ICTS | INTERNATIONAL CENTRE *for* THEORETICAL SCIENCES

TATA INSTITUTE OF FUNDAMENTAL RESEARCH

ICTS Seminar

Title : Learning to Grok: Emergence of in-context learning and skill composition in modular arithmetic tasks

Speaker : Aritra Das (University of Maryland, College Park, USA)

Date : Thursday, 15 January 2026

Time : 11:30 AM (IST)

Abstract : Large language models can solve tasks that were not present in the training set. This capability is believed to be due to in-context learning and skill composition. In this work, we study the emergence of in-context learning and skill composition in a collection of modular arithmetic tasks. Specifically, we consider a finite collection of linear modular functions $z = ax + by \bmod p$. We empirically show that a GPT-style transformer exhibits a transition from in-distribution to out-of-distribution generalization as the number of pre-training tasks increases. We find that the smallest model capable of out-of-distribution generalization requires two transformer blocks, while for deeper models, the out-of-distribution generalization phase is transient, necessitating early stopping. Finally, we perform an interpretability study of the pre-trained models, revealing highly structured representations in both attention heads and MLPs; and discuss the learned algorithms. Notably, we find an algorithmic shift in deeper models, as we go from few to many in-context examples.

Venue : Emmy Noether Seminar Room
Zoom Link: https://icts-res-in.zoom.us/j/99808148844?pwd=0MsJE56uLBE2nDOUFGQyucMjdSjFfo.1
Meeting ID: 998 0814 8844
Passcode: 152026