GDPR and Explainable AI

By Janet Wagner

The EU General Data Protection Regulation (GDPR) went into effect on May 25, 2018. The regulation has had a significant impact on how companies around the world handle the data of European citizens. The regulation includes numerous provisions regarding data protection and privacy. Several sections of the GDPR have led to a debate among AI industry professionals about the "right to explanation" mandate included in the regulation.

GDPR Explainability Clauses

GDPR Articles 13-15 and 21-22 outline requirements related to automated data processing and decision making. The basic concept is that when a decision is generated solely from automated processing (no human intervention), including profiling, the data subject has the right to receive an explanation of how the decision was rendered. This clause applies when a company is using automated processing on personal data to evaluate an individual (who resides in the European Union) based on the individual's attributes.

Automated data processing and decision systems typically use machine learning, a subset of AI. The intent of the "right to explanation" clauses in GDPR when it comes to AI algorithms and models are a subject of debate among AI industry professionals.

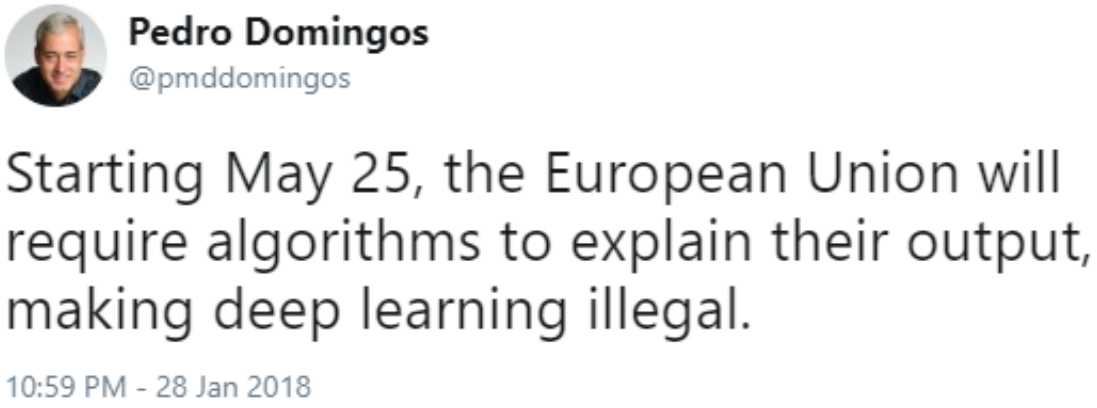
The Explainability Debate

When it comes to AI, "explanation" could mean several things: 1) How an algorithm works or how the system functions. 2) The factors or data that resulted in a decision by the algorithm or system that impacted an individual (a data subject). AI industry professionals disagree about whether "explanation" in the context of GDPR is referring to how the technologies work or the factors that led to the automated decision.

Dr. Sandra Wachter, research fellow at the Oxford Internet Institute, University of Oxford, has written about GDPR and AI. In a blog post, she said that "the GDPR is likely to only grant individuals information about the existence of automated decision-making and about "system functionality," but no explanation about the rationale of a decision." For example, a bank could use automated data processing for online credit card applications. If an applicant is denied approval of a credit card, it is likely that the bank would not be required to provide an explanation as to the rationale of that automated decision under GDPR.

Andrew Burt, chief privacy officer and legal engineer at Immuta, explains in an article for IAPP what GDPR "in practice" means for the AI community. In the article, he said that the GDPR 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."

A few months before GDPR went into effect, Pedro Domingos, professor of computer science at UW and author of "The Master Algorithm," published a controversial tweet that started a heated debate among the AI community:



Domingos received quite a bit of pushback on the idea that GDPR will make deep learning illegal.

GDPR and AI

The GDPR explanation requirements may not be cut and dry when it comes to AI. But there are projects that aim to produce explainable AI such as the DARPA Explainable AI (XAI) program and Local Interpretable Model-agnostic Explanations (LIME).

It may take legal cases to determine the correct interpretation of the explainability clauses in GDPR as it pertains to AI.