

Sarthak Jain

sjain1@seas.upenn.edu | 732-520-9871 | linkedin.com/in/sarthak-jain2/ | github.com/SarthakJain2 | U.S. Citizen

EDUCATION

University of Pennsylvania

Master of Science in Computer Science and Data Science, GPA: 4.0/4.0

Philadelphia, PA

Expected May 2027

- **Coursework:** Applied Machine Learning, Big Data Analytics, Internet and Web Systems

- **Organizations:** Developer and Researcher for AI @ Penn, Claude Builder Club (Placed Top 3 in yearly hackathon)

Rutgers University-New Brunswick

Bachelor of Science in Computer Science and Data Science, GPA: 3.91, Major GPA: 3.97

New Brunswick, NJ

September 2022 - May 2025

- **Coursework:** Data Structures, Algorithms, Computer Architecture, Artificial Intelligence, Databases, Linear Algebra

- **Organizations:** Backend Developer and Technical Lead at HackRU, ARESTY Research Assistant

WORK EXPERIENCE

Robinhood

Menlo Park, CA

Incoming Software Engineer Intern

Incoming Fall 2026

Amazon Web Services (AWS)

New York, NY

Software Development Engineer Intern

May 2025 - August 2025

- Developed an end-to-end system for automated report delivery via console using **Java** for backend, **React + TypeScript** for frontend, and **SQL** for querying, eliminating 50+ manual data requests/month and improving data accessibility for stakeholders
- Built a Redshift ingestion pipeline with real-time S3 triggers, ensuring daily refresh and automated transformation of report data
- Migrated 100K+ daily records from NoSQL **DynamoDB** to **Redshift** using **Lambda**, EventBridge, S3 replication, and SQS
- Created a custom API with presigned URLs to serve 90 days of reports across 10+ report types via a centralized internal console

Machine Learning Lab

New Brunswick, NJ

Undergraduate Researcher

December 2023 – May 2025

- Enhanced LLM evaluation by contributing to the Multitask Language Understanding with Symbol Replacement (MMLU-SR) dataset, showcasing over a 30% drop in model performance, highlighting the symbolic reasoning limitations of different LLMs
- Executed benchmark testing for MMLU-SR on Llama-3-70B and Gemini 1.5 Pro, demonstrating a performance decrease of 25%
- Published evaluation results at peer-reviewed **EMNLP GenBench Workshop** in collaboration with a leading USC research team

Rutgers Rail and Transit Research Lab

New Brunswick, NJ

Undergraduate Researcher

September 2023 – May 2025

- Collaborated on the Using Artificial Intelligence for Next-Generation Intelligent Transportation project that will analyze track safety challenges, path efficiency, and real-time positioning accuracy in 10+ low connection areas of the NJ Transit Rail System
- Applied **OpenCV** and **Python** to generate 3,000+ bounding boxes to validate model accuracy in detecting track safety hazards
- Developed a user-friendly full-stack **Tkinter** interface for crack visualization, video navigation, and defect validation workflows
- Co-authored research paper published in **Smart and Resilient Transportation Journal** on AI applications to enhance rail safety

Gateway (Acquired by Circle)

New York, NY

Software Engineer Intern

June 2024 - August 2024

- Developed protocol layer in a Web3 environment using **Rust**, creating 50+ Zero-Knowledge rollups for anchoring credential data
- Refined **TypeScript** SDK to improve developer experience for issuing and consuming verifiable credential data in 7+ languages
- Created 50+ data dashboards to show transaction activity on blockchain protocol layer using Dune Analytics and Flipside Crypto

PROJECTS

Rutgers ML Course Finder

October 2024

- | Python, PyTorch, Pinecone, React.js | **HackRU 2024 Winner**
- Implemented Pinecone for optimized storage and **60% faster retrieval** time of vector embeddings, facilitating real-time similarity searches and delivering probabilistic confidence scores for accurate matching of students' interests with course offerings
 - Developed website with React.js to search classes using interest-based keyword search and integrated user feedback functionality
 - Integrated similarity scoring using **5000+ vector embeddings** created by PyTorch, enabling context-aware retrieval of courses
 - Utilized BeautifulSoup4 for web scraping of relevant course information to improve the accuracy and accuracy of search results

XAI Precancer

January 2024 – May 2024

- | Python, Convolutional Neural Networks, PyTorch, NumPy, Jupyter Notebook
- Implemented DeepLabv2 for segmenting cell images to identify precancerous symptoms, enhancing early detection capabilities with a **96% validation accuracy** rate in a dataset of **1000+ histology images** and to be used at Robert Wood Johnson Hospital
 - Leveraged SoftMax function for precise predictions across **3 distinct tissue classes**, improving the model's diagnostic accuracy
 - Developed masks for each image, enabling visual identification of precancerous regions through detailed semantic segmentation

TECHNICAL SKILLS

Languages: Python, Java, Go, SQL, R, Rust, Typescript, JavaScript, C, C++, MATLAB

Technologies: React, AWS, NumPy, Pandas, PyTorch, Pinecone, MySQL, NoSQL, Docker, Git, REST APIs