M. Yafi Akhtar

582-203-8419 | yafi.7910@gmail.com | yafiakhtar.me | linkedin.com/in/yafiakhtar | github.com/yafiakhtar

Education

The Pennsylvania State University **Aug. 2023 – May 2027**
Bachelor of Science in Computer Science and Mathematics, GPA: 3.65 *University Park, PA*

Research

Undergraduate Research Assistant **Jan. 2025 – May 2025**
The Pennsylvania State University *University Park, PA*

- Designed and evaluated an ML pipeline for automated lecture segmentation using Whisper and OpenAI API, achieving 85% reduction in content retrieval time across a dataset of academic lecture recordings
- Applied NLP techniques including topic modeling, embedding search, cosine similarity to make unstructured lecture content semantically queryable with 92% retrieval accuracy
- Deployed the system to 250+ students during exam preparation, with adoption rates and query volume serving as proxies for measurable impact on study efficiency

Teaching

Learning Assistant - CMPEN 331 **Aug. 2025 – Present**
CSE @ Penn State *University Park, PA*

- Hold office hours three times a week for 300+ students in Computer Organization and Design, resolving doubts on instruction set architecture, memory systems, and MIPS/VHDL implementation
- Collaborated with course TAs to design and grade assignments

Experience

Software Engineer Intern **Jun. 2025 – Aug. 2025**
Jade Global *San Jose, CA*

- Built and shipped an internal lead generation tool aggregating data from LinkedIn, web scraping, and Snowflake Marketplace to surface job posting intelligence for consulting sales teams
- Designed and implemented REST API endpoints using Express.js and Node.js, integrating with Snowflake database for real-time data retrieval and processing
- Built chat assistant interface powered by Snowflake Cortex APIs, enabling sales teams to query lead data in natural language filtering by state, role, and company need, reducing reliance on manual SQL-based scoring

Web Director **Jan. 2026 – Present**
ACM @ Penn State *University Park, PA*

- Revived and scaled the club to 300+ active members, overseeing a web division of 50+ members
- Organized and taught weekly workshops covering web development fundamentals, guiding students through hands-on projects from ideation to completion and occasionally hosted industry speakers from professional web development roles to provide students with real-world career perspective
- Secured funding for events, AI tool subscriptions, and domain names, removing financial barriers for student projects

RoboX Team **Aug. 2024 – Dec. 2025**
RoboX @ Penn State *University Park, PA*

- Work on AI, computer vision, and robotics projects, gaining hands-on experience with OpenCV, Python, and machine learning models
- Develop image processing solutions for object detection, tracking, and edge detection, enhancing robotic system capabilities

Notable Projects

LLM Alignment Evaluation Framework | *Python, Ollama*

- Built a local evaluation pipeline using open-source LLMs to score model outputs for safety and quality, simulating RLHF-style alignment techniques
- Implemented a dual-model assistant-judge architecture to automatically assess responses across diverse prompts, improving output consistency without human annotation
- Designed modular prompt engineering and evaluation logic to enable lightweight, reproducible alignment experimentation

PaperTalk | *TypeScript, OpenAI, Next.js*

- Built a full-stack research assistant that lets users upload papers and interact with them through voice and text using RAG technology
- Implemented bidirectional voice interaction with real-time chat streaming and persistent conversation history, creating an interactive, assistant-like reading experience

Sentiment-Driven Market Volatility Index | *Python, MLlib, HDFS*

- Developed a scalable PySpark-based big data framework to process and analyze 8M+ financial text documents (80GB) from Reddit, NewsAPI, and SEC filings for real-time sentiment tracking
- Engineered a distributed sentiment analysis pipeline using spaCy, NLTK, and Spark MLlib, improving sentiment scoring accuracy to 65% through NLP methods
- Built an interactive dashboard visualizing sentiment–volatility relationships using Plotly and Matplotlib, facilitating rapid analysis for financial trend forecasting

Other: *Lead Generator (Jade Global), VidIQ (Penn State), Highlight (Chrome Extension)*

Technical Skills

Programming Languages: Python, C, Java, JavaScript, TypeScript, CSS, Shell

Technologies: Git, Docker, Kubernetes, AWS ECS, AWS EC2, FFmpeg, Linux, Snowflake, MongoDB, Supabase, Cloudflare, Vercel, React, Node.js, Express.js, Next.js, TailwindCSS

AI/ML: PyTorch, TensorFlow, Scikit-learn, NumPy, Pandas, spaCy, PySpark, LangChain, Ollama, OpenCV, pgvector

CS Coursework: Data Structures & Algorithms, Operating Systems, Systems Programming, Programming Language Concepts, Computer Organization & Design, Programming Models for Big Data, Object Oriented Programming

Math Coursework: Linear Algebra, Discrete Math, Real Analysis, Number Theory, Combinatorics, Numerical Analysis Methods, Statistics

Certifications

Machine Learning Specialization — *Coursera, Stanford University*

LLM Engineering — *Udemy*