

Aayan Deb

(901) 336-8104 | aayan.deb@outlook.com | GitHub: [aayandeb](#) | LinkedIn: [aayandeb](#)

EDUCATION

Rhodes College

Bachelor of Science in Computer Science; Minors in Mathematics and Economics

Memphis, TN

Expected Dec 2027

- **GPA:** 3.98/4.00
- **Awards:** ICPC Mid-Central Regional Venue Champion (2025), Tennessee DataFest Winner(2026), Rhodes Hackathon Champion (2026), Presidential International Scholar, Honor Roll
- **Coursework:** Machine Learning, Artificial Intelligence, Data Structures & Algorithms, Computer Systems, Object-Oriented Programming, Applied Regression, Advanced Economics, Parallel Systems, Advanced Algorithms, Software Engineering

EXPERIENCE

Machine Learning Engineer Intern | *St. Jude Children's Research Hospital*

Jun 2026 – Aug 2026

- Automated clinical data extraction and cohort search across 750+ patient exams and 60+ query combinations, using JavaScript and REST APIs getting backend endpoints directly, achieving 18x speedup (6 hrs to 20 mins).
- Engineered a model-ready dataset of 1,100+ CT scans by developing an automated patient-data de-identification pipeline using a locally deployed LLM with vLLM inference.
- Programmed reproducible 5-fold nnU-Net, Swin-UNETR, and Mamba-based medical image segmentation workflows on HPC clusters managed by SLURM; developed a custom PyTorch decoder for an 847M-parameter 3D CT foundation model.

Full Stack Developer Intern | *Rhodes College Summer Research Fellowship*

May 2025 – Aug 2025

- Designed PostgreSQL schemas and Node.js/Express REST APIs with JWT authentication; maintained product quality through 30+ automated tests, 95% coverage, and GitHub Actions CI/CD.
- Worked with faculty stakeholders in an Agile/Scrum team to determine user requirements and build a full-stack research platform for 12 combinatorial games and 50+ concurrent sessions with real-time synchronization.

Undergraduate Research Assistant | *Rhodes College*

Feb 2025 – Present

- Engineered a latent-space world model (Beta-VAE encoder + GRU dynamics) in PyTorch under the JEPA framework, improving next-state latent prediction accuracy by 18% across 50k+ simulated transition pairs.
- Integrated a DreamerV3-style predictive agent with the LIDA cognitive architecture across 4 action spaces, reducing planning inference latency by 25% and co-authoring a manuscript for publication.
- Devised an empirical validation suite benchmarking 3 synthetic data generator classes (copula, VAE, GAN) across 6 statistical metrics (Wasserstein, MMD, KS, moment, inferential) on datasets exceeding 100k samples.

Student Leader – Resident Assistant & Writing Fellow | *Rhodes College*

Aug 2025 – Present

- **Writing Fellow:** Improved writing proficiency among 45+ first-year students by 18% on rubric-based evaluations by delivering 12 targeted workshops and conducting 60+ one-on-one conferences with actionable structural feedback.
- **Resident Assistant:** Achieved a 95% satisfaction rating across a 120-student residence hall by mediating 20+ peer conflicts, organizing 14 developmental programs, and managing 15+ urgent situations.

PROJECTS

MARC for MATA: Transit App | *Next.js, TypeScript, React, Flask, MongoDB, LLM* | [live demo](#)

Mar 2026

- Deployed a full-stack transit app with a live map of 40+ buses, GTFS trip planning, and accessibility-issue reporting.
- Integrated LLM to auto-score issues, manage photo uploads with resolution queue. Crowned champion in Rhodes Hackathon.

Beyond the Ball: Spatial Features for Soccer Prediction | *Python, PyTorch, scikit-learn, pandas*

May 2026

- Tested whether StatsBomb 360 spatial features improve possession-outcome and shot-creation prediction on UEFA Euro 2020/2024 data; MLPs vs. tree and logistic baselines raised possession-level F1 by ~0.03.
- Currently extending with graph neural networks (GNNs) over player-position graphs and a vector database for possession-embedding similarity search.

Healthcare Analytics Dashboard (ASA DataFest) | *Python, DuckDB, XGBoost*

Apr 2026

- Constructed an end-to-end analytics platform over ~700K healthcare events: DuckDB warehouse, Streamlit dashboard, and high-utilizer k-means segmentation with PCA and care-journey network visualizations; awarded **Best Visualization**.
- Trained XGBoost/CatBoost classifiers on leakage-aware cohort features to predict delayed follow-up across four horizons, reaching ~0.85 AUCPR with patient-grouped validation.

TECHNICAL SKILLS

Languages: Python, Java, C++, TypeScript, JavaScript, SQL, Bash

ML/AI: PyTorch, TensorFlow, MONAI, scikit-learn, XGBoost, vLLM, LLMs, GNNs

Frameworks & Data: React, Next.js, Flask, Node.js, Express, PostgreSQL, MongoDB, DuckDB, pandas, NumPy

Tools & Infrastructure: Git, GitHub Actions, Docker, Linux, REST APIs, SLURM, LSF, HPC, Microsoft Foundry