Abhinav Raghunathan

130 E 37th Street, New York, NY 10016 | (609) 619-1344 | abhirags@seas.upenn.edu | GitHub: abhinavr2121 | abhiraghunathan.com

EDUCATION

University of Pennsylvania: School of Engineering and Applied Science *Philadelphia, PA*
- M.S.E. Data Science, GPA: 3.73 / 4.00. *Aug 2021 - Dec 2022*
- Penn Graduate Consulting Club (VP Management), Wharton Analytics Fellows (Technical Lead).

The University of Texas at Austin: Cockrell School of Engineering, College of Natural Sciences *Austin, TX*
- B.S. Computational Engineering, B.S. Mathematics, GPA: 3.71 / 4.00. *Aug 2017 - May 2021*
- Student Engineering Council, Texas Undergraduate Computational Finance (UCF).

EXPERIENCE

AI Scientist & Project Lead | KPMG AI Center of Excellence *(high-impact AI for all functions within Big Four firm)* *New York, NY*
- Leading development on solutions for data annotation and synthetic data generation for use with fine-tuning. *June 2024 - present*
- Directing proof-of-concept projects with startups offering AI solutions.
- Leading project development of audit-facing AI knowledge assistants.

AI Scientist, Credit Research | Vanguard IMFS *(high-impact AI within $8T+ AUM investment firm)* *Malvern, PA*
- Led product development for Vanguard's premier Generative AI product (fixed income investor-centric RAG) from scratch with a fully open-source LLM stack. Project cited by CIO in various conferences. *Jan 2023 - May 2024*
- Designed and built a system of LLMs featuring a novel, research-driven approach ("Mixture of Agents (MoA)") that analyzed and produced insight from 30,000+ documents in under 60 seconds.
- Resulting product outperformed most modern equivalents and had 100+ users across use cases in active/passive funds and financial advisory services, ~$50M positive impact.

Data Science Intern | FairPlay AI *(model risk mitigation for fair lending and insurance)* *Los Angeles, CA*
- Processed HMDA mortgage data from 1991-2021 (Python, Dask), resulting in 350M+ primary mortgage records, the largest longitudinal study on American mortgage fairness. *Jun 2022 - Dec 2022*
- Performed clustering analyses, correlations, and statistical inferences to identify historical lending bias.
- Research on the "State of Mortgage Fairness" (link) appeared on Forbes, USAToday, and Yahoo Finance (link).

Summer Analyst | Point72 Asset Management *(global hedge fund with $24B+ AUM)* *New York, NY*
- Applied statistical techniques (chi-squares, t-tests, etc.) to large datasets from third-party vendors (e.g., AppAnnie) to gain conviction on investment theses within video gaming, chiropractic, and furnishing sectors. *Jun 2020 - Aug 2020*
- Conducted in-depth research and financial analysis (DCF, comparables, three-statement models).
- Tracked ~100 tickers in the healthcare sector to quantify COVID-19 impact and identify mispriced stocks.

LEADERSHIP & INDUSTRY CONTRIBUTIONS

Founder | Ethical AI Database (EAIDB) (market research for B2B and B2C startups enabling responsible AI) *New York, NY*
- Categorized and researched 300+ companies within the responsible AI ecosystem (link). *May 2022 - present*
- Conducted market research and published findings, which have received 5,000+ downloads.
- Consulted for international government organizations such as Nordic Innovation and Australia's CSIRO.
- Spoke at various conferences and events like BGV's "Securing AI Summit" and Arthur's "AI Fest".

Technical Lead | Wharton Analytics Fellows, Teach for America *Philadelphia, PA*
- Clustered and segmented teacher profiles (DBScan, KMeans) to identify sources of churn. *Sep 2021 - Nov 2021*
- Designed an optimal candidate matching system with over 95% historical success using tree-based models.

PUBLICATIONS, PRESENTATIONS & WRITING

MoA is All You Need: Building LLM Research Teams using Mixtures of Agents | arXiv *Malvern, PA*
- Designed a novel approach using networks of specialized LLMs and SLMs to accomplish complex tasks. *September 2024*
- Co-authored Vanguard's first public language modeling paper on networks of specialized LLMs (link).
- Research is currently used in Vanguard's investor-focused RAG solutions.

Researcher, Writer | Various data science publications *Remote*
- Published articles in Towards Data Science, Towards AI, and Open Data Science Conference (ODSC). *Jul 2018 - present*

TEDx Speaker | TEDxUT Austin *Austin, TX*
- Presented on the dangers of algorithmic bias through modern-day examples in data science and ML (link). *Mar 2021*

SKILLS

Programming Languages: Python, R, JavaScript, SQL, MATLAB
Libraries: HuggingFace, Langchain, LiteLLM, LlamaIndex, OpenAI, LLM testing/evaluation libraries
Technologies: Amazon AWS, Microsoft Azure, Airtable, vector databases, language models, AI agents
Topics: LLM design, NLP, machine learning, responsible AI, financial modeling, statistical testing, project management
Spoken Languages: English, Tamil, Spanish, Hindi