# VEDANG LAD

vedanglad.com

### Education

Massachusetts Institute of Technology GPA 5.0/5.0 Master of Engineering in Electrical Engineering and Computer Science

### Massachusetts Institute of Technology GPA 4.8/5.0

Bachelor of Science in Electrical Engineering and Computer Science Bachelor of Science in Physics

### Publications

### **Conference** Publications

The Remarkable Robustness of LLMs: Stages of Inference? Vedang Lad, Wes Gurnee, Max Tegmark International Conference on Machine Learning (ICML): Workshop on Mechanistic Interpretability 2024

Estimating label quality and errors in semantic segmentation data via any model. Vedang Lad, Jonas Mueller International Conference on Machine Learning (ICML): Workshop on Data-centric Machine Learning 2023

### **Journal Publications**

Opening the AI black box: program synthesis via mechanistic interpretability. Eric Michaud<sup>\*</sup>, Isaac Liao<sup>\*</sup>, Vedang Lad<sup>\*</sup>, Ziming Liu<sup>\*</sup>, with Max Tegmark et al. MDPI: Entropy Journal 2024

### Streamlining Physics Problem Generation to Support Physics Teachers in Using Generative Artificial Intelligence.

Shams El-Adawy, Isaac Liao, Vedang Lad, Mohamed Abdelhafez, Peter Dourmashkin The Physics Teacher, Vol. 62, No. 7, 2024

### **Conference** Abstracts

### Physion++: Evaluating Physical Scene Understanding with Objects Consisting of Different **Physical Attributes in Humans and Machines**

Hsiao-Yu Tung, Mingyu Ding, Zhenfang Chen, Sirui Tao, Vedang Lad, Daniel Bear, Chuang Gan, Josh Tenenbaum, Daniel Yamins, Judith Fan, & Kevin Smith Proceedings of the Annual Meeting of the Cognitive Science Society 2023

### **Research** Experience

## ML Alignment & Theory Scholars (MATS)

Research Scholar

- Improved the interpretability of LLMs with gradient-based approaches under Jessica Rumbelow.
- Developed novel data-agnostic method for feature extraction and evaluation in large language models.

### **Tegmark AI Safety Group**

Research Assistant

- Studied the science of machine learning, or mechanistic interpretability, under Max Tegmark.

Cambridge, MA

May 2023 Cambridge, MA

Cambridge, MA

June 2024 – August 2024 Berkeley, CA

June 2024

September 2023 – May 2024

1

### **MIT Brain and Cognitive Sciences**

Undergraduate Researcher

- Investigated the gap in intuitive physics between humans and computer vision models under the guidance of Joshua Tenenbaum, Dan Yamins, and Judith Fan.

### MIT Kavli Institute with NASA NICER

Undergraduate Researcher

- Time-series analysis under Dheeraj Pasham studying black holes with the NASA telescope NICER.
- Implemented optimization algorithms to fit models to energy spectra to determine composition.

### **Industry Experience**

### Cleanlab May 2022 – July 2023 Machine Learning Engineer Remote - Developed and published a new ML algorithm for label error detection to improve ML data quality.

### Teaching Experience

**MIT** Teaching Assistant 8.01 Graduate Teaching Assistant Cambridge, MA Spring 2022 **MIT** Teaching Assistant 6.C01/6.C51 Lab Assistant Modeling with Machine Learning: Algorithms and Applications Cambridge, MA

### Invited Talks

### AI In The Classroom

Vedang Lad West Windsor Plainsboro Regional School District Professional Development Day (Scheduled: Feb 2025)

### Software Contributions

- Cleanlab Github Repository (9700+ stars) - Error detection algorithms for semantic segmentation

### Technical Skills

Languages: Python, Java, Julia, JavaScript, HTML/CSS, C, Assembly, Mathematica, MATLAB Developer Tools: VS Code, Jupyter, Pytorch, Tensorflow, Docker, Github, ROS, React

### Extracurricular

MIT Cross Country, Track & Field	2019 -	2024
NCAA Division III Athlete: 2x Team National Champion, 1x Team National Runner-Up		
Plainsboro Rescue Squad	2015 -	2023
EMT: NJ certified EMT volunteering over 2500+ hours to local community.		

Last Updated: December 1, 2024

May 2021 – January 2022

Cambridge, MA

Cambridge, MA

Fall 2023

### December 2021 – May 2022