Machine Learning with Personal Data:

Profiling, Decisions and the EU General Data Protection Regulation

Dimitra Kamarinou, Christopher Millard
Centre for Commercial Law Studies
Queen Mary University of London
<d.kamarinou, c.millard>@qmul.ac.uk

Jatinder Singh
Computer Laboratory
University of Cambridge
jatinder.singh@cl.cam.ac.uk

Abstract

As machine learning promises to revolutionise a number of sectors, for example by predicting users' preferences in online marketplaces and uncovering hidden links in health data, we should not forget that machine learning models are based on data and such data may relate to people. This paper gives a brief overview of some of the considerations and challenges regarding using machine learning models to process individuals' personal data under the EU General Data Protection Regulation. Specifically, we consider the impact of using machine learning in carrying out profiling and automated decision-making, we look at the responsibilities of those who control personal data, and at the rights data subjects have.

1 Introduction

Machine learning shows much promise in enabling new insights and efficiencies in a wide variety of areas, by assisting and providing the means for automating decision-making processes. However, machine learning is driven by data; as we aspire to reap the benefits of machine learning we should bear in mind the data available for analysis may be personal data subject to specific regulations. In the European Union, processing of **personal data**, i.e. *"any information relating to an identified or identifiable natural person*" (Data Protection Directive [4]) is subject to fundamental data protection principles, including that personal data should be processed in a lawful, fair and transparent way. In this paper, we consider questions regarding data protection in the context of using machine learning for profiling, and automated decision-making based on such profiles.

1.1 The EU General Data Protection Regulation

Following years of deliberations, in May 2016 the European Parliament and the Council of Ministers agreed on the final text of the *General Data Protection Regulation* (GDPR) [7]. The GDPR comes into direct effect in EU Member States in the spring of 2018. It represents a modernisation of EU data protection rules, replacing the *Data Protection Directive* (DPD) of 1995 [4]. The GDPR is generally considered expansive in terms of the rights of **data subjects** (the identified or identifiable natural people to whom the personal information relates to), and in terms of the obligations and penalties imposed on **data controllers**, that is those who control the **processing** of personal data (with processing defined very broadly to include any 'operation' on personal data from initial collection by any means through to final

destruction).

The GDPR gives data protection authorities stronger powers, including the power to issue substantial penalties to those who infringe data subject rights; in certain cases up to the value of Euro 20M or 4% of global turnover, whichever is higher. In the face of potential penalties of this magnitude, there may well be some reluctance to adopt technologies which have the complexity of machine learning with personal data.

1.2 Data processing

Under the GDPR, a key requirement is that personal data processing should be based on one of several specific lawful grounds, such as the data subject's consent to processing for a particular purpose or where processing is necessary for the controller to comply with a legal obligation. Additional requirements may exist for special categories of personal data, such as data revealing racial or ethnic origin, or genetic data, etc.

Processing must also follow a number of fundamental principles, including that data should be collected for specified, explicit and legitimate purposes and not be further processed in a way that is incompatible to those purposes (purpose limitation principle) and that processing should be limited to what is necessary for the specific purposes specified by the data controllers (data minimization principle).

In machine learning, where data (from existing datasets or collected real-time) are used to train the algorithm, and because some machine learning techniques and applications are better suited to large datasets, attention should be paid to avoid processing unnecessary personal data. Compliance with the data minimization principle may limit the application of such techniques and may result in a less representative picture of data subjects.

2 Automated decision-making, including profiling

In practice, automated decision-making will often entail profiling, where the profiles guide the decision-making process. This is explicitly recognised in the GDPR, where profiling is defined as a sub-category of automated processing, and refers to the use of personal data to evaluate certain personal aspects of natural people to analyse and predict certain aspects of their life.

It has been argued that one of the underlying principles of the DPD is that '*fully automated assessments of a person's character should not form the sole basis of decisions that significantly impinge upon the person's interests'* (Bygrave [2]). This principle appears to continue to be reflected in Article 22 of the GDPR, which also covers profiling of people based on their health, location and movement, as data subjects have the right not to be subject to decision-making based solely on automated processing, where it significantly affects them in some way (see §2.2). This is significant in relation to machine learning, given that proponents of the technology emphasise its ability to automate and facilitate decision making processes.

2.1 Profiles

Individuals' personal data may be processed not only to create descriptive profiles about them but also to '*check [their profiles] against predefined patterns of normal behaviour'* (Coudert [3]) so as to determine whether they fit or deviate from them. The building of such profiles will be subject to the GDPR rules governing the processing of personal data.

2.1.1 Individuals or groups?

Article 22 seems only to apply to profiling of individual data subjects and not groups. That is, a profile might be based on some pre-existing group, e.g. the students in a class, or an

'assumed' group such as individuals deemed to have a specific credit risk profile based merely on their residence within a particular postcode. In such cases it could be argued that the protection against such decisions under Article 22 of the GDPR would be applicable, as the provision does not limit 'profiling' as such to individual profiling, but only requires that the *decision* based on such profiling is addressed to an individual.

2.1.2 Anonymization?

Article 22 of the GDPR only applies to data subjects and it could be argued that the protection against solely automated decision-making might not apply if the data processed are anonymized (Savin [8]), even if such decisions may impact an individual. However, using anonymized data, whether alone or combined with other data, to 'single out' individuals, or to infer information about them in order to take decisions affecting them (see §2.2 below), could be incompatible with the purpose for which the data were originally collected (§1.2). For example, personal data collected through tracking technologies (such as cookies) may be aggregated, anonymized and combined with other data (personal or not), using analytics to make predictions about an individual's interests or behaviour.

2.2 Decisions

Decisions refer to the *use* of the profile, i.e. the determinations and conclusions about data subjects based on such profiles. Under the DPD it has been argued that a 'decision' has to be interpreted broadly and GDPR Recital 71 clearly states that a 'decision' may include a 'measure'. The concept of 'measure' could include, for example, '*the targeted marketing of specific medical products against cancer based on the search made by an individual on the internet'* (EU Commission [5]).

Importantly, a decision has to produce *legal effects or similarly significantly affect the data subject,* for example an automatic rejection of an online credit application (GDPR, Recital 71). The effects can be both material and / or non-material (non-economic), potentially affecting the data subject's dignity, integrity or reputation. This may have an impact on the ways in which machine learning processes can be deployed.

3 Derogations: permitted automated decision-making

Notwithstanding the rule just described, Article 22(2) of the GDPR does permit automated decision-making in certain circumstances. These are when the decision

(a) is necessary for entering into, or performance of, a contract between the data subject and a data controller;

(b) is authorised by Union or Member State law to which the controller is subject (...);

(c) is based on the data subject's explicit consent.

In all three cases, suitable measures must be in place to safeguard data subjects' rights. Moreover, the first exception will be interpreted narrowly as only covering processing that is *necessary* for a particular contractual arrangement and the second will only apply where there is a specific legal framework covering the processing in question. In the case of the third exemption, for consent to be specific and informed, data subjects will require a meaningful explanation about how their data is processed. Depending on the degree of explanation required, it is clear this can raise a number of considerations in a machine learning context, where it can be difficult to explain the inner-workings of machine learning processes (see §4.1).

4 Rights and obligations

Under the GDPR (as under the DPD) data subjects have various individual rights, and also some protections that derive from obligations imposed on data controllers. The following are likely to be of particular significance in relation to machine learning.

4.1 Transparency

When personal data are collected from the data subject, the GDPR imposes on the data controller the obligation, at the time when personal data are obtained, to provide the data subject with information regarding the existence of automated decision-making and *'meaningful information about the logic involved'*. Complying with this obligation may be problematic in a machine learning context. Does it mean that whenever machine learning is used for profiling the data controller must provide information regarding the existence and type of machine learning algorithms used or the role of different service providers forming part of the 'machine learning' supply chain? Indeed, various components of the machine learning supply chain, including algorithms and pre-learned models, could involve one or more third parties, e.g. through cloud and other 'as a service' offerings (e.g. machine learning as a service). Moreover, does the term 'logic' refer to the data set used to train the algorithm, the way the algorithm itself works in general, for example the mathematical / statistical models on which an algorithm is based, or to the way the learned model worked in the particular instance when processing the data subject's personal data? There is a wide range of different machine learning algorithms, some are comparatively opaque in terms of their function and design, while others are more amenable to allowing humans to track how they work. It follows that the inner-workings of the various components may not be visible to data controllers, let alone regulators (see below). An added level of opacity may derive from the confidential nature of the underlying code, which could be protected by intellectual property rights.

Moreover, the *meaningfulness* aspect of the provision should be assessed from a data subject's perspective. So it remains to be seen whether it is important for controllers to know how those algorithms and models have been designed, whether their initial training data set was based on personal or anonymized data, and the sources of such data - and how much of these inner-workings need to be explained to data subjects. It may be that although such detailed information about the technical workings of a machine learning process may not be helpful to data subjects, data controllers might be required to disclose it to regulators in the context of an audit or investigation. Indeed, Bygrave has argued that the logic should: '*be documented and (…) the documentation be kept readily available for consultation and communication…*' (Bygrave [2]). Even if it is difficult to keep documentation, it may be feasible to describe (albeit in broad terms) the way in which the system was constructed, the data selected, the algorithms trained/tested, and the outputs evaluated.

Note also, that Article 13(2)(f) of the GDPR requires that information on the decision-making process should be provided at the time that data subjects' personal data are obtained. Machine learning can be a highly dynamic process, where different algorithms and approaches may be tried, and therefore it may be difficult for data controllers to predict and explain at the time personal data are collected the precise nature of the algorithms employed.

4.2 Intervention rights

When decisions based on profiling are permitted, under the GDPR (Article 22(2) (a) and (c)) data subjects have the right (a) to obtain human intervention on the part of the controller, (b) to express their point of view with regards to a decision made about them and (c) to contest the decision that has significantly affected them.

There are a number of difficulties with this in a machine learning context. It is unclear as to which stage human involvement is required: in practice (at least at present!) people are involved in the design, training, and testing of a system incorporating machine learning. Also, it might be inappropriate for certain approaches to be used to take measures / decisions when a human is absent, e.g. machine providing medical diagnosis without a doctor reviewing the facts.

As regards the data subjects' right to express their point of view, there should be a possibility to do this prior to a decision being made (or a measure being taken). It follows that in a machine learning context the data controller should implement appropriate measures to

prevent any machine learning driven-process from making a final decision before the data subject is consulted. This may have an impact on system design, and pose challenges in situations where decisions are taken in response to data in real time.

If a data subject exercises the right to contest the decision, a human should be tasked with reviewing the decision. Having said that, it is not clear who this 'human' should be and whether he / she will be able to review a process that may have been based on third party algorithms, pre-learned models or data sets including other individuals' personal data or on opaque machine learning models.

4.2.1 Appeal to a machine?

A number of studies have shown that human decision-making is often based on established prejudices and stereotypes (Lowry & MacPherson [6], Wood et al. [9]). Given the ongoing work on issues of bias, fairness and transparency in machine learning, there is the potential that machine learning algorithms can be made to disregard discriminatory factors more effectively than humans. If that is true, an interesting question arises as to whether individuals could appeal to a machine against a decision made by humans, instead of the current norm of appealing to a human against a decision made by a machine? There may also be scope for machine-learned models that drive decisions subsequently to be reviewed and tested by other algorithms for audit purposes.

4.3 Deletion and rectification

According to GDPR Recital 72, it appears that creating profiles is also subject to the requirement that there be a legal ground for processing and the obligation to comply with the data protection principles. In relation to the DPD, the Article 29 Working Party (EU data protection regulators acting collectively) advised in 2013 that '*data subjects should also have the right to access, to modify or to delete the profile information attributed to them'* [1]. If this is correct, then, as a prerequisite to exercising such rights, data subjects have the right to know what profiles have been created about them.

The exercise by individuals of their rights to rectify inaccurate or incomplete personal data, or erase personal data, could have complex 'knock-on' impacts on machine learning processes. Consider the situation where an individual may become aware that his or her personal data has been incorporated into a machine learned model. The individual may then decide to exercise the right to request erasure of, or correction to some or all of that data. That may in turn have an impact on the legal basis for continuing to use the model to the extent that it still incorporates the personal data in question. In particular, might a data controller then be obliged either to stop using the model or to go back and retrain the model either without including the data that have been removed or using only the modified version of the data?

4.4 Data Protection Impact Assessments

Where a type of processing is likely to result in a high risk for the rights and freedoms of data subjects, data controllers are required to carry out a Data Protection Impact Assessment (DPIA). A DPIA is required particularly in cases of *'a systematic and extensive evaluation of personal aspects relating to natural persons which is based on automated processing, including profiling'* (GDPR Article 35(3)(a)). Under the GDPR, the DPIA must cover, among other things, the security measures aimed at ensuring the protection of personal data and compliance with the Regulation.

The GDPR refers, in particular, to cases where processing of data subjects' personal data for profiling purposes may give rise to discrimination on the basis of any of the special categories of data (e.g. racial or ethnic origin, religion, health status, etc.) or to measures having such a discriminatory effect. It imposes on the data controller the obligation to use appropriate mathematical and statistical procedures for the profiling to ensure that factors which result in inaccuracies in personal data are corrected and the risk of errors is minimised

(GDPR Recital 71). Unfair discrimination in a machine learning environment may be caused by a direct or indirect bias introduced in the profiling process due to deficiencies in the quality and quantity of the data available to train and test the algorithm, as well as problems with data sources and labelling. Therefore, data controllers should implement appropriate measures to mitigate the risk of algorithms working on incomplete or unrepresentative data which may generate spurious correlations that result in unjustifiable decisions.

Even though not explicitly mentioned in this provision, the 'security measures' mentioned here could require data controllers to implement the principles of data protection by design and by default (under GDPR Article 25) both at the time of the determination of the means of processing (for example, when deciding to use machine learning algorithms to process personal data) and at the time of processing itself. So, in a machine learning context for example, in order to comply with the data minimization principle discussed above, data controllers may have to decide, at the time of collection, which personal data they are going to process for profiling purposes. Then, they will also have to provide the algorithm with only the data that are strictly necessary for the specific profiling purpose, even if that leads to a narrower representation of the data subject and possibly a less fair decision for him/her.

Again, machine learning systems may be composite in nature, having been designed by a party other than the data controller and with input data derived from a range of separate data providers, and machine learning processes may run in a cloud environment that may itself involve multiple service providers. Therefore, the data controller may struggle to implement the appropriate technical and organisational measures required by the GDPR to comply with the data protection principles.

5 Conclusion

EU data protection law assumes that automated decision-making processes are risky and that individuals need to be protected from such processes. Specific protections include the right to be informed about automated decision-making, including profiling, as well as rights to have a human review a machine decision. With advances in machine learning research, an interesting possibility is that machines may surpass certain limitations of human decision makers and provide us with decisions that are demonstrably fair.

Acknowledgments

We acknowledge the financial support of Microsoft through the Microsoft Cloud Computing Research Centre (MCCRC). We thank members of the MCCRC team for the useful comments and feedback. Responsibility for views expressed remains with the authors.

References

[1] Article 29 Working Party. (2013) Advice paper on essential elements of a definition and a provision on profiling within the EU General Data Protection Regulation. http://bit.ly/2fIei8K (accessed 3 November 2016).

[2] Bygrave, L. (2001) Automated Profiling, Minding the Machine: Article 15 of the EC Data Protection Directive and Automated Profiling. *Computer Law & Security Review* 17(1):17-24.

[3] Coudert, F. (2010) When video cameras watch and screen: Privacy implications of pattern recognition technologies. *Computer Law and Security Review* 26:377-384.

[4] Directive 95/46/EC of the European Parliament and of the Council on the protection of individuals with regard to the processing of personal data and on the free movement of such data (Data Protection Directive), OJL 181/31 (23 November 1995).

[5] EU Commission. (2014) The EU data protection Regulation: Promoting technological innovation and safeguarding citizens' rights. SPEECH 14/175 http://europa.eu/rapid/press-release_SPEECH-14-175_en.htm?locale=en (accessed 3 November 2016).

[6] Lowry, S. and McPherson, G. (1988) A blot on the profession. *British Medical Journal* 296 (6623):657-658.

[7] Regulation (EU) 2016/679 of the European Parliament and of the Council on the protection of natural persons with regard to the processing of personal data and on the free movement of such data,

and repealing Directive 95/46/EC (General Data Protection Regulation) (27 April 2016) OJ L119/1 (4 May 2016).

[8] Savin, A. (2014) Profiling and Automated Decision Making in the Present and New EU Data Protection Frameworks. (paper presented at 7th International Conference Computers, Privacy & Data Protection, Brussels, Belgium): 1-14.

[9] Wood, M., Hales, J. et al. (2009) A test for racial discrimination in recruitment practice in British cities. *Department for Work and Pensions Research Report No 607*, pp. i-69.