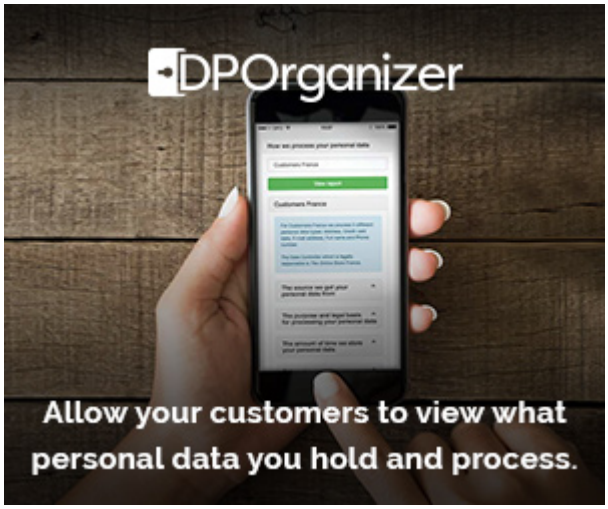
Is there a "right to explanation" for machine learning in the GDPR?

Jun 1, 2017 Save This ()



(http://bit.ly/2s47Qxm)

Andrew Burt

Much has been made about the coming effects of the GDPR — from how organizations collect data to how they use that data and more. But as machine learning gains a more prominent role across organizations, a key question is emerging, and it's baffling lawyers, scholars and regulators alike: How does the GDPR affect machine learning in the enterprise?

As in other areas, the GDPR is less than clear. And as a result, the idea that the GDPR mandates a "right to explanation" from machine learning models — meaning that those significantly affected by such models are due an accounting of how the model made a particular decision — has become a controversial subject. Some scholars, for example, have spoken out vehemently (https://poseidon01.ssrn.com/delivery.php?ID=332092117088003102097108120088097121026011068081017086005005105078092114078092101108063055059052058025 against the mere possibility that such a right exists. Others, such as the UK's own Information Commissioner's Office, seem to think the right is pretty clearly self-evident (https://ico.org.uk/for-organisations/data-protection-reform/overview-of-the-gdpr/individuals-rights/rights-related-to-automated-decision-making-and-profiling/).

Ultimately, I have some good news for lawyers and privacy professionals . . . and some potentially bad news for data scientists.

Starting With the Text

a data subject's right to access data), and Articles 21 and 22, which falls under Section 4 (dealing specifically with the data subject's right to object to and opt out of automated decision-making).

So let's start with Articles 13–15.

Each article covers a separate aspect of a data subject's right to understand how her or his data is being used. Importantly, each article also contains identical language stating that, in the case of automated decision-making, the data subject possesses the right to access "meaningful information about the logic involved, as well as the significance and the envisaged consequences of such processing for the data subject."

Two points about this language are worth making right off the bat.

First, the fact that the same language is repeated multiple times, in one of the most important sections of the GDPR, is not for nothing. It's safe to assume that the drafters wanted it to be clear that this provision is, as the legendary, but fictional, Ron Burgundy might say, kind of a big deal.

Second, it's also important to note that this provision makes a critical distinction. On the one hand, it mandates that "meaningful information about the logic" of automated systems be made available. On the other hand, the provision also distinguishes "the significance and the envisaged consequences of such processing" as something apart from the logic itself.

So what does "meaningful information" and the "significance" of processing mean for data subjects in practice? Articles 21 and 22 give us some perspective.

Article 21 outlines a data subject's right to object to the processing of her or his data, and Article 22 specifically spells out that right with regards to automated processing, stating that "the data subject shall have the right not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her."

> In plain English, the text suggests that a data subject is entitled to enough information about the automated system that she or he could make an informed decision to opt out.

Taken together, Articles 21 and 22 suggest that the right to understand "meaningful information" about and the "significance" of, automated processing is related to an individual's ability to opt out of such processing. In plain English, the text suggests that a data subject is entitled to enough information about the automated system that she or he could make an informed decision to opt out.

Looking to the Recitals for Context

The Recitals — where the drafters tried to explain the regulation in nonbinding language — shed some light onto the text we've just examined. Recital 71 explains that automated processing "should be subject to suitable safeguards, which should include specific information to the data subject and the right to obtain human intervention, to express his or her point of view, to obtain an explanation of the decision reached after such assessment and to challenge the decision."

This is perhaps the clearest the GDPR gets on the right to explanation. Indeed, it's hard to get more straightforward than to state data subjects have a right to "obtain an explanation of the decision reached after such assessment," as Recital 71 does.

So why leave this language for the nonbinding Recitals, and not include it in the text of the regulation itself?

that the only reason this language was left to the Recitals is because the drafters didn't mean for a full right to explanation to exist in the binding text. But I believe that interpretation takes things too far. In my view, there's a difference between meaningful information about the logic of a decision and the significance of that decision, as required by Articles 13–15, and the explanation of an individual decision itself, described by Recital 71 — but that difference is not as great as it seems.

In some cases, explaining exactly how an individual decision was made is incredibly difficult in machine learning systems, meaning that the safeguards suggested in Recital 71 are less of a distinction in kind and more of a distinction in degree. That is, while the binding provisions of the GDPR state that individuals are entitled to meaningful information about the logic and significance of machine learning systems, the drafters also suggest that this right should include the right to understand each individual decision in the Recitals. But they don't mandate that level of transparency in the regulation.

What is mandatory is enough insight into the logic of a model and the significance of that logic so that a data subject would have the context necessary to intelligently opt-out.

So What Does This Mean in Practice?

Let's start with the fact that high-quality machine learning is difficult. Incredibly difficult. In fact, and the only thing harder than training good models is explaining them. MIT Technology Review's Will Knight does a good job of describing the problem (https://www.technologyreview.com/s/604087/the-dark-secret-at-the-heart-of-ai/) as it frequently occurs in practice: "A network's reasoning is embedded in the behavior of thousands of simulated neurons, arranged into dozens or even hundreds of intricately interconnected layers. ... These outputs are fed, in a complex web, to the neurons in the next layer, and so on, until an overall output is produced."

It's because of this complexity that understanding exactly why, and how, a machine learning model has made a particular decision is so challenging.

> All of which is to say that if you're a privacy professional, you're going to find it difficult to implement these requirements in practice once the GDPR comes into effect.

All of which is to say that if you're a privacy professional, you're going to find it difficult to implement these requirements in practice once the GDPR comes into effect. To that end, below are a few pointers designed to help privacy professionals enable the deployment of machine learning systems while complying with the GDPR:

- **Start by getting technical:** You'll want to get a technical description of the model and the data it was trained on, which you can work off to build your data subject-friendly explanation. Is the model using, for example, a neural net, a support vector machine, or logistic regression? Different models tend to contain different levels of opacity, and model choice can help determine the best ways to explain the underlying logic. You'll also want to get a basic

did the data come from, for example? How many features does the model select for? Discuss all of this with your technical experts.

- **Understand deployment:** Next you'll want to develop a thorough understanding of how the model will be used in practice. What decision is it going to be used to make? What are the consequences of a false positive? Or of a false negative? This will help you explain more than just the logic of the decision, but also the significance of the model, as required by Articles 13-15.

- **Educate the data subject:** Finally, you'll want to put these answers together with the general aim of educating a data subject who might seek to opt out of this type of decision. What might they want to understand if they'd like to opt out of the model making the decision? You might even want to make a list of the major points required to intelligently opt out — what are the benefits of allowing the automated-processing, for example, versus the downsides of opting out — and make sure that the information you're presenting about the model addresses as many of these points as possible.

With these answers in hand, you should be able to construct a basic explanation of how the model is working, which can help you to put together a general explanation that satisfies the "logic" and "significance" requirements set forth in Articles 13–15, and place them in the context of the "opt-out" requirements in Articles 21–22.

Sound difficult? Surely. The road to GDPR compliance won't be straightforward, and it won't be easy. But with the GDPR's enforcement date less than a year away, we've still got plenty of time to think through the most difficult challenges it poses, and to ensure we're ready for May 2018.

Author

Andrew Burt

Share This

Tags

EU (/tag/eu) Big Data (/tag/big-data) Privacy Law (/tag/privacy-law)

Privacy Opinion (/tag/privacy-opinion)

Pease International Tradeport, 75 Rochester Ave, Suite 4
Portsmouth, NH 03801 USA • +1 603.427.9200