

Federica Gerace: Data-centric AI: Play with Datasets to Improve Machine Learning Performances

An emerging research line in machine learning is focusing on the role played by data in learning problems. Quite recently, the statistical physics community has started working in the direction of extending its methods to include data structure. In this talk, I will provide some insights on how this extension can contribute to analyzing several machine learning frameworks, especially those meant to mitigate the need of new labelled data. Among them, I will particularly focus on Transfer Learning, by showing how an analytically tractable model of source-target dataset pairs can lead to useful guidelines for Machine Learning practitioners.