

Legal Significance of Explainable AI and Its Practice

*Sieh-Chuen Huang**

Abstract

This article attempts to clarify whether or which aspects of the “explainable AI”, a research hotspot in the data science community, can meet the “explainability” or “right to explanation” required by the legal domain. First, by analyzing recent research in the data science field regarding “explainable AI”, the two connotations of “explainability” are found. One is the interpretation brought out by the researchers after understanding (interpretability). And the second is transparency, which is achieved by using methods such as decomposition to show “explanation producing system”. Next, this article turns eyes to discussions related to “explanation” in legal domain. The word “explanation” is often used when regulations and judicial decisions require information related to algorithms. But it is more often seen that, instead of “explanation”, adjacent concepts such as information access, disclosure, due process, etc. are used. However, there is still considerable debate on whether regulations such as GDPR can derive the “right to explanation” and what its connotation is. After comparing the idea of “explanation” in both data science and law, this paper argues that, when a higher level of explanation is required (for example, when reviewing public sector decisions), exogenous approaches such as surrogate models developed by the data scientists do not satisfy “meaningful information” defined by law and hence are not legally qualified explanations. The information provided by AI producers should at least include an overview of the training data, the type of model, the most important

* Professor, College of Law, National Taiwan University.

E-mail: schhuang@ntu.edu.tw

factors, and the effectiveness of the model. The above information consisting of “production system of interpretation” may comply with the “meaningful information” of Article 15 of the GDPR. On the other hand, the weight of each factor or the source code is not included in the information that should be legally disclosed. Finally, with regard to the judicial AI that may appear in the future, this article takes the relevant research on legal analytics as an example to illustrate the relationship between the processing and explainability, so as to benefit users such as judges and lawyers to properly exercise the “right to explanation”.

Keywords: Explainability/ Interpretability, Right to Explanation, Model-Centric Interpretation, Subject-Centric Interpretation, Legal Analytics, Global Interpretability, Local Interpretability