

Rishik Sarkar

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

EDUCATION

Cornell University

Master of Engineering in Computer Science

New York, NY

Aug 2024 – May 2025

- GPA: 3.9/4.0; Merit Scholarship
- *Relevant Coursework:* Applied Machine Learning, Machine Learning Engineering, Virtual & Augmented Reality, HCI & Design, Introduction to Computer Vision, Natural Language Processing, 3D Interaction Design, Machine Learning for Health, Productizing Machine Learning

Rutgers University-New Brunswick

Bachelor of Science in Computer Science, Cognitive Science

New Brunswick, NJ

Sep 2020 – May 2024

- GPA: 3.9/4.0; Honors: Summa Cum Laude, SAS Honors Program, Phi Beta Kappa, Dean's List
- *Relevant Coursework:* Data Structures, Design and Analysis of Computer Algorithms, Principles of Programming Languages, Numerical Analysis & Computing, Principles of Data Management, Formal Languages & Automata, Introduction to Artificial Intelligence (Graduate)

TECHNICAL SKILLS

- **Languages:** Python, Java, JavaScript, TypeScript, C++, C, C#, Kotlin, Dart, Rust, SQL, MATLAB, Scheme, Swift
- **Web Technologies:** HTML, CSS, Flask, Next.js, Tailwind CSS, Spring
- **Frameworks & Libraries:** PyTorch, TensorFlow, scikit-learn, Keras, OpenCV, Pandas, Tkinter, NumPy, NLTK, SpaCy, Unity (with UPM), Hadoop
- **Databases:** MySQL, MongoDB, SQLite, JDBC, Supabase
- **DevOps & Tools:** Docker, Kubernetes, Jenkins, AWS, Git, Jira, Jupyter
- **Concepts:** Computer Vision, NLP, Data Science, Algorithms, Parallel Computing, AR/VR Interaction Design

INDUSTRY EXPERIENCE

Software Engineer (Contract)

AllAboutID

Oct 2024 – Nov 2024

Remote

- Built a secure, streamlined asset management system for an interior design startup, leveraging MongoDB Atlas for data storage and SVG handling for dynamic furniture integration in the API.
- Developed a responsive frontend using Next.js and Tailwind CSS to enhance user experience and asset visualization.

ML Full-Stack Developer Intern

Provenir

Jun 2023 – Dec 2023

Parsippany, NJ

- Engineered an automated credit risk decisioning solution by integrating Decision Trees, Random Forests, XGBoost, and RNNs into FLAML using scikit-learn and TensorFlow, achieving 95% prediction accuracy.
- Collaborated with a team of 5 engineers to enhance AI explainability via SHAP and LIME visualizations, facilitating data-driven decision-making.
- Implemented 100+ unit tests using MockMvc and streamlined deployment on Minikube, boosting software reliability by 20%.
- Refined API endpoints for artifact generation and log retrieval to enable seamless model monitoring and continuous learning.

SAT/ACT Tutor

The Princeton Review

Feb 2023 – Jun 2023

Princeton, NJ

- Leveraged innovative teaching technologies and instructional design to enhance student performance in standardized test preparation.
- Provided extensive after-hours support, ensuring deep comprehension and boosting test confidence.

RESEARCH EXPERIENCE

Independent Researcher

Cornell University, XR Collaboratory

Jan 2025 – Present

New York, NY (On-site)

- Conducting independent research under Professor Harald Haraldsson at Cornell Tech; focused on designing interaction techniques for AR/VR creator tools.
- Utilizing Large Language Models (LLMs) and Unity UPM packages to develop selection and manipulation tasks for AR/VR head-mounted displays.
- Developing a mindmap creation system to assist individuals with ADHD and developers in organizing complex information during 3D XR development.

Research Assistant

Princeton University (CCNP)

Sep 2023 – Aug 2024

Hybrid

- Developed a Python-based ETL pipeline to transform and consolidate over 800 Excel, CSV, and JSON files from five clinical studies into a unified SQLite database.
- Integrated automated schema generation with key constraints for clinical assessment data—capturing scales such as HDRS, BPRS, PANSS, and SCID-5—to record mental health conditions (e.g., depression, BPD).
- Designed a Tkinter-based GUI that enables non-technical researchers to execute custom SQL queries and perform advanced CRUD operations via Pandas.

ML Research Intern

Abraira Lab

May 2022 – Jun 2023

New Brunswick, NJ

- Employed Motion Sequencing (MoSeq2) to preprocess and generate over 10,000 high-quality training samples for an unsupervised ML model in a computational neuroethology study.
- Analyzed behavioral syllables and corrected anomalous keypoints, resulting in a 60% improvement in data quality.

PROJECTS

ReddiGist — Next.js, Flask, NLP, Reddit API, TypeScript

Oct 2024 – Present

- Developed a Reddit discussion analyzer that processes up to 5,000 comments per request using multi-threaded retrieval and n-gram detection (with case preservation).
- Implemented upvote-weighted scoring, position-based ranking, and a multi-stage filtering pipeline for substring deduplication and adaptive thresholding.
- Optimized performance via LRU caching, rate limiting, and batch processing; data persistence managed through Supabase.
- Planning future integration of RoBERTa for semantic analysis and the development of an analytics dashboard.

MiniTorch — Python, PyTorch, CUDA, Numba

Aug 2024 – Dec 2024

- Re-implemented the Torch API as part of Cornell's Machine Learning Engineering course.
- Developed from-scratch autodifferentiation and gradient computation to enable backpropagation for neural network training.
- Constructed a custom tensor library for multi-dimensional array operations critical to deep learning.
- Optimized performance using CUDA for GPU acceleration and Numba for JIT compilation.

Protoclear — NLP, LlamaIndex, Chroma

Aug 2024 – Dec 2024

- Developed an automated compliance toolkit for IRB regulations as part of Cornell's Product Studio incubator.
- Utilized TF-IDF and rule-based NER to extract keywords from IRB documents and map them to specific sections.
- Implemented a decision tree wizard to generate context-specific compliance recommendations.
- Integrated LlamaIndex with a Chroma vector database to establish a Retrieval-Augmented Generation (RAG) workflow for transparent, context-aware guidance.

PIU Prediction Using Time-Series & Physical Activity Data — Python, scikit-learn

Aug 2024 – Dec 2024

- Developed an ML pipeline to predict Problematic Internet Use severity in children and adolescents using physical activity data.
- Leveraged accelerometer actigraphy and fitness assessments from the Healthy Brain Network dataset; performed feature engineering including circadian rhythm metrics and PCA.
- Utilized supervised models (MLPs, XGBoost, LightGBM, TabNet) to achieve a top validation QWK score of 0.985; refined methodologies through a Kaggle competition.

Neko Sekai — Next.js, Tailwind CSS, TypeScript Feb 2024 – May 2024

- Developed an interactive web app where users can feed, pet, and customize an animated cat sprite.
- Implemented daily goal tracking, a leveling system, and an in-app currency mechanism to unlock new features.

Invasion of the Bot-Grabbers — Python, Pandas, PyTorch, Matplotlib Sep 2023 – Dec 2023

- Developed a grid maze simulation for a graduate-level AI course to enable a bot to navigate and rescue crew members while avoiding aliens.
- Implemented search algorithms (BFS, A*, D* Lite) for initial path-finding.
- Enhanced decision-making by integrating Bayesian networks and later an ACTOR-CRITIC reinforcement learning framework using PyTorch.

Tch.ai — Next.js, Tailwind, Flask, Keras, OpenCV, Pandas, MySQL Apr 2023 – Jul 2023

- Developed a full-stack web application deploying a Keras image classifier and tokenizer to recommend songs based on mood predictions derived from facial expressions or text.
- Trained the model on the FER-2013 dataset with preprocessing via OpenCV and a Haar Cascade classifier (~96% training accuracy, >70% validation accuracy).
- Designed a Next.js/Tailwind frontend supporting multiple genre selection methods and integrated a remote MySQL database for user account management.

LEADERSHIP AND SERVICE

Product Studio Collaborator Sep 2024 – Present

Weill Cornell Medicine, Cornell University (CWID: ris4016)

New York, NY

- Collaborating as part of Cornell's Product Studio course to develop health tech solutions that streamline HIPAA/IRB compliance for researchers and entrepreneurs.

Chief Technology Officer Sep 2023 – Aug 2024

Health Model United Nations

New Brunswick, NJ

- Orchestrated the design, development, and maintenance of a conference website using Next.js and Tailwind CSS, ensuring real-time updates and a user-friendly interface.
- Spearheaded communication infrastructure, social media strategy, and fundraising initiatives; actively participated in recruiting board members.

Co-Founder Mar 2022 – May 2023

The Verbose Project

New Brunswick, NJ

- Co-founded a non-profit tutoring platform for first-year college students, producing "Introduction to Python" tutorial videos.
- Provided supplemental tutoring at Rutgers learning centers to enhance foundational programming skills.

EXTRACURRICULAR AFFILIATIONS

- *XR Collaboratory at Cornell Tech* - Independent Researcher (2024)
- *The Phi Beta Kappa Society* - Member (2023)
- *Undergraduate Student Alliance of Computer Scientists* - Member (2021–2023)
- *Cognitive Science Club* - Member (2021–2023)

CERTIFICATIONS

- Human Subjects Research (IRB) Training, CITI Program - Completed Sep 2024 (Record ID: 64849869)
- Deep Learning Specialization, DeepLearning.AI (Coursera) - Completed Oct 2022
- Machine Learning Specialization, Stanford University (Coursera) - Completed Jun 2022
- AI For Everyone, DeepLearning.AI (Coursera) - Completed Mar 2022

PUBLICATIONS / ARTICLES

- Eisdorfer, J.T., Thackray, J., Theis, T., et al. (2023). *The Behavior Biomarker Scale (BBS): A machine-vision approach for automated locomotor recovery evaluation at millisecond timescales*. **bioRxiv**. <https://doi.org/10.1101/2023.10.31.564826>
- Sarkar, R. (2023, August 2). *Student insights: The Dark Side of Chatbot Therapy*. **Critical AI**. <https://criticalai.org/2023/07/31/student-insights-the-dark-side-of-chatbot-therapy/>

LANGUAGES

- English (Native), Bengali (Native), Hindi (Advanced), Japanese (Intermediate), German (Beginner)