# Provable Neural Scaling Laws

**Abstract:**

The training of modern large language models (LLMs) is guided by so-called neural scaling laws that inform strategic decisions about model size and dataset size within a given computational budget. These laws describe how model performance changes with respect to both model size and sample size and have been empirically observed to follow power-law relationships. However, the origins of these scaling laws remain unclear: are they fundamental principles of neural networks?

In this talk, I will present a solvable setting (associative memory models) where neural scaling laws naturally emerge. Importantly, the exponents of these power laws are determined analytically by the long-tailed nature of the data distribution and the optimization procedures employed during training. This is the first work which rigorously establishes neural scaling laws, in a regime which is relevant to the understanding of LLMs.