The Poisoned Chalice: The Legal Right to Explanation of AI Systems

Hofit Wasserman-Rozen

 The increasing use of artificial intelligence, either as an alternative to human decision-making or as a decision-support tool, raises concerns about transparency, accountability, and fairness. In response, regulators worldwide are promoting the "right to explanation," intended to provide the subject of a decision with an explanation for an autonomous machine decision that affects them. To implement this right, regulators are turning to various technological practices for creating "explanations" as defined by machine learning experts, collectively known as eXplainability or XAI.

This article critically examines the suitability of XAI practices for implementing the regulatory "right to explanation." It analyzes the concept of "explanation," part of the notion called "reason-giving," in law. By reviewing its instances in public, private, and international law, it extracts the core purposes underlying this legal concept. The analysis reveals that the substantive purposes of legal reason-giving include better decision quality, respect for human autonomy, delivering on the right to due process, and strengthening the authority of the decision-making system.

The article then examines XAI, exploring its development motivations and various practices. Comparing XAI to the underlying purposes of legal explanation reveals challenges and potential risks in implementing the right to explanation in its current state. The article concludes by calling for a purposive interpretation of this right to guide regulators and developers in creating explanations that serve the "right to an explanation" and promoting institutions that will make this right more than just a dead letter in AI system regulation.