

What is "Explainable AI"?

*Satoshi Hara¹

1. Osaka University

In this talk, I will present a brief overview of "Explainable AI". The recent developments of AI technologies enable computers to perform comparably or even better than the human on some tasks such as image recognition. However, the "black-box" nature of AI is considered as a crucial shortcoming of the current AI. That is, current AI only outputs its prediction or recognition results but it does not explain why and how such outputs are obtained in a "human interpretable" manner. To alleviate the "black-box" nature, several efforts have been made to AI to be "explainable" in the last few years. I will introduce these recent developments on "Explainable AI".

Keywords: Artificial Intelligence, Explainable AI, Machine Learning