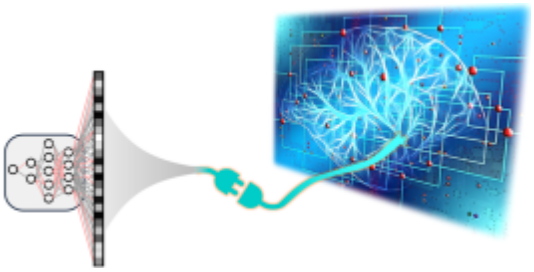
ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Distributed Computing

Prof. R. Wattenhofer

Emergent architectures: Neuro-symbolic AI



Despite the breakthrough made by Large language models (LLMs), they struggle with out-of-distribution generalizations. Particularly, the pre-trained state-of-the-art Transformer language models fail at compositional tasks such as multi-step compositional reasoning and analogical reasoning [1, 2]. On the other hand, our latest research results demonstrate the power of neuro-symbolic AI in solving these challenging tasks [3, 4]. Neuro-symbolic AI domain seeks complementary approaches that beneficially combine deep learning advancements with symbolic computations to endorse their strengths and supplement their weaknesses. In this project, one of the key objectives is to develop a post-attention neural architecture with a strong focus on neuro-symbolic AI approaches. The developed architecture will be evolved and evaluated with the aim of improving out-of-distribution generalization, length generalization, and sample efficiency for a range of synthetic and real-world applications. Other inputs or directions are welcomed.

Requirements: Strong motivation, ability to work independently, and interest in conducting exciting theoretical and/or empirical research. Solid background in mathematics and deep learning. Good programming skills in Python (and ideally with libraries such as Pytorch). Expertise in Computer Vision and/or LLMs is an advantage.

The thesis will be performed at the IBM Research-Zurich lab in Rüschlikon. If you are interested in this challenging position on an exciting new topic, please send your most recent curriculum vitae including a transcript of grades.

Contact

- Giacomo Camposampiero: gcamposampie@ethz.ch
- Dr. Abbas Rahimi: abr@zurich.ibm.com

References

[1] Nouha Dziri, Ximing Lu, Melanie Sclar, Xiang Lorraine Li, Liwei Jiang, Bill Yuchen Lin, Sean Welleck, Peter West, Chandra Bhagavatula, Ronan Le Bras, Jena D. Hwang, Soumya Sanyal, Xiang Ren, Allyson Ettinger, Zaid Harchaoui, and Yejin Choi. Faith and fate: Limits of transformers on compositionality. In *Thirty-seventh Conference on Neural Information Processing Systems*, 2023.

[2] Jonathan Thomm, Giacomo Camposampiero, Aleksandar Terzic, Hersche Michael, Bernhard Schölkopf, and Abbas Rahimi. Limits of transformer language models on algorithmic learning. In *Thirty-eight Conference on Neural Information Processing Systems*, 2024.

[3] Michael Hersche, Mustafa Zeqiri, Luca Benini, Abu Sebastian, and Abbas Rahimi. A neuro-vector-symbolic architecture for solving raven's progressive matrices. *Nature Machine Intelligence*, 5(4):363–375, 2023.

[4] Giacomo Camposampiero, Michael Hersche, Aleksandar Terzić, Roger Wattenhofer, Abu Sebastian, and Abbas Rahimi. Towards learning abductive reasoning using vsa distributed representations. In *International Conference on Neural-Symbolic Learning and Reasoning*, pages 370–385. Springer, 2024.