Mikail Khona

Personal Info

email: mikail@mit.edu

Twitter : @KhonaMikail

Current Experience

May 2023 - September 2023 **NTT Research and Harvard University**
Research Scientist Intern
Supervisor: Hidenori Tanaka
Studying reasoning and planning in transformer-based language models (LLMs) using synthetic tasks. Developing mechanistic interpretability technique to reverse engineer transformers on synthetic algorithmic tasks.

Education and Research

2018 - 2024 (expected) **Massachusetts Institute of Technology**, MA
PhD candidate in Physics
Advisor: Ila Fiete, Secondary: Mehran Kardar
Graduate research in theoretical/computational systems neuroscience and deep learning.

2014 - 2018 **Indian Institute of Technology (IIT), Bombay**, India
Bachelor of Technology in Engineering (GPA: 9.6/10)
Major: Engineering Physics (+honours in Physics)
minor: Mathematics

Publications

- Khona, Mikail *et al* **Toward a mechanistic understanding of stepwise inference in transformers: A synthetic graph navigation model**, NeurIPS 2023: R0-FoMo: Robustness of Few-shot and Zero-shot Learning in Foundation Models [ICLR 2024 link]
- Rahul Ramesh, Khona, Mikail, *et al* **How Capable Can a Transformer Become? A Study on Synthetic, Interpretable Tasks** NeurIPS 2023: R0-FoMo: Robustness of Few-shot and Zero-shot Learning in Foundation Models [ICLR 2024 link]
- Khona, Mikail*, Rylan Schaeffer*, *et al* **Self-Supervised Learning of Representations for Space Generates Multi-Modular Grid Cells** NeurIPS, 2023.
- Khona, Mikail, Fiete, Ila. **Attractor and Integrator Networks in Neuroscience**. Nature Reviews Neuroscience, 2022. *[link]*
- Ziming Liu*, Khona, Mikail*, *et al* **Growing Brains: Co-emergence of Anatomical and Functional Modularity in Recurrent Neural Networks**. *[link]*. Also at NeurIPS 2023: Unifying Representations in Neural Models Workshop
- Schaeffer, Rylan, Khona, Mikail, *et al* **No Free Lunch from Deep Learning in Neuroscience: A Case Study through Models of the Entorhinal-Hippocampal Circuit**. NeurIPS. 2022. *[link]*
- Schaeffer, Rylan, Khona, Mikail, *et al* **Reverse-engineering recurrent neural network solutions to a hierarchical inference task for mice**. NeurIPS. 2020. *[link]*
- Khona, Mikail*, Chandra, Sarthak*, Fiete, Ila. **Spontaneous emergence of topologically robust grid cell modules: A multiscale instability theory**. Submitted.*[link]*
- Duan, Sunny*, Khona, Mikail*, Bertagnoli, Adrian*, Fiete, Ila. **See and Draw: Generation of complex compositional movements from modular and geometric RNN representations**. Proceedings of Machine Learning Research. *link*
- Khona, Mikail*, Chandra, Sarthak*, *et al* **Winning the lottery with neurobiology: faster learning on**

many cognitive tasks with neural connectivity constraints. Neural Computation (2023). *[link]*

- G. Madirolas, A. Al-Asmar, L. Gaouar, L. Marie-Louise, A. Garza-Enriquez, M. Khona, C. Ratzke, J. Gore, A. Pérez-Escudero. **A taste for numbers: Caenorhabditis elegans. foraging follows a low-dimensional rule of thumb**. Nature communications biology (2023).*[link]*
- Schaeffer, Rylan*, Bordelon Blake*, Khona, Mikail*, *et al* **Efficient Online Inference for Nonparametric Latent Variable Time Series**. UAI. 2021. *[link]*
- Rylan Schaeffer, Khona, Mikail, Zachary Robertson, *et al* **Double Descent Demystified: Identifying, Interpreting Ablating the Sources of a Deep Learning Puzzle**, NeurIPS 2023 Workshop on Attributing Model Behavior at Scale. *[arXiv link]*
- Rylan Schaeffer, Berivan Isik, Victor Lecomte, Mikail Khona, Yann LeCun, Andrey Gromov, Ravid Shwartz-Ziv, Sanmi Koyejo **An Information-Theoretic Understanding of Maximum Manifold Capacity Representations**, NeurIPS 2023 workshop: Information-Theoretic Principles in Cognitive Systems [link]

Peer-Review Conference and Workshop Posters

- Rylan Schaeffer, Mikail Khona, Nika Zahedi, Ila R Fiete, Andrey Gromov, Sanmi Koyejo **Associative Memory Under the Probabilistic Lens: Improved Transformers and Dynamic Memory Creation**, Associative Memory and Hopfield Networks in 2023
- Khona, Mikail, Schaeffer, Rylan, and Fiete, Ila. **Self-Supervised Learning of Efficient Algebraic Codes generates Grid Cells,** NeurIPS Self-Supervised Learning: Theory and Practice Workshop, 2022.
- Khona, Mikail, Chandra, Sarthak, Konkle, Talia and Fiete, Ila. **Modelling the development of the primate visual cortical hierarchy**. Cosyne Abstracts 2022, Lisbon, Portugal.
- **Khona, Mikail**, Chandra, Sarthak, Acosta, Francisco, Fiete, Ila **The emergence of discrete grid cell modules from smooth gradients in the brain**. Cosyne Abstracts 2021.
- Khona, Mikail, Xu, Qianli and Fiete, Ila. **A model of oscillatory gating of information flow between neural circuits as a function of local recurrence**. Cosyne Abstracts 2020.
- Schaeffer, Rylan, Khona, Mikail, and Fiete, Ila. **No Free Lunch from Deep Learning in Neuroscience: A Case Study through Models of the Entorhinal-Hippocampal Circuit,** ICML AI4Science Workshop. 2022.

Publications in prep/to appear

- Mikail Khona*, Sarthak Chandra*, Talia Konkle, Ila Fiete. **Self-organized emergence of modularity, hierarchy, and topography from competitive synaptic growth in a developmental model of the visual pathway**

Awards / Achievements

2022 - 2023 K. Lisa Yang ICoN Graduate Student Fellow ($100k)

2021 - 2022 MathWorks Science Fellowship (one of 20 across the School of Science at MIT) ($100k)

2018 - 2019 Seigel Fellowship, Department of Physics ($100k)

2016 - 2018 Institute Academic Award for the highest GPA among undergraduates in the Physics department at IIT Bombay (9.95/10)

2016 - 2017 DAAD-WISE scholarship for an undergraduate project in Germany in 2017 [declined].

2014 An All India Rank of 562/1.4M (Percentile 99.96) in the **IIT - JEE** 2014.

2014 INSPIRE Scholarship for Higher Education - A scholarship awarded by the Government of India to meritorious students in high school who plan to pursue a degree in the natural sciences.

Relevant courses

IIT-Bombay Mathematics and Statistics: Real analysis, Complex analysis, Differential equations, General Topology, Abstract Algebra, Lie groups and Lie Algebras, Stochastic processes. Physics: Statistical physics, Advanced statistical physics, Quantum mechanics sequence (I,II,III).

MIT Mathematics: Probability Theory, Computational neuroscience. Physics: Statistical physics for biology, Systems Biology

Technical Skills

advanced Deep learning with Python: Pytorch

advanced Scientific computing with Python (NumPy, SciPy, NetworkX, etc..) and MATLAB

Academic Services

general Reviewer for Physical Review letters, iScience, Cell Reports

2023 Reviewer for NeurReps: Symmetry and Geometry in Neural Representations, Workshop, NeurIPS 2022, Reviewer for NeurIPS 2022 Workshop: InfoCog, Reviewer for NeurIPS AI4Science workshop, Reviewer for Associative Memory and Hopfield Networks

2022 Reviewer for NeurReps: Symmetry and Geometry in Neural Representations, Workshop, NeurIPS 2022, Reviewer for NeurIPS 2022 Workshop: Self-Supervised Learning - Theory and Practice, Reviewer for NeurIPS AI4Science workshop.

Teaching and Mentoring

Fall 2019 8.01L: Physics I

Spring 2021 8.592: Introduction to Biological Physics

Fall 2021 Physics Mentorship program, Physics Department, MIT

Fall 2023 8.03 Waves and Oscillations