Legal Perspectives for Explainable Artificial Intelligence in Medicine - Quo Vadis?

Cătălin-Mihai PESECAN* and Lăcrămioara STOICU-TIVADAR

Department of Automation and Applied Informatics, University Politehnica Timişoara, Vasile Pârvan Blvd, no. 2, 300223 Timişoara, Romania.
E-mails: catalin.pesecan@student.upt.ro; lacramioara.stoicu-tivadar@upt.ro;

* Author to whom correspondence should be addressed;

Abstract
Explainable Artificial Intelligence (XAI) can offer an insight into the inner workings of AI models. The new EU Artificial Intelligence Act that came into force in August 2024 and will be fully applicable in August 2026, classifies the AI used in medical domain as "high-risk". For high-risk applications the requirements are "to ensure … operation is sufficiently transparent to enable deployers to interpret a system's output and use it appropriately. An appropriate type and degree of transparency shall be ensured with a view to achieving compliance with the relevant obligations of the provider and deployer". In this work we present how XAI methods can help in explaining medical AI models. We present a mapping for 3 types of models (for tabular data classification, for image data classification and for diagnostic prognosis data). In order to understand for example images, we can deploy techniques like Grad-CAM. For tabular data we can use both LIME or Grad-CAM. The first method generates a new dataset consisting of perturbed samples and offers local approximations. Grad-CAM will generate heatmaps based on the gradient from the last layer (because it contains the most information) of a convolutional neural network. Explainable Artificial Intelligence methods come in multiple flavors and options and can offer different perspectives. Multiple XAI methods can offer a broader perspective for the models used in the medical area. It is also very important to make sure that the medical experts trust and understand the explanations, so the evaluation of each method before integrating it with the medical experts can help them to accept the models.

Keywords: Artificial Intelligence (AI); Explainable Artificial Intelligence (XAI).