only used for the purpose for which it was collected. In general, organisations are also permitted to use personal information for a secondary purpose that would be 'reasonably expected' by the individual. This raises the question of whether information being used as input data for an AI system can be considered a 'reasonably expected secondary purpose', given that in many instances, the outcome of doing so would be unknown to the individual. Just as AI can highlight patterns and relationships in data unforeseen by humans, it could also

[19] Information Accountability Foundation, *Artificial Intelligence, Ethics and Enhanced Data Stewardship*, 20 September 2017, p 6.

reveal new potential uses for that information. Combining this with the issues of purpose specification above, organisations are likely to find it difficult to ensure personal information is only used for the purpose it was collected for when using AI technologies.

The assumption that people, particularly young people or 'digital natives', are becoming less concerned about their information privacy may prompt the idea that a reasonably expected secondary purpose for use of information would be quite broad. This is not necessarily the case. The Boston Consulting Group found that for 75% of consumers in most countries, privacy of personal information remains a top issue, and that people aged 18-24 are only slightly less cautious than older generations.[20] This indicates that people are not by default becoming less concerned about how their personal information is being used just because technology is becoming ubiquitous, and therefore may not always regard the use of their personal information by AI as a reasonably expected secondary purpose.[21] AI is likely to blur the distinction between what is considered a primary and secondary purpose to the extent that the practicality of the use limitation principle may need to be reconsidered.

Taken together, the purpose specification, collection limitation and use limitation principles are significantly challenged by AI. Mass data collection, often by means that are not obvious to individuals; vague or misleading collection notices; and an assumption that people are more comfortable with the secondary use of their information than they actually are, lead to a situation in which the current understanding of information privacy through these principles may no longer be effective. However, AI also brings with it opportunities to revolutionise the way traditional privacy principles are realised. For instance, training a machine learning algorithm on massive amounts of data in a secure environment before being released could in turn allow for increased data security.

Widespread use of AI will prompt us to change the way we apply traditional privacy principles – whether this is an improvement or a degradation on the standards of privacy protection however, remains to be seen. By considering privacy as a foundational element within an ethical framework for developing AI, there is potential for organisations to improve collection notice practices and enable individuals to have a more nuanced and informed interaction with organisations regarding the use – and secondary use – of their information.

4.7. Transparency and consent

Our current understanding of information privacy rests on the ability of individuals to exercise choices regarding the information others have about them and what is done with it. Yet, the complexity surrounding AI can mean that processes are unclear to individuals' whose information is being used, making truly informed and meaningful consent unattainable. For instance, deep learning techniques can pose challenges to transparency, as providing an explanation about how conclusions are drawn can at times be difficult even for those initially developing the algorithms, let alone the average individual. Organisations will struggle to be transparent in their AI practices, or to obtain consent, if they cannot communicate the processes to citizens.

There is much research upon the emergence of a 'privacy paradox', in which people express concern for their privacy, but in practice continue to willingly contribute their information via the systems and technologies they use.[22] One interpretation of this paradox indicates that even when informed, individuals

[20] John Rose, Christine Barton, & Rob Souza, 'The Trust Advantage: How to Win with Big Data', *Boston Consulting Group*, November 2013, available at: https://www.bcg.com/publications/2013/marketing-sales-trust-advantage-win-with-big-data.aspx.
[21] For instance, a Pew Research Center survey of 1,002 adult users conducted in 2013 found that 86% had taken steps online to remove or mask their digital footprints, and 68% believed that current laws were not good enough in protecting online privacy. See *Anonymity, privacy, and security online*, Pew Research Centre, 2013.
[22] Patricia A. Norberg, Daniel. R. Horne & David A. Horne, 'The privacy paradox: Personal information disclosure intentions versus behaviors', *Journal of Consumer Affairs*, Vol. 41, No.1, 2007, pp 100–126; Bettina Berendt, Oliver Gunther & Sarah Spiekermann 'Privacy in e-commerce: Stated preferences vs. actual behavior', *Communications of the ACM*, Vol. 48, No. 4, 2005, pp 101–106.

often have no choice but to enter an 'unconscionable contract' to allow their data to be used.[23] In this sense, many may feel resigned to the use of their data because they feel there is no alternative, rather than positively welcoming it.[24] An increasing complexity of the networks and systems we use, combined with the widening variety of data collection methods renders a binary yes/no response to consent at the beginning of a transaction less and less meaningful in the modern world.[25] While AI technologies encourage many of these challenges, they also have the potential to be the solution, by presenting novel ways of explaining what is happening to an individual's data within each layer of processing, or enabling individualised platforms for people to exercise consent.

One potential way to increase transparency and also scrutinise, challenge and restrain decision making that has occurred without human involvement is being explored in the 'right to explanation'. Such a right would provide individuals with the ability to question decisions that affect them, which have been made on a purely algorithmic basis.[26] Despite the current technological challenge to enable this, many key figures in the AI community see transparency of decisions, or 'explainability', as integral to developing and maintaining trust in the evolving relationship between humans and intelligent machines.[27]

There is much work already being done to build algorithms that can explain how and why they came to produce their output.[28] With this kind of ability, AI could potentially facilitate transparency, in that it would be able to clearly explain decisions and be tested for bias – a process that is not always achievable for human decision makers. From a legal and policy perspective, this right is being explored in Article 22 of the European Union General Data Protection Regulation. It remains to be seen how effective this will be, with some critics arguing that there remain "serious practical and conceptual flaws," as the right only applies to decisions that are solely automated, which is rarely the case.[29]

4.8. Discrimination

Information privacy is generally regarded as an enabling right, meaning that a lot of its value lies in its ability to enable other human rights to be realised, such as the rights to freedom of association and freedom of expression. Privacy protections can also assist in the protection against discrimination by placing controls on how information about a person can be collected, used and disclosed. For example, information regarding an individual's ethnic origin or sexual orientation has stronger protections under privacy law. This is due to the inherent sensitive nature of the information, and aims to minimise the risk of harm that can be caused by making decisions based upon it. One of the most prominent ethical issues of AI with immediate ramifications is its potential to discriminate, perpetuate biases, and exacerbate existing inequalities. Because algorithms are trained on existing data, they can end up replicating unwanted patterns of unfairness due to the data they have ingested.[30]

Further, those building the systems may unknowingly introduce their own human biases into the functionality. Because AI challenges the ability of information privacy to operate as it has done historically, the safeguard against discrimination that information privacy provides as an enabling right risks becoming dismantled. Interestingly, AI technology also has the potential to minimise discrimination if developed with

[23] Sylvia Peacock, 'How web tracking changes user agency in the age of Big Data; the used user', *Big data and society*, Vol. 1, No. 2, 2014, available at: http://m.bds.sagepub.com/content/1/2/2053951714564228.
[24] ICO, *Big data, artificial intelligence, machine learning and data protection,* 2017, p 24.
[25] Ibid., p 30.
[26] Toby Walsh, *It's Alive! Artificial Intelligence from the logic piano to killer robots*, Latrobe University Press, 2017, pp 150-151.
[27] For example, such as Ruslan Salakhutdinov (Director of AI research at Apple and Associate Professor at Carnegie Mellon University) in Will Knight, 'The Dark Secret at the Heart of AI', *MIT Technology Review,* 11 April 2017, available at https://www.technologyreview.com/s/604087/the-dark-secret-at-the-heart-of-ai/.
[28] For example, see the Privacy Preservation work done by Data61 and CSIRO at https://www.data61.csiro.au/en/Our-Work/Safety-and-Security/Privacy-Preservation.
[29] Lilian Edwards & Michael Veale, 'Enslaving the Algorithm: From a 'Right to an Explanation' to a 'Right to Better Decisions'?', *IEEE Security & Privacy*, 2017, p 5.
[30] Ibid., p 2.

consideration of these issues – by removing or supporting the human element of many decision-making processes, innate human biases can be avoided.

5. Accountability and governance

Governance and oversight are championed in information privacy law to ensure appropriate structures are in place that prevent a power imbalance between citizens and government. This relies on regulators ensuring that personal information is being handled appropriately. The challenges to our understanding of information privacy outlined in the sections above are replicated when it comes to effectively regulating AI technology.

The difficulty of regulating technology has been discussed elsewhere in depth,[31] however some considerations with particular relevance to AI and information privacy include:

- AI technology is not confined to one state or jurisdiction, making it difficult to create and maintain good privacy practices and governance across borders.
- Determining who owns the data, where it is stored and who has responsibility for it is a complex task for regulators.
- Good governance needs to be based on an understanding of the technology. As AI continues to develop rapidly, the long-established gap between the law and technology is widening while the complexity and wide-reaching application of AI continues to grow.
- The extent to which government should regulate AI, noting that the absence of a regulatory framework for AI in relation to information privacy is a regulatory decision in itself.

Good governance frameworks can be used to promote good design, structure and oversight of AI technologies and how they interact with privacy. By creating an environment in which general rights and protections are enshrined, regulation can prompt the development of automated systems that are underpinned by information privacy, consistent with a Privacy by Design approach to privacy protection.

Privacy governance cannot be achieved solely through top-down control from regulators; those who control the data, and those building the technology, should themselves be involved in the design of privacy-enhancing systems.[32]

6. Conclusion

We already live in a world of big data, and the expansion of computational power through AI stands to drastically alter the landscape of information privacy. A connected life through IoT devices and smart cities technology – fuelled by AI – promises a wealth of potential benefits, including more dynamic use of resources, increased efficiency and a higher standard of living. The possibilities that AI technology could provide in healthcare, the justice system and government services are immense. Yet, as many technologies before it, AI presents social, technological and legal challenges to how we understand and protect information privacy.

This paper has stepped through some of the key information privacy considerations of AI, and how AI will require our established understanding of personal information to be revisited. However, while the long-held principles of information privacy may need to be reconceptualised, the emergence of AI does not mean that privacy will cease to matter or exist. Privacy provides an important framework for making ethical choices about how we develop, use and regulate new technologies. It will also continue to be integral to

[31] See Michael Kirby, 'The fundamental problem of regulating technology', *Indian JL & Tech*, Vol. 5, 2009.
[32] Information Accountability Foundation, *Artificial Intelligence, Ethics and Enhanced Data Stewardship*, 20 September 2017, p 15.

how we mediate our identities, develop a sense of self, and realise other important rights including freedom of speech and association. Answering the privacy questions raised by AI will be essential to its long-term success.

Moving forward, our understanding of AI and privacy may see a shift in focus from the collection aspect of information privacy, toward emphasising safeguards to ensure information is handled ethically and responsibly once it is obtained. Attempts to control or limit collection of data are likely to become increasingly difficult as data-collecting technology becomes ubiquitous. As such, shifting the emphasis toward 'ethical data stewardship' over data once it is collected has been posited as an option. This would require a genuine commitment to transparency and accountability through good governance practices.

Government has an important role to play in creating an environment in which a commitment to developing safe and fair AI can be balanced with technological progress.[33] The right balance necessitates a consultative, interdisciplinary approach, as excessive, inappropriate or misplaced regulation could slow the adoption of AI or fail to address its true challenges. Leveraging existing information privacy frameworks, as well as re-imagining traditional concepts will be a key component in building, using and regulating AI.

7. Additional reading

A list of further resources has been compiled here for further introductory reading. Please note that these resources were last accessed in May 2018.

- Alex Campolo, Madelyn Sanfilippo, Meredith Whittaker & Kate Crawford, AI Now 2017 Report, *AI Now*, 2017, available at: https://ainowinstitute.org/AI_Now_2017_Report.pdf.
- Matt Chessen, 'The AI Policy Landscape', *Medium*, March 2017, available at: https://medium.com/artificial-intelligence-policy-laws-and-ethics/the-ai-landscape-ea8a8b3c3d5d.
- Matt Chessen, 'What is Artificial Intelligence? Definitions for policy-makers and non-technical enthhusiats', *Medium*, April 2017, available at: https://medium.com/artificial-intelligence-policy-laws-and-ethics/what-is-artificial-intelligence-definitions-for-policy-makers-and-laymen-826fd3e9da3b.
- DL4J Introduction to Deep Learning and Neural Networks resources, available at: https://deeplearning4j.org/ai-machinelearning-deeplearning.
- Information Commissioner's Office, UK, *Big Data, artificial intelligence, machine learning and data protection*, 2017, available at: https://ico.org.uk/media/for-organisations/documents/2013559/big-data-ai-ml-and-data-protection.pdf.
- Will Knight, The Dark Secret at the Heart of AI, *MIT Technology Review* , 11 April 2017, available at: https://www.technologyreview.com/s/604087/the-dark-secret-at-the-heart-of-ai/.

[33] This approach is currently being explored in the European Union General Data Protection Regulation. See Article 35 in particular.