Explaining responsibility

de Groot, Aviva; Bayamlioğlu, Emre; van Schendel, Sascha; Nyrup, Rune; Veale, Michael; Binns, Reuben; van Otterlo, Martijn

Publication date:
2019

Document Version
Publisher's PDF, also known as Version of record

Link to publication in Tilburg University Research Portal

Citation for published version (APA):
de Groot, A., Bayamlioğlu, E., van Schendel, S., Nyrup, R., Veale, M., Binns, R., & van Otterlo, M. (2019). *Explaining responsibility*. Abstract from TILTing perspectives conference, Tilburg, Netherlands.

EXPLAINING RESPONSIBLY

Panel proposal for TILTING Perspectives 2019

Proposal for track "AI, Robotics and Responsibility" (Track leader: Merel Noorman/ M.E.Noorman@uvt.nl)

The interdisciplinary scholarly discourse on 'the right to explanation' of AI-infused decision making processes goes beyond the GDPR's sphere of application, as it addresses understandability needs that are recognized on a global scale. Processes of analyzing, profiling, and predicting human behaviour support decisions in all sectors of society, from credit scoring to fraud detection to decisions on who to hire - or even arrest. The increasing complexity of the technologies used, industry intricacies, and network effects all add to inscrutability and assessment challenges of these applications, on individual level as well as societal levels. The decreasing awareness of ubiquitous automation processes in the background of people's lives raises additional concerns. Increasingly, it is noted that issues of obscurity cannot be 'explained away,' or explained at all on an individual level. While most agree there is a pressing need to make these systems safe, fair, and 'democratically understandable,' there seems to be, at least temporarily, some competition between those that argue for scrutability at higher levels and the ones researching individual explanatory potential.

In the meantime, in theory and practice, different approaches and methodologies towards 'explainable AI 2.0' are being designed and tested. The GDPR functions as a catalyst as controllers already need to comply with requirements for explainability. Explanations should be understandable and meaningful. The latter term precisely triggers the above mentioned competition, as it is far from self-evident what a ‘meaningful’ explanation is. What counts as an honest, time-stamped translation of a complex and dynamic computational process? Who gets to decide what that is? Can explanations be misused to obfuscate abuse of power?

In the absence of commonly understood and accepted evaluative standards it is hard to assess the beneficence, usefulness and pitfalls of these developing explanatory methodologies. This conundrum might inform us to stop talking about 'responsible explanations' and instead speak of 'explaining responsibly.' As a field of research, it needs to be interdisciplinary. Law, philosophy, data science, cognitive sciences, STS and humanities each have valuable theory and experience to bring to the table.

This panel provides such a table, and aims to start the discussion in acknowledgment of the seemingly irreconcilable, acute needs for both individual explanations and high level governance strategies.

Confirmed panelists: Reuben Binns, Michael Veale, Martijn van Otterlo, tentative: Rune Nyrop. The panel will be presented, chaired and the discussion hosted by Emre Bayamlıoğlu, Aviva de Groot, and Sascha van Schendel. A 'test-case' will be designed by Aviva, Emre and Sascha, and discussed by the panelists.

Contact: Aviva de Groot, Aviva.deGroot@uvt.nl