Varun Rao

443-695-7581 | varunvellorerao@gmail.com | linkedin.com/in/varunvrao | github.com/DurjaMan27

EDUCATION

University of Maryland, College Park
Aug 2023 - May 2026

B.S., Computer Science (ML - Honors), B.S., Mathematics, Minor in Robotics — *GPA: 3.98*

- *Extracurriculars*: Hack4Impact-UMD (Engineer), Apex Quant (Senior Quant Analyst), UMIACS Researcher
- *Coursework*: Intro to AI/ML, Advanced Data Structures & Algorithms, Real Analysis, Parallel Computing, Computational Methods, Combinatorics & Graph Theory, Robotic Path Planning, Linear Alg, Multivariable Calc

EXPERIENCE

Machine Learning Intern
June 2025 – Aug 2025

Capital One — *McLean, VA*

- Improved CI/CD deployment efficiency by 18% by developing an error detection and suggestion pipeline for a proprietary LLM development tool; leveraged AWS resources to validate deployment infrastructure combinations
- Increased domain-specific LLM accuracy by leading the development/testing of a novel synthetic data generation method for RAG frameworks based on hierarchical LDA topic modeling; led crucial progress in data orchestration, extensive retrieval recall evaluation, and benchmarking (awaiting paper submission & acceptance for early '26)
- Conducted 30+ hours of experimentation for NIPS-accepted RAG pipeline reranking research; earned full-time RO

Software Engineering Intern
May 2024 – Aug 2024

WordX - pWin.ai (formerly AIS) — *Reston, VA*

- Developed a fully automated data ingestion pipeline supporting an internal LLM used by 10+ organizations by collaborating with engineers to integrate the system into a retrieval-augmented generation (RAG) architecture
- Saved 3+ hours of manual work per document by utilizing advanced prompt engineering techniques and multi-agent systems to extract data from PDFs, convert them to JSON, and deliver them to production data stores
- Independently researched and implemented 8-10 various tools for optimized text retrieval including AzureDevOps, OpenAI Studio, cv2, pytesseract, Autogen, XML, while communicating progress regularly with PMs and tech leads

PUBLISHED RESEARCH

MAD-RAG (Capital One - NeurIPS 2025) | Paper
Dec 2025

- Contributed to a novel approach focused on augmenting existing RAG architecture by utilizing entire documents for retrieval and scoring based on relevance while using parallel, multi-agentic processes for adept summarization
- Performed extensive literature review and experimentation to verify novelty of approach, ensuring 27% reduction in chunk retrieval hallucinations and 22% increase in recall efficiency; advised by managers Juan Origgi, Qi Yu

LLMs Meet Finance (IEEE IDS 2025) | Paper
May 2025

- Co-authored a peer-reviewed paper on fine-tuning LLMs for finance, leveraging Qwen2.5 and DeepSeek-R1 with SFT, DPO, and RL techniques to improve performance on FinLLM Leaderboard, beating popular models by 7%
- Conducted numerous experiments and tuning operations across 20+ datasets to document and compare base model vs. fine-tuned model performance regarding various financial tasks; advised by Youran Sun and Dr. Haizhao Yang
- Paper accepted to IEEE IDS 2025 conference in New York City, earning 1st Prize in the FinRL Contest - Task 3

PROJECTS

3Domain | *OpenCV, Three.js, Mediapipe, Trimesh, Open3D*
Jul 2025

- Built an interactive web app that uses real-time hand tracking to rotate, reorient, zoom, and manipulate 3D scanned objects using only webcam-based gestures, with additional support for user-uploaded models in an .obj format.
- Converted original Python prototype to a full-stack JavaScript web app using Node.js, Express, OpenCV, and MediaPipe; computed 3D angular/linear velocities and axis of rotation to map gestures to precise object control.

TreeTrader | *MongoDB, Express.js, React, Node.js, yfinance, Google Gemini, LangChain*
Aug 2024

- Developed a full-stack paper trading application focused on encouraging long-term, intelligent investment strategy
- Constructed a MongoDB backend to store stock data from Yahoo! Finance API, allow for low-latency access
- Utilized professional trading indicators and Gemini to provide buy/sell/hold recommendations for new investors

TECHNICAL SKILLS

Languages: *(Proficient)* Python, JavaScript, TypeScript, HTML/CSS, C, Java, OCaml, *(Familiar)* SQL, R, Rust, C++
Tools/Tech: AWS (Lambda, SageMaker, EC2), MongoDB, Django, OpenCV, sklearn, PyTorch, Keras, ROS2, Linux
Certifications/Awards: AWS Certified Cloud Practitioner (2024), UMBC Hackathon 2023 (Best Use of MongoDB)
Interests: Football, Tennis, Formula One, Science-Fiction Literature, Video-Games, Sports Betting, Supercars