

**Title:**

Transformers, codes and locality notions

**Abstract:**

We consider two independent questions: (i) when can Transformers learn efficiently? (ii) when can Reed-Muller codes communicate reliably? In particular, it was conjectured from the 1960s that the answer to the second question is at any rate below Shannon capacity, while it is believed by some that the answer to the first question is also quite comprehensive in favor of Transformers. In this talk, we present a proof of the Reed-Muller code conjecture, as well as a conjecture on the characterization of Transformers learning capabilities. In both cases, the crux of the arguments relies on proper notions of "distribution locality". The talk mostly focuses on the Transformer case. Based on joint works with S. Bengio (Apple), A. Lotfi (EPFL), C. Sandon (EPFL), O. Saremi (Apple).