

VEDANT GAUR

Philadelphia, PA // San Francisco, CA

☎ 650-762-9475 ✉ vedantg@wharton.upenn.edu 🌐 vedantgaur.com 📺 [vedantgaur](#) 📺 [vedantgaur](#)

Education

University of Pennsylvania, Jerome Fisher M&T Program

B.S.E., Computer Science (Engineering) and B.S. Economics, Finance (Wharton)

May 2027

Philadelphia, PA

University of Pennsylvania

M.S.E., Computer Science

Technical Skills

Languages: Python, Java, C/C++, LaTeX, HTML/CSS, JavaScript, SQL, OCaml, R, React

Soft Skills: Leadership, Research, Public Speaking & Communication, Journalism Writing, M&A (Mergers & Acquisitions)

Technologies/Frameworks: Python | (PyTorch, Tensorflow, NumPy, Pandas), AI, Deep Learning, NLP, CV, Ubuntu, AWS, Google Cloud, GitHub, JUnit, WordPress, Excel, Powerpoint, PostgreSQL, MongoDB

Relevant Coursework

- | | | | | |
|-----------------|---------------------------------|---------------------|------------------------------|-----------------------|
| • Discrete Math | • Databases/
Network Systems | • Stats/Probability | • Finance/Acct
Management | • Entrepreneurship |
| • DS&A | | • Operating Systems | • Operations | • Convex Optimization |
| • AI/ML | | • Deep Learning | | • Indian Philosophy |

Professional Experience

Amazon

June 2025 – August 2025

SDE Intern

Seattle, WA

- Engineered and deployed ML pipelines for intent classification, sentiment analysis, and routing prediction on scrubbed, multi-channel customer transcripts, for transfer reduction and contact routing improvement opportunities.
- Integrated RAG and LLM-based task classification to support natural language querying and actionable insight generation.
- Established automated data preprocessing, model retraining, and monitoring workflows with GDPR-compliant storage (Gluon S3), reducing manual analysis time by 80% and increasing actionable contact reduction from 5% to a projected 30%.

StanfordNLP

July 2024 – May 2025

Researcher

Stanford, CA

- Researching multi-trigger classification in mechanistic interpretability under Dr. Jan-Philipp Fränken to improve transparency in neural networks.
- Led model performance evaluations, open-sourced Anthropic's paper and met single-classification accuracy AUROC of $\sim 98\%$.

Princeton PLI

June 2022 – May 2025

Lead Author & Researcher

Princeton, NJ

- (June 2024 - Present) Collaborating with Dr. Tanya Goyal and Dr. Leqi Liu on advanced LLM benchmarking, enhancing cross-model evaluation reliability by 20%, published at EMNLP 2025 (first author).
- (June 2022 - July 2023) Worked with Dr. Nikunj Saunshi (now at Google Research) on symbolic math reasoning in LLMs, published at ACL 2023 (first author).
- Improved benchmark evaluation robustness to 80% accuracy for symbolic reasoning tasks.

Wharton Analytics Fellows

September 2023 – Present

Senior Analyst

Philadelphia, PA

- [IKEA] Led development of an LLM to enhance internal dashboard navigation, optimizing user experience and streamlining data access across departments.
- [Zillow] Modeled marketing spend for ZHL submissions, improving efficiency by 10%. Presented to Zillow's C-Suite and contributed to RapidMMM refinement.
- [Fox] Developed clustering algorithms for viewer personas, increasing targeted viewership by 15%.

BOLD AI

November 2021 – January 2023

Founder & CEO

- Led clinical testing in Bay Area clinics for a low-cost tool for detecting stutters in speech, working with 20+ SLPs and academics.
- Achieved an F1 score of 0.87 with Transformer models, surpassing earlier approaches.

Research Publications

[Preprint/NeurIPS] *Multi-Trigger Classification for Mixed Functional Mappings*

July 2024 – May 2025

[EMNLP 2025] *When should we (not) trust automatic LLM benchmarks?*

June 2024 – May 2025

[NeurIPS ATTRIB] *Learned Meta Token Reasoning*

September 2024 – December 2024

[NeurIPS ATTRIB] *Weak-to-Strong In-Context Optimization*

September 2024 – December 2024

[ACL 2023] *Reasoning in large language models through symbolic math word problems*

July 2022 – July 2023

[MIT IEEE URTC 2022] *Symbolic Math Reasoning with Language Models*

July 2022 – October 2022

[Caltech // JSR] *Lucas-Kanade Optical Flow Machine Learning Implementations*

June 2021 – December 2021

Projects/Hackathons (trunc.)

Hackathons | HackMIT (*2x win*), PennApps, xAI, AGI House, CalHacks, CalHacks LLMs, YHack

Deep Learning Live Interview Platform | Python, React, MERN Stack, AWS

June 2023

RAG-Based Research Literature Review Tool | Pytorch, Pinecone

June 2024

LLM-Based Resume Correction and Finetuning | Python, Pytorch, etc., Google Cloud Console

October 2023