ISAAC LIAO

iliao@mit.edu https://iliao2345.github.io

EDUCATION

Carnegie Mellon University

Aug 2024 - June 2029 (Expected)

PhD in Machine Learning. Advisor: TBD

Massachusetts Institute of Technology GPA: 5.0/5.0

Sep 2023 - May 2024

Master of Engineering in Electrical Engineering and Computer Science. Advisor: Max Tegmark

Massachusetts Institute of Technology GPA: 5.0/5.0

Sep 2019 - Jun 2023

Bachelor of Science, Double major in Computer Science and Physics

Select coursework: Information theory, Bayesian modeling and inference, Statistical learning theory

EXPERIENCE

Tegmark AI Safety Group

Sep 2023 - May 2024

Graduate Researcher, Supervised by Max Tegmark

- Posed and tested hypotheses about the internal workings of neural networks, including LLMs.
- Simplified recurrent neural networks into standard forms using symmetry transformations.

MIT 8.01 Classical Mechanics I

Sep 2023 - December 2023

Teaching Assistant, Supervised by Peter Dourmashkin

• Collected RAG data for LLM used to generate physics problems to teach \sim 700 students.

Beneficial AI Foundation

Jul 2023 - Aug 2023

Research Consultant, Supervised by Max Tegmark

• Spearheaded the below publication: Generating Interpretable Networks Using Hypernetworks.

Soljačić Group

Jun 2020 - Jun 2023

Undergraduate Researcher, Supervised by Marin Soljačić

• Spearheaded the below publication: Learning to Optimize Quasi-Newton Methods.

RESEARCH PUBLICATIONS

Not All Language Model Features Are Linear.

Submitted to NeurIPS 2024.

Josh Engels, Isaac Liao, Eric J. Michaud, Wes Gurnee, and Max Tegmark.

• Discovering that LLMs represent temporal data on circular manifolds.

Opening the AI Black Box: Program Synthesis via Mechanistic Interpretability. arXiv, 2024. Eric J. Michaud, Isaac Liao, Vedang Lad, Ziming Liu, Anish Mudide, et al.

• Reducing trained RNNs into interpretable python through a series of simplifying steps.

Learning to Optimize Quasi-Newton Methods.

TMLR 2023.

Isaac Liao, Rumen Dangovski, Jakob Nicolaus Foerster, and Marin Soljačić.

• Learning an optimizer for optimizing neural networks with theoretical guarantees.

Streamlining Physics Problem Generation to Support Physics Teachers in Using Gen. AI.

Shams El-Adawy, Isaac Liao, Vedang Lad, Mohamed Abdelhafez, et al.

The Physics Teacher, 2024.

• How to use an LLM to generate physics problems suitable for teaching.

Generating Interpretable Networks Using Hypernetworks.

arXiv 2023.

Isaac Liao, Ziming Liu, and Max Tegmark.

• Designing a graph neural network to generate good weights for another neural network.

PROJECTS

Bayesian Recommendation Systems

Feb 2023 - May 2023

- Made >2% RMSE improvement on the Netflix Prize Dataset for user-product recommendation systems.
- Created a Bayesian extension of the alternating least squares algorithm for large matrix completion.

Expressive Capacity of Sparse Neural Networks

Sep 2022 - Dec 2022

• Theorems showing that sparse neural networks can take less memory to represent the same computation.

AWARDS AND HONORS

International Physics Olympiad: Silver Medal. 2nd in Canada.	July 2019
International Physics Olympiad: Honorable Mention. 5th in Canada.	July 2018
MIT Battlecode: 1st place on one-man team, \$8000 prize. Swarm intelligence competition.	Jan 2022
MIT Battlecode: 7th place on one-man team, \$1000 prize.	Jan~2021
MIT Battlecode: 1st place of newbie division on one-man team, \$500 prize.	Jan 2020