ID contributo: **10** Tipo: **non specificato**

The math of training large neural networks, with some analogy to physics

Recently, the theory of infinite-width neural networks led to the first technology, muTransfer, for tuning enormous neural networks that are too expensive to train more than once. For example, this allowed us to tune the 6.7 billion parameter version of GPT-3 using only 7% of its pretraining compute budget, and with some asterisks, we get a performance comparable to the original GPT-3 model with twice the parameter count. In this talk, I will explain the core insight behind this theory. In fact, this is an instance of what I call the *Optimal Scaling Thesis*, which connects infinite-size limits for general notions of "size"to the optimal design of large models in practice. I'll end with several concrete key mathematical research questions whose resolutions will have incredible impact on the future of AI.

Relatore: YANG, Greg (Microsoft Research)