



ELLIS talk

What functions do self-attention blocks prefer to represent?

Surbhi Goel (Microsoft)

Host: Marco Mondelli

Self-attention, an architectural motif designed to model long-range interactions in sequential data, has driven numerous recent breakthroughs in natural language processing and beyond. In this talk, we will focus on studying the inductive bias of self-attention blocks by rigorously establishing which functions and long-range dependencies they statistically represent. Our main result shows that bounded-norm Transformer layers can represent sparse functions of the input sequence, with sample complexity scaling only logarithmically with the context length. Furthermore, we propose new experimental protocols to support this analysis, built around the large body of work on provably learning sparse Boolean functions.

Based on joint work with Benjamin L. Edelman, Sham Kakade and Cyril Zhang.

Monday, December 13, 2021 04:00pm - 06:00pm

IST Austria Campus Zoom Link: <https://istaustria.zoom.us/j/94397239114?>



3Tg1bkpVUDc5TXIZWG1paWpUdz09 Meeting ID: 943 9723 9114 Passcode

This invitation is valid as a ticket for the IST Shuttle from and to Heiligenstadt Station. Please find a schedule of the IST Shuttle on our webpage: <https://ist.ac.at/en/campus/how-to-get-here/>. The IST Shuttle bus is marked IST Shuttle (#142) and has the Institute Logo printed on the side.