ellis unit | TURIN Talk

Politecnico di Torino

ARTIFICIAL INTELLIGENCE HUB
POLITECNICO DI TORINO

VANDAL

http://vandal.polito.it

www.ellis.eu

Damien Teney

Idiap Research Institute, Martigny, Switzerland

July 16h, 2024
starting at 3:00 PM

Vandal Lab, Covivio - Corso Ferrucci 112
Turin, Italy

Inductive Priors in Trained and Untrained Neural Networks

While much of the current work in AI is about pushing the scale of data and models upwards, much remains to be understood about the basic building blocks of these models. In particular, our understanding of the generalization capabilities of neural networks is still incomplete. Prevailing explanations are based on implicit biases of gradient descent (GD) but they cannot account for the capabilities of models obtained with gradient-free methods, nor for the simplicity bias that appears even in untrained models. In this talk, we will uncover some inductive biases inherent to various architectural components of neural networks that explain these capabilities. We will present how our recent work analyzed trained and untrained networks to characterize the influence of various components on a model's inductive prior. This approach yields, for example, an explanation for the well-known "simplicity" bias that does not rely on the regularization effect of (S)GD. The talk will also cover ongoing work on the applications of these findings for controlling the behaviour of trained models.

Damien Teney leads the Machine Learning Group at the Idiap Research Institute in Martigny, Switzerland. He has done extensive work at the intersection of computer vision, machine learning, and natural language processing. In particular, he contributed to the development of early methods for automatic image captioning and visual question answering. His research group at Idiap now focuses on understanding and expanding the generalization capabilities of machine learning models, both small and large. He is particularly interested in the robustness of models across distribution shifts. He has previously been affiliated with the University of Adelaide, Carnegie Mellon University, the University of Bath, the University of Innsbruck, and the University of Liege in Belgium, where he is originally from.