Yuv Boghani

State College, PA-16803 • (203) 721-1985 • yuvboghani@gmail.com • www.linkedin.com/in/yuv-boghani

EDUCATION

The Pennsylvania State University August 2022 - May 2026
College of Engineering, EECS B.S Computational Data Science

SKILLS

Languages: Python, C, C++, Java, SQL, R, JavaScript, TypeScript, Bash/Shell, Verilog.
Data & ML: PyTorch, TensorFlow, PySpark, scikit-learn, Pandas, NumPy, Matplotlib, OpenCV, XGBoost, RAG, LLMs.
Web & Backend: React, FastAPI, Flask, Spring Boot, Node.js, Express, WebSockets, REST APIs.
Databases & Tools: AWS, GCP, Docker, CI/CD, Git, GDB, Valgrind, HDFS, NoSQL, PostgreSQL, MongoDB, Redis.
Relevant Coursework: Operating Systems, Computer Architecture, Data Structures & Algorithms, Systems Debugging, NLP, Object-Oriented Programming, Classical Machine Learning, Deep Learning, Computer Vision, Relational Databases.

WORK EXPERIENCE

Data Science Intern June 2025 - August 2025
Dr. Reddy's Laboratories - North America *Princeton, NJ*
- Developed a data pipeline using Python & SQL to calculate the 'cost of poor execution' by performing ETL on GBQ data.
- Analyzed COPE distribution to identify key risk segments, saving an estimated 180+ hours of manual effort annually.
- Leveraged an LLM to accelerate building a multi-view dashboard in React & JavaScript for visualizing COPE metrics.
- Optimized legacy data analysis code for historical inventory tracking by implementing Dask for parallel processing.

NLP Intern June 2025 - August 2025
Titan - Tata Group *Remote*
- Built a real-time voice agent using OpenAI's Realtime API and a custom FastAPI server for conversational AI.
- Integrated Twilio & WebSockets to stream low-latency, bi-directional audio to both landline and mobile phone networks.
- Built a RAG pipeline using LangChain & a Redis vector database to provide the agent with context-aware responses.
- Developed interruption handling and semantic caching to enhance conversational flow and significantly reduce API latency.

Research Intern May 2024 - August 2024
BITS Pilani - Dr. Aneesh Chivukula *Hyderabad, India*
- Developed a modular PyTorch framework to optimize an image classifier via a comparative analysis of CNN architectures.
- Implemented Tensor Methods for lossless compression on a dataset of 1.5 million high-resolution images.
- Deployed the image classifier into a production app, achieving 98% validation accuracy in predicting structural damage.

Data Analyst January 2024 - May 2024
The Daily Collegian *State College, PA*
- Automated a data pipeline using GraphAPI, Express, and Axios to track user engagement statistics for a Facebook page.
- Derived actionable insights from engagement metrics & demographics, boosting engagement by 15% in 4 months.

Research Assistant October 2022 - August 2024
Human-Technology Interaction Lab, Penn State - Dr. Yiqi Zhang *State College, PA*
- Built an XGBoost classification model using eye gaze & simulator data to predict driver takeover in AVs with 94% accuracy.
- Analyzed takeover quality vs. interface design using Tobii2 gaze data; presented findings at a university research exhibition.

Treasurer September 2022 - August 2025
Indian Cultural and Language Club, Penn State *State College, PA*
- Spearheaded fundraising and organized events for over 10,000 attendees, generating event revenues of $30,000.

PROJECTS

ChatDnD: Multi-Agent AI Dungeon Master | Python, LangGraph, React, MongoDB May 2025 - August 2025
- Built a multi-agent Dungeons & Dragons game using LangGraph, featuring specific agents for storytelling & game logic.
- Developed a real-time backend with Python & FastAPI, managing interactions for 5 concurrent players via WebSockets.
- Implemented a RAG pipeline with MongoDB to ground the AI's responses in the official D&D 5e ruleset, ensuring accuracy.
- Designed a dynamic frontend in React & JavaScript, featuring a live chat interface and character sheet management.

Bayesian Portfolio Optimization Framework | Python, PyTorch, NumPy, and Pandas January 2025 - May 2025
- Optimized an LSTM for time-series stock forecasting to an MSE of 0.030, using AlphaVantage, Yahoo & NASDAQ data.
- Quantified the correlation of external factors such as stock level data and sector indices with stock price.
- Synthesized external factors & correlation scores using Bayesian Networks to classify stocks by stability for portfolios.

High-Performance Distributed Storage System | C, C++ January 2024 - May 2024
- Built a linearized file system for a multi-disk storage system, featuring a custom malloc with segregated free lists.
- Programmed a multithreaded C client with pthreads & mutexes to handle high-concurrency distributed operations.