Editorial

Machine learning with personal data: is data protection law smart enough to meet the challenge?

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

Almost seven decades after Alan Turing conceived of 'intelligent machines', there has recently been a surge of interest in machine learning and algorithmic decision-making. The popular imagination has been stirred by high-profile events such as the victory of IBM's supercomputer, Watson, in the US quiz show Jeopardy, and Google Deepmind's deep learning program AlphaGo's victory in the ancient Chinese game Go. Meanwhile, machine learning processes are being deployed in contexts as varied as fraud prevention, medical diagnostics, and the development of autonomous vehicles. The underlying technologies are increasingly accessible to data controllers, with major cloud computing providers including Amazon, IBM, Google, and Microsoft offering low-cost, scalable, cloud-supported machine learning services and tools, with a particular focus on data mining and other types of predictive analytics.

Regulation of 'automated individual decisions' is not new to data protection law and was addressed explicitly in the 1995 Data Protection Directive (DPD).[1] The 2016 General Data Protection Regulation (GDPR) extends the protection against decisions made solely on the basis of automated processing to cover not only profiling of data subjects but also any other form of automated processing.[2] All of the data protection principles apply to such processing, but perhaps most significant are the requirements of the first principle, which stipulates that processing of personal data must be lawful, fair, and transparent. Although that may appear straightforward, the practical application to machine learning of each element of this principle is likely to be challenging.

Article 22(1) of the GDPR gives data subjects the right not to be subject to decision-making, including profiling, based solely on automated decision-making that produces legal effects concerning them or similarly affecting them. Personal data used for automated decisions, including profiling, should only be collected for specified, explicit, and legitimate purposes, and subsequent processing that is incompatible with those purposes is not permitted. Machine learning is data driven, typically involving both existing data sets and live data streams in complex training and deployment workflows.[3] It may be difficult to reconcile such dynamic processes with purposes that are specified narrowly in advance.

In terms of lawfulness, Article 22(2) of the GDPR does contain some specific exemptions from the prohibition on automated decision-making, including contractual necessity and consent. In those cases, however, Article 22(3) provides that the data controller 'shall implement suitable measures to safeguard the data subject's rights and freedoms and legitimate interests, at least the right to obtain human intervention on the part of the controller, to express his or her point of view and to contest the decision'. Again, this may look simple, but in practice how can informed consent be obtained in relation to a process that may be inherently non-transparent (a 'black box')? Even if an algorithmic process can in theory be explained, what if it is impossible to do that in a way that is intelligible to a data subject? To be sufficiently 'specific', will a separate consent be required for each situation in which personal data are to

* Editor-in-Chief.

** Managing Editor.

*** Editor.

[1] Directive 95/46/EC, art 15 and Recital 41. For a helpful analysis of these provisions, see Lee Bygrave, 'Automated Profiling, Minding the Machine: Article 15 of the EC Data Protection Directive and Automated Profiling', (2001) 17(1) Computer Law & Security Review17.

[2] Regulation (EU) 2016/679, art 22 and Recital 71. art 22 GDPR appears to be broader in scope than art 15 DPD because the GDPR covers 'a decision based solely on automated processing, including profiling' whereas the DPD covers only 'a decision . . . which is based solely on automated processing of data intended to evaluate certain personal aspects relating to him, such as his performance at work, creditworthiness, reliability, conduct, etc'.

[3] Singh and others, 'Responsibility and Machine Learning: Part of a Process', paper delivered at the MCCRC Symposium on Machine Learning: Technology Law & Policy <https://queenmaryuniversit907-public.sharepoint.com/Pages/Symposium-2016.aspx>

be processed for automated decision-making, for example, in particular employment, financial, or medical contexts? As for 'human intervention', it may not be feasible for a human to conduct a meaningful review of a process that may have involved third-party data and algorithms (which may contain trade secrets), prelearned models, or inherently opaque machine learning techniques.

In terms of fairness, bias may be introduced into machine learning processes at various stages, including algorithm design and selection of training data, which may embed existing prejudices into automated decision-making processes. For example, underrepresentation of a minority group in historic data may reinforce discrimination against that group in future hiring processes or credit-scoring. Profiling based on postal codes or even magazine subscriptions may become a proxy for selection based on race or gender.[4] Identifying and controlling for such biases is a critical challenge in designing and evaluating the fairness of machine learning processes.

As regards transparency, GDPR Articles 13(2)(f) and 14(2)(g) oblige data controllers to inform data subjects (at the time of data collection) regarding 'the existence of automated decision-making' and to provide 'meaningful information about the logic involved, as well as the significance and the envisaged consequences of such processing...'. Again, it is difficult to see how the second part of this requirement can be satisfied, especially in cases where a machine learning process involves multiple data sources, dynamic development, and elements that are opaque, whether for technological or proprietary reasons. Presumably, what will constitute 'meaningful information' about 'logic' must be evaluated from the perspective of the data subject. Disclosure of the full code of algorithms and detailed technical descriptions of machine learning processes are unlikely to help. A high-level, non-technical, description of the decision-making process is more likely to be meaningful. There may also be a tension between the right to generic information about a decision-making process, and the apparently more specific right 'to obtain an explanation of the decision reached ... and to challenge the decision' (GDPR, Recital 71). Although not directly binding, this Recital may embolden regulators and courts to try to compel data controllers to provide explanations of specific outcomes in particular cases, and not merely 'meaningful information' about 'logic' in general.

So, is data protection law, and in particular the GDPR, up to the challenge of regulating machine learning with personal data? Some commentators foresee a bleak, indeed almost dystopian, future, in which the growing use of algorithms increases inequality and threatens democracy.[5] Others present a more nuanced outlook in which automated decision-making, while not without significant risk, may be made subject to accountability and governance mechanisms that will facilitate outcomes in which anticipated benefits outweigh potential harms. For example, technical tools might be developed which can be applied to automated decision-making processes to audit and verify compliance with data protection and other legal requirements.[6]

Finally, while 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.[7] 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. In that event, the current data protection requirement that automated decisions should be subject to an appeal to a human may need to be reversed. A right to appeal to a machine against a decision made by a human may in the end prove to be the more effective remedy.[8]

doi:10.1093/idpl/ipx003

[4] Joshua Kroll and others, 'Accountable Algorithms' (2017) University of Pennsylvania Law Review (forthcoming). <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2765268##>

[5] See, for example, Cathy O'Neil, *Weapons of Math Destruction* (Allen Lane 2016).

[6] Kroll (n) 4.

[7] One widely reported study of judges' behaviour at a parole board in Israel revealed that it was much more likely for a parole application to be granted in the early morning or after lunch than in the middle of the day when the judges were hungry. 'Extraneous Factors in Judicial Decisions' (2011) 108(17) Proceedings of the National Academy of Sciences USA 6889.

[8] See Dimitra Kamarinou and others, 'Machine Learning with Personal Data', Queen Mary University of London Legal Studies Research Paper 247/2016<https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2865811>