

Hajo Greif

Politechnika Warszawska

The Meanings of Transparency in Artificial Intelligence

Concerns over epistemic opacity and demands for epistemic transparency abound in contemporary debates on Artificial Intelligence (AI). However, it is far from clear to what extent these concerns and demands refer to the same set of problems, and how they relate to each other. This paper sketches the landscape of meanings of epistemic opacity versus transparency in AI, first, by distinguishing between technical requirements, ethical and political demands, and epistemic norms of transparency.

For the sake of responding to the former two demands, the epistemic norms of transparency need to be explicated. Second and conversely, it will have to be determined whether transparency and opacity refer to properties either of AI algorithms or of the models to which they pertain. This decision has implications as to whether and how epistemic opacity can be resolved: Whereas complex algorithms might be made more transparent for human observers, the opacity of AI models might be ‘essential’ in certain respects.