Akkshay Sundara Rajan

+1 (765) 772-5548 | akkshaysr0107@gmail.com | github.com/akkshay0107 | linkedin.com/in/akkshaysr

Education

Purdue University West Lafayette, IN

Bachelor of Science, Computer Science and Mathematics (Double Major) *Aug 2024 — (Expected) Dec 2027*

- Dean's List (4x), Semester Honors (4x), **3.92 GPA**
- **Undergraduate Teaching Assistant** - Numerical Methods
- INOI National Finalist, Codeforces Expert

Experience

Applied AI Intern Jun 2026 — Present

Anunta *Remote*

- Architected asynchronous microservices using FastAPI and RabbitMQ to queue, batch, and compress incoming user queries for an enterprise IT resolution agent, **cutting LLM inference costs by 20%**.
- Deployed llama.cpp inference infrastructure for a guardrails microservice as an isolated ClusterIP Kubernetes service.
- Built an agentic redaction pipeline using LangGraph to prevent cross tenant data leakage in a RAG pipeline.
- Developed automated agent evaluation CI benchmarks in DeepEval with LLM-as-a-judge scoring for accuracy.

Project Lead Sep 2025 — Present

Machine Learning @ Purdue - Student Organization *West Lafayette, Indiana*

- Led 6 developers in building an RL agent to play Pokemon battles, achieving **86% winrate** against baseline heuristics.
- Designed a transformer model from scratch in Pytorch, with a custom tokenizer and FlashAttention backend.
- Optimized model throughput to **40 actions/s** using query coalescing, pinned memory and multi-threaded parallel envs.
- Dockerized and deployed the inference architecture to Google Compute Engine, handling 20+ concurrent battles per node.

Undergraduate Researcher / ML Engineer Aug 2025 — Dec 2025

Dr. Philip Low's Lab *West Lafayette, Indiana*

- Designed and scaled distributed protein structure inference (NVIDIA Boltz) across 8 HPC nodes via SLURM.
- Built an automated postprocessing pipeline (OpenMMDL) to evaluate the stability of predicted protein-ligand complexes, simplifying candidate selection for drug discovery research.

Software Engineering Intern Jun 2025 — Aug 2025

Siemens Energy *Gurugram, India*

- Built an early-warning system (decision trees, scikit-learn) to detect high-risk gas turbine temperatures with **91% recall**.
- Built an FFT-based diagnosis system to classify turbine failure modes by matching sensor data against past failures.
- Developed an ETL pipeline (Python, SQL, Azure) to batch process and analyze **10 million sensor data records**.
- Automated the visualization of multi-sensor feeds in a dashboard (Plotly, Flask, Dash), **saving 1 hour** per inspection.

Projects

sift - Multimodal Semantic Search Engine github.com/akkshay0107/sift

- Engineered a local semantic search engine using Qdrant for multimodal retrieval across text, image, audio, and video files.
- Integrated multimodal ingestion pipelines using OCR, audio transcription, and dynamic CUDA/Metal dispatch.
- **Reduced application memory consumption by 40%** using quantized inference and a singleton model inference layer.

jepamania - Autonomous Driving World Model github.com/akkshay0107/jepamania

- Pretrained a self driving world model via self supervised learning (JEPA) on **10 hours** of simulation data.
- Built a vectorized CEM trajectory planner in Jax, evaluating 100s of trajectories in a **realtime 20 Hz control loop**.
- Built a post training sandbox environment for online reinforcement learning, training the model to adapt to new tracks.

Skills

- **Programming Languages**: Python, C++, Rust, Java
- **ML/Data**: PyTorch, Langchain, Scikit-learn, HuggingFace, Pandas, Qdrant, SQL, Postgres, Weights & Biases, MLflow
- **Backend/Cloud**: FastAPI, Docker, Redis, Celery, Azure, AWS, Google Cloud Platform, SLURM, Bash