

EDUCATION

Carnegie Mellon University PhD in Machine Learning. Advisor: Albert Gu	<i>Aug 2024 - June 2029 (Expected)</i>
Massachusetts Institute of Technology GPA: 5.0/5.0 Master of Engineering in Electrical Engineering and Computer Science. Advisor: Max Tegmark	<i>Sep 2023 - May 2024</i>
Massachusetts Institute of Technology GPA: 5.0/5.0 Bachelor of Science, Double major in Computer Science and Physics Select coursework: Information theory, Bayesian modeling and inference, Statistical learning theory	<i>Sep 2019 - Jun 2023</i>

EXPERIENCE

Tegmark AI Safety Group Graduate Researcher, <i>Supervised by</i> Max Tegmark <ul style="list-style-type: none"> Posed and tested hypotheses about the internal workings of neural networks, including LLMs. Simplified recurrent neural networks into standard forms using symmetry transformations. 	<i>Sep 2023 - May 2024</i>
MIT 8.01 Classical Mechanics I Teaching Assistant, <i>Supervised by</i> Peter Dourmashkin <ul style="list-style-type: none"> Collected RAG data for LLM used to generate physics problems to teach ~700 students. 	<i>Sep 2023 - December 2023</i>
Beneficial AI Foundation Research Consultant, <i>Supervised by</i> Max Tegmark <ul style="list-style-type: none"> Spearheaded the below publication: Generating Interpretable Networks Using Hypernetworks. 	<i>Jul 2023 - Aug 2023</i>
Soljačić Group Undergraduate Researcher, <i>Supervised by</i> Marin Soljačić <ul style="list-style-type: none"> Spearheaded the below publication: Learning to Optimize Quasi-Newton Methods. 	<i>Jun 2020 - Jun 2023</i>

RESEARCH PUBLICATIONS

Not All Language Model Features Are Linear. <i>Josh Engels, Isaac Liao, Eric J. Michaud, Wes Gurnee, and Max Tegmark.</i> <ul style="list-style-type: none"> Discovering that LLMs represent temporal data on circular manifolds. 	Submitted to NeurIPS 2024.
Opening the AI Black Box: Program Synthesis via Mechanistic Interpretability. <i>Eric J. Michaud, Isaac Liao, Vedang Lad, Ziming Liu, Anish Mudide, et al.</i> <ul style="list-style-type: none"> Reducing trained RNNs into interpretable python through a series of simplifying steps. 	arXiv , 2024.
Learning to Optimize Quasi-Newton Methods. <i>Isaac Liao, Rumen Dangovski, Jakob Nicolaus Foerster, and Marin Soljačić.</i> <ul style="list-style-type: none"> Learning an optimizer for optimizing neural networks with theoretical guarantees. 	TMLR 2023.
Streamlining Physics Problem Generation to Support Physics Teachers in Using Gen. AI. <i>Shams El-Adawy, Isaac Liao, Vedang Lad, Mohamed Abdelhafez, et al.</i> <ul style="list-style-type: none"> How to use an LLM to generate physics problems suitable for teaching. 	The Physics Teacher , 2024.
Generating Interpretable Networks Using Hypernetworks. <i>Isaac Liao, Ziming Liu, and Max Tegmark.</i> <ul style="list-style-type: none"> Designing a graph neural network to generate good weights for another neural network. 	arXiv 2023.

PROJECTS

Bayesian Recommendation Systems <ul style="list-style-type: none"> Made $\geq 2\%$ 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. 	<i>Feb 2023 - May 2023</i>
Expressive Capacity of Sparse Neural Networks <ul style="list-style-type: none"> Theorems showing that sparse neural networks can take less memory to represent the same computation. 	<i>Sep 2022 - Dec 2022</i>

AWARDS AND HONORS

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