Timothy H. Kostolansky

timothy.h.kostolansky@gmail.com | tim0120.github.io

Education

Massachusetts Institute of Technology *Cambridge, MA*
Master of Engineering in Computer Science and Engineering *May 2024*
Bachelor of Science in Computer Science and Engineering *May 2023*
Bachelor of Science in Physics *May 2023*

Relevant Coursework: Machine Learning, Deep Learning, Reinforcement Learning, Statistical Inference, Natural Language Processing, Algorithms, Robotics (Manipulation, Task and Motion Planning), Quantum Physics, Relativity, Statistical Physics

Work and Research Experience

MIT CSAIL, Algorithmic Alignment Group *July 2023 – Present*
Graduate researcher for AI alignment lab led by Asst. Prof. Dylan Hadfield-Menell

- Developing and testing methods to extract a constitution which describes language models and preference datasets
- Using language modeling, clustering, textual semantic similarity, and contextual bandit methods to find a set of principles which describes a language model's behaviors in safety-relevant situations
- Paper in progress, Github to be released publicly, Master's work at https://dspace.mit.edu/handle/1721.1/156804

Supervised Program for Alignment Research (SPAR) *June – October 2024*
Mentored by Jake Mendel from Apollo Research

- Decomposing and reverse engineering neural networks that learn Boolean circuits
- Using probing, causal abstraction, and Boolean function measures (e.g., influence) in order to determine how small neural networks represent the parts of a Boolean circuit that it is trained on
- Blog/paper in progress, pending results

Second Spectrum Incorporated *June – August 2022*
Software engineer for a sports data company that uses computer vision to track athletes in game film

- Upgraded and refactored video data pipelines from professional sports streams to the company's S3 servers
- Used Temporal.io to protect from failure over long-running protocols

MIT Laser Interferometer Gravitational Wave Observatory (LIGO) *February 2021 –August 2021*

- Updated prototype designs for the Fast Shutter System (protects high-sensitivity measuring equipment)
- Use of numerical physics simulation with Mathematica and hands-on work with designing and building shutter prototype

Activities and Leadership Experience

MIT Science Policy Review *April 2021 – September 2023*
Technology director for a policy journal that publishes science policy reviews authored by members of the MIT community

- Maintaining and updating the Review's website, uploading articles and covers

MIT Varsity Basketball *September 2019 – March 2022*
NCAA Division III athlete, competed with full course load, two-time NEWMAC Academic All-Conference selection

Japanese National Basketball Team *June 2019 – August 2019*

- Selected for National Team and trained at Ajinomoto National Training Center in Tokyo
- Competed in the 2019 William Jones Cup in Taiwan, earned bronze medal

Skills

Code: Python 3, Julia, MATLAB, Mathematica, TypeScript, bash
Frameworks: PyTorch, NumPy, pandas, ROS, React, Node.js
Languages: English (native), Japanese (proficient), Italian (learner)
Interests: meditation, chess, basketball, tennis, running, ortholinear keyboards