Expanding the Artificial Intelligence-Data Protection Debate

Fred H. Cate
Indiana University Maurer School of Law, fcate@indiana.edu

Christopher Kuner
Vrije Universiteit Brussel

Orla Lynskey
London School of Economics

Christopher Millard
Queen Mary University

Nora Ni Loideain
University of London

See next page for additional authors

Authors

Fred H. Cate, Christopher Kuner, Orla Lynskey, Christopher Millard, Nora Ni Loideain, and Dan Jerker B. Svantesson

Editorial

Expanding the artificial intelligence-data protection debate

Christopher Kuner,* Fred H. Cate,** Orla Lynskey,** Christopher Millard,** Nora Ni Loideain,** and Dan Jerker B. Svantesson**

Artificial intelligence (AI) has developed rapidly in recent years. From narrow applications to translate documents, filter email, and recognize faces and voices to more ambitious uses, such as, in the words of the European Commission's recent report *Artificial Intelligence for Europe*, 'helping us to solve some of the world's biggest challenges: from treating chronic diseases or reducing fatality rates in traffic accidents to fighting climate change or anticipating cybersecurity threats',[1] the capabilities of AI now and in the foreseeable future promise widespread and substantial benefits for individuals, institutions, and society. At the same time, these technological innovations raise important issues, including significant questions about the tension between AI and data protection laws.

In recent months, a great deal of ink has been spilled on AI and data protection. Some nations have issued what appear to be duelling reports, with governments focusing on how to advance AI through national and regional AI strategies and incentives, while data protection authorities address the importance of ensuring that privacy is protected in the AI context. Industry, advocacy groups, and academics have added to the debate. Most agree that AI is important and often beneficial on the one hand, but that data privacy must be protected on the other. But that is often as far as the consensus extends.

Data protection is challenged by the often rapid development and deployment of AI. At the same time, protecting data privacy is more important than ever given the speed, impact, difficulty of assessing and explaining many AI tools. This conundrum heightens the importance of expanding the focus of the debate from mere compliance with existing laws to the need to consider other approaches to enhance the quality of data protection and effective governance in the face of AI and other emerging digital tools. We wrote about this issue a year ago and we return to it now to highlight the importance of this critical subject and the growing need to expand the debate over the adequacy of existing data protection approaches to address the serious data privacy issues that AI presents.[2]

AI is in widespread use today

AI is not a new or futuristic concept. As the EC has noted: 'Artificial intelligence (AI) is already part of our lives—it is not science fiction. From using a virtual personal assistant to organise our working day, to travelling in a self-driving vehicle, to our phones suggesting songs or restaurants that we might like, AI is a reality.'[3] Or in the words of the UK House of Lords in its recent AI report, 'AI is a tool which is already deeply embedded in our lives'.[4]

This is an important point. When we talk about the AI context, we aren't referring to something hypothetical or futuristic. If there is confusion about how to apply existing data protection laws and tools to AI—or whether they apply at all—the impact is already being felt. We are literally building the boat while already sailing in it.

AI has an insatiable appetite for data

Most AI tools use substantial amounts of data. With few exceptions, more data is better than less, and there is

* Editor-in-Chief.

** Editors.

[1] Communication from the Commission, Artificial Intelligence for Europe, COM (2018) 237 final, <http://ec.europa.eu/newsroom/dae/document.cfm?doc_id=51625>.

[2] Christopher Kuner and others, 'Machine Learning with Personal Data: Is Data Protection Smart enough to Meet the Challenge?' (2017) 7, (1) 1–2, <https://academic.oup.com/idpl/article/7/1/1/3782694>.

[3] Communication from the Commission (n 1).

[4] House of Lords Select Committee in Artificial Intelligence, AI in the UK: Ready, Willing and Able?, HL Paper 100 (2018), <https://publications.parliament.uk/pa/ld201719/ldselect/ldai/100/100.pdf>.

almost never enough. As Professor Viktor Mayer-Schönberger recently noted in 'Foreign Affairs', even large companies are in need of more data to develop and deploy AI, as 'the quality of [AI applications] would deteriorate absent sufficient data, leading to inefficient transactions and reduced consumer welfare'.[5]

Data is necessary not only for AI to achieve its full potential and to prevent monopolization of critical AI, but also to guard against bias or error. If we don't have the underlying data, it is far more difficult to detect or remediate discriminatory outcomes. Moreover, large, multinational data sets are essential for AI to serve underserved segments of the population.

AI's need for personal, even sensitive data is widely recognized. As the Norwegian Data Protection Authority explained: 'Most applications of artificial intelligence require huge volumes of data in order to learn and make intelligent decisions.'[6] In fact, rather than sample data, AI often works by, in the words of the UK Information Commissioner, 'collecting and analysing *all* of the data that is available'.[7]

We in the data protection community may wish it were otherwise but, given the extraordinary proliferation of existing AI and the promise of the technology for the future, we need to come to grips with this reality.

AI challenges traditional data protection norms

Most data protection laws reflect principles established in 1980—38 years ago—in the OECD Guidelines on the Protection of Privacy and Transborder Flows of Personal Data: collection limitation, data quality, purpose specification, use limitation, security safeguards, openness, individual participation, and accountability.[8] AI challenges many of these, not just because of its demand for data, but because of how it uses data.

Knowledge and articulation of purposes for processing is required by the purpose specification and use limitation principles, which respectively provide that personal data should be collected for specified purposes and then used only for those or other compatible purposes. The challenge is how to comply with these requirements in the context of AI when data may potentially yield unforeseen and sometimes unpredictable results, by advanced algorithms that are not always directed by or initially understood by their programmers and may increasingly be created by computers.

Implicit in the OECD Guidelines, and made explicit in the EU's General Data Protection Regulation (GDPR) and other modern data protection laws, is another widely shared principle: data minimization, ie that 'Personal data must be adequate, relevant and limited to what is necessary in relation to the purposes for which those data are processed'.[9] However, with data, in the words of the Singapore Personal Data Commission, constituting the 'basic building block of the digital economy',[10] the concept of data minimization stands in tension with developing AI technologies. It is difficult to know in advance 'what is necessary' in a world of 'surprising correlations' and computer-generated discoveries. The challenges of defining a purpose for processing and only keeping data for that purpose are exacerbated because, as the Norwegian DPA has noted, 'it is not possible to predict what the algorithm will learn', and the 'purpose may also be changed as the machine learns and develops'.[11]

AI also challenges retention limits because deleting or restricting the use of data after its original purpose has been fulfilled or upon request by an individual could strip organizations and society of the potential benefits of using that data for AI development, deployment, and oversight. Data is essential if these models are to perform optimally. Yet, keeping data for longer periods or indefinitely may violate current data protection laws.

The openness and individual participation principles require that data processing be transparent and that individuals are informed about uses of their personal data. Providing transparency in the context of AI is not easy. As Professor Paul Ohm has stressed, when a program 'thrives on surprising correlations and produces

[5] V Mayer-Schönberger, and T Range, 'A Big Choice for Big Tech: Share Data or Suffer the Consequences' *Foreign Affairs* (September/October 2018) 52.

[6] Artificial Intelligence and Privacy, Datatilsynet (Norwegian Data Protection Authority) at page 4 (January 2018), <https://www.datatilsynet.no/globalassets/global/english/ai-and-privacy.pdf>.

[7] Big Data, Artificial Intelligence, Machine Learning and Data Protection, UK Information Commissioner's Office, p 11 (Version 2.2 - 2017) (emphasis added, <https://ico.org.uk/media/for-organisations/documents/2013559/big-data-ai-ml-and-data-protection.pdf>.

[8] The Guidelines were revised in 2013. See OECD Revised Guidelines on the Protection of Privacy and Transborder Flows of Personal Data (2013), <http://oecd.org/sti/ieconomy/oecd_privacy_framework.pdf>.

[9] GDPR, recital 39; art 5(1)(c).

[10] Singapore Personal Data Protection Commission, 'Discussion Paper on Artificial Intelligence (AI) and Personal Data—Fostering Responsible Development and Adoption of AI', (5 June 2018) p 2, <https://www.pdpc.gov.sg/-/media/Files/PDPC/PDF-Files/Resource-for-Organisation/AI/Discussion-Paper-on-AI-and-PD—050618.pdf>.

[11] Artificial Intelligence and Privacy (n 6) at 18.

inferences and predictions that defy human understanding . . . . [h]ow can you provide notice about the unpredictable and unexplainable?'[12]

Many regulators, policymakers, businesses, attorneys, and academics are working hard to find ways to address the challenges presented by AI to data protection laws. These are important initiatives and obviously necessary in light of the urgent need for users of data to comply with existing data protection laws. However, as we have seen, the tension between those laws and AI is so fundamental that any effort to reconcile them runs the risk of substantially weakening data protection or substantially interfering with the benefits of AI or both. None of these results is desirable given the importance of AI and of personal privacy.

Over the past decade there has been considerable attention given to how data protection law might be modernized to work better not only in the face of AI, but the growth of big data, blockchain,[13] the Internet of Things, social media, and other phenomenon that were not anticipated when the OECD Guidelines were originally published in 1980. The advent of AI may well require rethinking of fundamental data protection principles, not just because they pose an unnecessary burden to the use of AI tools, but because they do too little to protect privacy in this critical field.

In the light of AI developments, more attention may also need to be given to under-developed data protection tools, such as risk management, accountability, data review boards, and the importance of speedy redress, as well as tools not yet even imagined. The GDPR is focusing new attention on some of these; the challenge will be applying them in the face of rapidly changing technologies, including AI.

The role of humans and the protection of humanity

One aspect of re-examining traditional data protection principles and thinking about new data protection mechanisms is considering the role of individuals in overseeing technology. Notwithstanding their roots in fundamental rights, many data protection frameworks have approached the role of individuals in a very transactional way—we are given notice and sometimes choices regarding data collection, we can obtain access, we can bring complaints, and throughout the process, consent is a way around many other substantive privacy obligations. If that is unworkable in the face of AI (and big data and widely distributed sensors and other technologies), then what is the role of individuals?

As the speed, accuracy, and impact of AI increases, the role of human oversight likely will need to change as well. What is the most effective role for human intervention in the face of increasingly autonomous and advanced AI? Moreover, human decision-making is sometimes unexplainable or irrational. AI, if developed and used appropriately, offers the potential for decision-making that is not only speedier and more accurate than that of humans, but also less biased and more rational. As we wrote in 2017:

> [W]hile considerable attention has been given to the dangers of embedding unfairness in algorithmic decision-making processes, it should not be forgotten that human decision-making is often influenced by bias, both conscious and unconscious, and even by metabolism. Indeed, while it may be extremely difficult to ensure complete transparency in automated decision-making processes, even well-intentioned human decision makers are susceptible to prejudices of which even they are unaware. This suggests the intriguing possibility that it may in future be feasible to use an algorithmic process to demonstrate the lawfulness, fairness, and transparency of a decision made by either a human or a machine to a greater extent than is possible via any human review of the decision in question.[14]

Whether or not humans understand the details of how AI works, we can assure that it is developed according to legal and ethical principles. Humans are essential to evaluating its results and providing redress in the case of incorrect or unfair decisions. We need to confront frankly and openly questions about 'fairness', a term we use a great deal but rarely define. What does it mean for AI to be fair?

A key component of this question is what factors should data protection experts, whether within companies or regulatory agencies, consider when evaluating fairness. For example, some AI is likely to eliminate some jobs. Some will cause significant shifts in industries (for example, reducing individual car ownership or undermining the need or traditional public transport). Is that a component of 'fairness' in a data protection analysis? If data protection officials do not consider these and similar impacts, who will?

[12] Paul Ohm, 'Changing the Rules: General Principles for Data Use and Analysis' (2014) Privacy, Big Data, and the Public Good: Frameworks for Engagement, p 100.

[13] Christopher Kuner and others, 'Blockchain versus Data Protection', *International Data Privacy Law* (2018) 8 (2), pp 103–04.

[14] 'Machine learning with personal data: is data protection smart enough to meet the challenge?', *International Data Privacy Law* supra at 2 (citations omitted).

The EC wrote in 2018 that 'Like the steam engine or electricity in the past, AI is transforming our world, our society and our industry.'[15] We need to consider to what extent that transformation should extend to data protection law itself and the tools that give it meaning. Otherwise, we run the risk of leaving data privacy inadequately protected—or the benefits of AI inadequately developed—in our rapidly transforming world.

doi:10.1093/idpl/ipy024

[15] Communication from the Commission (n 1) at 237.