Varshini Subhash

LinkedIn | Google Scholar | GitHub | Website varshinisubhash@g.harvard.edu

Education

Harvard University Cambridge, Massachusetts
M.E in Computational Science and Engineering, GPA: 3.91/4.0 *Aug 2021 - May 2023*

Massachusetts Institute of Technology Cambridge, Massachusetts
Cross-Registered Student, Courses: 6.036, 6.8610, 6.S986, 6.9041 *Aug 2021 - May 2023*

Manipal Institute of Technology Manipal, India
Bachelor of Technology in Mechanical Engineering, CGPA: 9.09/10.0 *Aug 2014 – July 2018*

Research Publications

- **Varshini Subhash***, Anna Bialas*, Weiwei Pan, Finale Doshi-Velez, "Why do universal adversarial attacks work on large language models?", *New Frontiers in Adversarial Machine Learning Workshop, ICML 2023.*
- Amber Nigam*, Jie Sun*, **Varshini Subhash**, Paolo Antonio S. Silva, MD, "Identifying the Risk of Diabetic Retinopathy Progression Using Machine Learning on Ultrawide Field Retinal Images", *International Workshop on Health Intelligence, AAAI Conference on Artificial Intelligence 2024.*
- Zixi Chen*, **Varshini Subhash***, Marton Havasi, Weiwei Pan, Finale Doshi-Velez, "What Makes a Good Explanation?: A Harmonized View of Properties of Explanations", *Trustworthy and Socially Responsible Machine Learning Workshop, NeurIPS 2022.* | [arXiv]
- **Varshini Subhash**, Karran Pandey, Vijay Natarajan, "GPU Parallel Algorithm for Computing Morse-Smale Complexes", *IEEE Transactions on Visualization and Computer Graphics | IEEE VIS Conference 2020.* [IEEE Xplore]
- Abhijath Ande, **Varshini Subhash**, Vijay Natarajan, "Tachyon: Efficient Shared Memory Parallel Computation of Extremum Graphs", *Computer Graphics Forum, 2023*
- **Varshini Subhash**, "Can Large Language Models Change User Preference Adversarially?" | [arXiv]

Research Experience

Harvard University Cambridge, Massachusetts
Student Researcher | Advisors: Dr. Weiwei Pan & Prof. Finale Doshi-Velez *February 2022 - May 2023*

- Synthesized mathematical properties needed for good explanations and quantified trade-offs between them.
- Extracting user properties from explanations deployed in human-centered (HCI) settings.

Massachusetts Institute of Technology Cambridge, Massachusetts
Student Researcher | Advisors: Dr. Weiwei Pan & Prof. Yoon Kim *September 2022 - May 2023*

- Proposed a novel geometric hypothesis explaining the effectiveness of universal adversarial attacks on large language models like GPT-2. Used dimensionality reduction and white-box analysis as supporting evidence.

Stanford Existential Risks Initiative Cambridge, Massachusetts
ML Alignment Theory Scholar | Advisors: Stuart Armstrong & Rebecca Gorman *Nov 2022 - Dec 2022*

- Demonstrated and interpreted adversarial red teaming and probing on dialogue models like GODEL & ChatGPT.

Indian Institute of Science Bangalore, India
Research Assistant | Advisor: Prof. Vijay Natarajan | Project Page | Code *June 2019 - August 2021*

- Designed the **first** fully GPU parallel algorithm for Morse-Smale complex computation – improved upon the state-of-the-art by up to **8.6x**, with algorithmic improvements up to **577.7x** and **5.4x**.

Indian Institute of Science Bangalore, India
Research Assistant & Intern | Advisor: Prof. Ramsharan Rangarajan | Code *Jan 2018 - February 2019*

- Improved performance of a parallel mesh optimization algorithm DVR – reduced mesh optimization time by **47.4%**, enabled **100%** scalability with a **40×** speedup for mesh sizes ∼**14 million**.
- Implemented 'Provably Good Mesh Generation' by Bern et al. – developed open-source software for adaptive mesh refinement. Improved obstacle problem accuracy by an **order of magnitude**.

Indian Institute of Technology Mumbai, India
Research Intern | Advisor: Prof. Arindrajit Chowdhury | Project Page *May 2017 - June 2017*

- Developed a spray ignition setup for hypergolic propellant combustion in rocket propulsion.

Relevant Coursework and Skills

- **Courses**: Introduction & Advanced Topics in Data Science (AC 209a/b), Introduction to Machine Learning (MIT 6.036), Advanced Scientific Computing (AM 205), Systems Development for Computational Science (AC 207), Ethics for Engineers (MIT 6.9041), Advanced Natural Language Processing (MIT 6.8610), Probabilistic Machine Learning (AM 207), Large Language Models & Beyond (MIT 6.S986).
- **Skills**: C++, Python, CUDA, PyTorch, Machine Learning, Natural Language Processing, Data Science.

Work Experience

Tonita New York City
Research Engineer *August 2023 - Present*
Early employee of a startup building intelligent commerce search using language models.

basys.ai Cambridge, Massachusetts
Research Data Scientist *Dec 2022 - April 2023*
Developed machine learning models with **81%** classification accuracy and **93%** deployment accuracy to detect diabetic retinopathy automatically and in a timely manner using computer vision.

NVIDIA Cambridge, Massachusetts
Deep Learning Performance Intern *May 2022 - Aug 2022*
Developed and implemented two GPU-parallel algorithms for sliding window inference in 3D U-Net segmentation model. Obtained ∼**22%** performance improvement in testbed implementation of NVIDIA's MLPerf benchmark for the model.

Deloitte Bangalore, India
Business Analyst *Aug 2018 - June 2019*
Led cloud deployment of Windchill configurations on client servers, performance tuning and part classification.

Awards & Honors

- Research on universal adversarial attacks on large language models featured by **Science News**. *2023*
- Nominated for **Forbes 30 Under 30 – Boston**. *2024*
- Research adapted as a graduate machine learning course – CS6216: Advanced Topics in Machine Learning (Spring 2023) at National University of Singapore (NUS). *2023*
- Recipient of the **Adobe Research Women-In-Technology Scholarship 2022** – awarded a cash prize of $10,000 for accomplishments in academics and research in Computer Science. [Feature] *2022*
- Selected as an **ML Alignment Theory Scholar** and awarded $6000 by the Stanford Existential Risks Initiative. *2022*
- Selected to represent Harvard University at the **Grace Hopper Celebration 2022**. *2022*
- Selected as a **Google CS Research Mentorship Program Scholar 2021**. *2021*

Projects

- **Algorithmic Bias in Recidivism Risk-Assessment for Criminal Justice** | *Report*
  Predicted risk of recidivism in criminal justice using Lasso-regularized logistic regression on the COMPAS dataset. Detected biased predictions with and without race as a predictor and determined optimal classification thresholds.
- **Homelessness in the United States**
  Predicted homelessness trends in the US by comparing multi-linear, polynomial & Lasso-linear regression, random forests and boosting models. Obtained best predictive performance across 33 states from random forests and boosting.
- **Machine Learning for Medical Diagnosis**
  Developed machine learning models for pathology classification in chest X-rays and evaluated performance.
- **Parallel Matrix Factorization for Recommender Systems**
  Implemented parallel matrix factorization for gradient descent with a 2.7× speedup and runtime benefit of 424 secs.
- **End Gender-Based Violence** | *Project Page* | *Podcast* | *Feature*
  Detected a sharp rise in domestic violence in the US due to COVID-19 using interactive visualizations.
- **Fourier Transforms** | *Code* | *Project Page*
  Computed and visualized Fourier Transforms (3Blue1Brown) for input signals and extracted constituent pure signals.

Teaching & Technical Volunteering

- **Reviewer**, Regulatable Machine Learning Workshop, NeurIPS 2023.
- **Workshop Organizer**, Regulatable Machine Learning Workshop, NeurIPS 2023.
- **Course Developer & Teaching Fellow**, CS 181, Introduction to Machine Learning (Spring 2023), by Weiwei Pan.
- **Teaching Fellow**, CS50 - Introduction to Computer Science (Fall 2021), by David Malan.
- **Teaching Assistant**, Brave Behind Bars - Introduction to Computer Science (Summer 2022) | TEJI, MIT.

Invited & Contributed Talks

- ‘Why do universal adversarial attacks work on large language models?’, Poster Presentation at New Frontiers in Adversarial Machine Learning Workshop, ICML 2023.
- ‘Identifying the Risk of Diabetic Retinopathy Progression Using Machine Learning on Ultrawide Field Retinal Images’, Poster Presentation at American Diabetes Association’s 83rd Scientific Sessions and the MIT-MGB (Mass General Brigham) AI Cures Conference, 2023.
- ‘Why do universal adversarial attacks work on LLMs?’ Spotlight Talk at New England NLP Meeting Series, 2023.
- Research Seminar: ‘GPU Parallel Computation of Morse-Smale Complexes’, Flagship Pioneering Intelligence, 2023.
- ‘What makes a good explanation?’, Lightning Talk at Women in Data Science (WiDS) Conference, Cambridge 2023.
- ‘What makes a good explanation?’, Spotlight Talk at Trustworthy Embodied AI Workshop, NeurIPS 2022.
- Panelist, Harvard IACS Graduate Admissions Information Panel 2022.
- Panelist, Harvard IACS Research & Thesis Panel, Graduate Student Orientation 2022.
- Women in High Performance Computing (WHPC) Lightning Talk at the Supercomputing Conference 2021.
- ‘GPU Parallel Computation of Morse-Smale Complexes’, ACM ARCS Symposium 2021. [Slides] [Poster]
- ‘GPU Parallel Computation of Morse-Smale Complexes’, IEEE VIS 2020 Conference. [Talk] [Preview]

Social Impact & Service

- **Brave Behind Bars, MIT** | *Teaching Assistant* | [Washington Post] *May - Aug 2022*
  Taught Computer Science and mentored incarcerated individuals. Featured in Washington Post.
- **Harvard Square Homeless Shelter** | *Spring Break Volunteer* *March 2022*
  Volunteered and ran all overnight operations and administration of the homeless shelter for a day.
- **Vizathon 2021** | *Organizer* | [Webpage] *May 2021*
  Visualization hackathon with ∼400 registrations.
- **Humans of AI Podcast** | *Volunteer* | [Webpage] *Jan 2021 - Sept 2021*
  Volunteered with backend operations of a podcast which interviewed AI researchers.
- **She Belongs Podcast** | *Co-Founder & Co-Host* | [YouTube] [Spotify] [Medium] *Sept 2020 - Aug 2021*
  Discusses gender inequity and why women belong at the table. Over 2.4k views on YouTube.
- **Coronavirus Visualization Team** | *Project Planning Co-Director & Co-Lead* | [Webpage] *May 2020 - Aug 2021*
  Directed projects, founded one on gender-based violence and visually depicted a rise in violence.