Rishik Sarkar

New York, NY | +1-732-783-8669 | rishiksarkar02@gmail.com | github.com/RishikSarkar | rishiksarkar.com

Education

Cornell University Aug 2024 – May 2025
Master of Engineering in Computer Science *GPA: 3.9/4.0*

Rutgers University-New Brunswick Sep 2020 – May 2024
Bachelor of Science in Computer Science (Honors), Cognitive Science *GPA: 3.9/4.0*

Experience

Machine Learning Engineer Mar '25 – Present
Instalily AI (Google AI Accelerator) *New York, NY*

- Engineered a **RAG**-based agent and **MCP** server for United Rentals using **Azure AI Search** and a custom **LangChain**-style framework; optimized **embedding pipelines** and **vector index** queries, reducing average retrieval latency by 40%.
- Developed a **Twilio**-powered voice agent with **Gemini Live API** and **Azure Search**, integrating **ADK** and **MCP** to cut rental conversion time by 50% through intelligent voice response routing.
- Designed and deployed a clienteling agent with **ADK/MCP** backend and **Next.js** frontend, enabling iMessage-based luxury outreach via **BlueBubbles**; drove 40% of total sales via autonomous outreach execution.
- Created an autonomous Jira issue resolution agent using **GitHub APIs**, **semantic search**, and **CI/CD** pipelines, enabling fully automated ticket handling in minutes and improving DevOps turnaround time.

Independent Researcher Jan 2025 – May 2025
Cornell XR Collaboratory *New York, NY*

- Conducted **ML**-driven **AR/VR** research for **Quest** platforms; developed a **Unity** package featuring point- and grab-based 3D interactions using **AR Foundation** and **XR Interaction Toolkit**.
- Built **.NET**-based microservices integrating **Ollama** LLMs to automate 3D object selection and real-time mind-map generation, streamlining interactive **XR** workflows.

ML Full-Stack Developer Intern Jun '23 – Dec '23
Provenir (Fintech) *Parsippany, NJ*

- Built an automated credit-risk decision engine using **Decision Trees**, **Random Forests**, **XGBoost**, and **RNNs** via **scikit-learn** and **TensorFlow**, achieving 95% accuracy on historical customer data.
- Improved model transparency using **SHAP/LIME**, wrote 100+ unit tests with **MockMvc**, and containerized deployments on **Minikube**, boosting system reliability by 20% and reducing latency by 98%.
- Optimized **API** endpoints for artifact generation and log retrieval to support real-time risk monitoring, resulting in a 135% increase in application conversion rates.

ML Research Intern May 2022 – Jun 2023
Abraira Lab *New Brunswick, NJ*

- Processed 10,000+ behavioral samples with **MoSeq2** in **Python**; applied filtering and segmentation pipelines to support unsupervised clustering for rodent behavior analysis.
- Corrected **keypoint detection** anomalies, improving data quality by 60% and enhancing model accuracy.

Projects

MiniTorch | ***Python, PyTorch, CUDA, Numba*** Aug 2024 – Dec 2024

- Rebuilt core **Torch API** from scratch with support for **autodiff**, **broadcasting**, and **gradient ops**; implemented dynamic computation graph to support full backpropagation.
- Developed a high-performance tensor library using **CUDA kernels** and **Numba JIT** compilation to enable multi-dimensional operations with parallel GPU execution.

Protoclear | ***Next.js, FastAPI, TF-IDF, NER, Chroma*** Aug 2024 – Dec 2024

- Created an IRB compliance tool using **TF-IDF** and rule-based **NER** to analyze human-subject risks in protocols.
- Built a **RAG** pipeline using **LlamaIndex** and **Chroma** to deliver context-aware compliance suggestions.

Technical Skills

Languages & Frameworks: Python, Java, C/C++, C, JavaScript/TypeScript, SQL, FastAPI, Flask, .NET, Unity, Next.js
AI/ML & Data: ADK, MCP, PyTorch, TensorFlow, scikit-learn, Keras, LangChain, LlamaIndex, RAG, Pandas, NumPy
DevOps & Infra: Docker, Kubernetes, Jenkins, CI/CD, AWS, Azure, GCP, MongoDB, PostgreSQL, Git, Jupyter