Curriculum Vitae

Abinitha Gourabathina
PhD Student @ MIT EECS

(abinithago.github.io)
abinitha@mit.edu

Research Interests

My research focuses on reliable, responsible, and trustworthy Machine Learning, and my work spans group robustness, GenAI Agents, chain-of-thought, backdoor attacks, and more. I use tools and frameworks from Optimization, Probability, and Statistics. I am open for collaborations and can be reached via email.

Education

Massachusetts Institute of Technology (MIT), Cambridge, MA — 2024–2028 (Expected)
PhD, Electrical Engineering and Computer Science — 2028
SM, Electrical Engineering and Computer Science — 2026
GPA: 5.0/5.0

Princeton University, Princeton, NJ — 2019–2023
BSE, Operations Research & Financial Engineering
Thesis: What Seems to be the Problem? Stigmatizing Language in Patient Medical Notes
Magna Cum Laude
Minors in Computer Science; Cognitive Science; Linguistics; Statistics & Machine Learning
GPA: 3.9/4.0

Professional Positions

MIT LIDS, Research Assistant — 09/2024–05/2029
— *Advisors: Prof. Collin Stultz & Prof. Marzyeh Ghassemi*
— Computational Cardiovascular Research Group (CCRG) and Healthy ML research group
— Researching methods to investigate the latent spaces of foundational models, audit generative models in clinical settings, and improve model performance for safety-driven goals like model abstention and robustness to backdoor attacks

IBM Research, AI Research Intern — 06/2025–08/2025
— *Mentor: Prasanna Sattigeri*
— Conducted research into LLM abstention (models refusing to answer) and assessing abstention performance with chain-of-thought outputs
— Developed novel hallucination framework by framing model errors as answering the *wrong* question rather than answering the question incorrectly
— Created new state-of-the-art methods by *inverting* reasoning traces as a measure of model correctness, beating existing baselines by more than 5%

Bridgewater Associates, Investment Associate / Investment Associate Intern — 09/2023–07/2024, 06/2022–08/2022
— Studied macroeconomic investment strategies, focusing on systematizing quantitative strategies for sustainable equities portfolios

Carisk Partners, Data and Product Intern — 03/2020–05/2022
— Built several machine-learning models to estimate recovery time for patients from lower income minority communities by utilizing clinical text data and adapting traditional NLTK approaches to medical jargon using ontology linking
— Conducted NLP-based data analysis of patient chatting platforms from the bottom-up, presenting findings to team of medical professionals on patient engagement

Vanderbilt University, REU Summer Research Intern 06/2021–08/2021
— *Advisor: Bradley Malin, Mentor: Zhiyu Wan*
— Developed game-theoretic framework to maximize the amount of COVID-19 data shared while maintaining HIPAA compliance and modeling the resulting privacy risk using deep learning models
— Conducted research project funded by the National Library of Medicine (NLM) and the National Science Foundation Research Experiences for Undergraduates (REU)

Publications

- The Medium is the Message: How Non-Clinical Information Shapes Clinical Decisions in LLMs.
  Abinitha Gourabathina, Walter Gerych, Eileen Pan, Marzyeh Ghassemi. 2025. In Proceedings of the 2025 ACM Conference on Fairness, Accountability, and Transparency (**FAccT** '25). Association for Computing Machinery, New York, NY, USA, 1805–1828. https://doi.org/10.1145/3715275.3732121

- PanDa Game: Optimized Privacy-Preserving Publishing of Individual-Level Pandemic Data Based on a Game Theoretic Model.
  Abinitha Gourabathina, Zhiyu Wan, J. Thomas Brown, Chao Yan, Bradley A. Malin, 2023. In **IEEE TNB**, vol. 22, no. 4, pp. 808-817, Oct. 2023, https://doi.org/10.1109/TNB.2023.3284092

- The MedPerturb Dataset: What Non-Content Perturbations Reveal About Human and Clinical LLM Decision Making.
  Abinitha Gourabathina, Yuexing Hao, Walter Gerych, Marzyeh Ghassemi. 2025. arXiv preprint arXiv:2506.17163.

- What seems to be the problem? Stigmatizing language in patient medical notes.
  Abinitha Gourabathina (Senior thesis, Princeton University). Princeton DataSpace.
  http://arks.princeton.edu/ark:/88435/dsp01cv43p110t

- Corpus Analysis of English Modifying Adverbs
  Abinitha Gourabathina, Christiane Fellbaum
  Chapter in Khan, T., Bapuji, M., & Satapathy, A. (Eds.). (2021). Alternative Horizons in Linguistics: A Festschrift in Honour of Prof. Panchanan Mohanty. LINCOM GmbH.

Under Review

- Robustness Beyond Known Groups with Low-rank Adaptation
  Abinitha Gourabathina, Hyeown Jeong, Teya Bergamaschi, Marzyeh Ghassemi, Collin Stultz. 2026. arXiv preprint arXiv:2602.06924.

- Subtyping with ConCEPT: Contrastive Clinical Embeddings from Patient Trajectories
  Teya Bergamaschi, **Abinitha Gourabathina**, Collin Stultz. 2026.

- Answering the Wrong Question: Reasoning Trace Inversion for Abstention in LLMs
  Abinitha Gourabathina, Inkit Padhi, Manish Nagireddy, Subhajit Chaudhury, Prasanna Sattigeri. 2026.

- Poison-then-Hide: Finetuning-Activated Backdoor Attack on Pretrained Vision Encoders
  Qixuan Jin, **Abinitha Gourabathina**, Vinith Menon Suriyakumar, Walter Gerych, Marzyeh Ghassemi. 2026.

Selected Academic Awards, Teaching, and Service

Best Independent Work Prize from Princeton University Center of Statistics and Machine Learning	2023
Sigma Xi Research Award	2023
Recipient of Eric F. S. Pai '83 Summer Research Grant	2022

Teaching

COS401: Machine Translation, Head Teaching Assistant Spring 2023
— Taught lectures, organized practicum sections for bridging coding experience, and helped create problem sets
— Course taught by Lecturer Srinivas Bangalore

ORF245: Fundamentals of Statistics, Teaching Assistant Spring 2021
— Hosted weekly office hours and graded problem sets
— Course taught by Professor Matias Cattaneo

Invited Talks

ML Tea Talk, MIT	2025
FAccT '25 Main Conference Presentation, ACM	2025
Social and Ethical Responsibilities of Computing (SERC) Research Symposium, MIT	2025
IMES Seminar Series, MIT	2025

Departmental Service

Graduate Women in Course 6 (EECS) Executive Board, MIT	2025–Present
Laboratory of Information & Decision Systems (LIDS) Student Committee, MIT	2025–Present

Professional Service

Program Chair for LIDS Student Conference, MIT	2025–Present
Organizer of GW6 Research Summit, MIT	2025–Present
Reviewer for TS4H Workshop, NeurIPS ’25	2025
Program Committee for AIES Conference, AIES ’25	2025