

# Riju Datta

Tampa, FL 33625 / (813) 957-6827 / [dattarij@alumni.upenn.edu](mailto:dattarij@alumni.upenn.edu) / [LinkedIn Profile](#) / [GitHub Profile](#)

## Education

University of Pennsylvania, School of Engineering and Applied Science

Graduation: December 2024

GPA: 3.56/4.00

Master of Science in Engineering, Robotics / Bachelor of Science in Engineering, Computer Science / Minor in Statistics

**Relevant Coursework:** Intro to Robotics, Principles of Deep Learning, Learning in Robotics, Machine Perception, Big Data Analytics, Engineering Entrepreneurship, Database and Information Systems, Computer Org. and Design, ML on Edge Devices, Operating Systems, Engineering Negotiations

## Skills

- **Languages:** Python, Java, C/C++, JavaScript, SQL, R, HTML, Bash, Verilog, MATLAB, OCaml
- **Frameworks:** PyTorch, MuJoCo, TensorFlow, NumPy, pandas, Matplotlib, scikit-learn, AWS DynamoDB, Spark, Node.js, React.js, FastAPI

## Work Experience

GRASP Lab / Research Assistant (Perception, Action, and Learning Group)

May 2024 – Present

- Assembled a visual imitation learning pipeline using ManiSkill2, a benchmark for generalizable manipulation tasks, to integrate the [im2contact](#) model with behavioral cloning policies for high-precision, contact-rich tasks, such as peg insertion and kit assembly
- Wrote a custom PyTorch dataset to ingest demonstration episodes of robot arm trajectories and output per-frame RGB/depth images collected at variable camera poses, object segmentation masks, and end-effector position pixel coordinates

Parabole.ai / Startup Software Engineering Intern

May 2023 – August 2023

- Implemented a multimodal pipeline using Hugging Face's LayoutLMv2 large language model for token classification and key-value extraction, customizing it to accurately process purchase order documents, including detailed line-item recognition
- Preprocessed over 250 images using the Azure Form Recognizer SDK for document structure analysis and PyTesseract for OCR and built a data pipeline to convert images into tokenized formats
- Created novel PyTorch dataloaders, enabling efficient model training over 16 epochs, and tuned hyperparameters to improve model performance, achieving consistent extraction of key values from purchase orders

Accenture Federal Services / Software Engineering Analyst

May 2022 – August 2022

- Developed an object detection API using CLIP embeddings and scikit-learn, enhancing the accuracy of CCTV image frame analysis and streamlining data categorization systems for better information retrieval
- Deployed a FastAPI web service with dedicated endpoints for training data collection and model inference, enabling automated data processing, seamless content management, and improved workflow efficiency

Urbanowicz Lab, Perelman School of Medicine / Student Researcher

June 2021 – August 2021

- Designed an innovative multi-class AutoML pipeline for allele biomarker classification, assessing the performance of 13 ML algorithms using the scikit-learn library and visualizing results with confusion matrices and performance metrics
- Conducted comprehensive statistical analyses on binary SNP datasets to identify features correlated with congenital heart disease, achieving high precision (95.5%, 97.9%, 99.1%) and ROC-AUC scores (0.955, 0.979, 0.990) and presenting [key research findings](#) at Fall 2021 PURM Research Expo

## Projects

- **[Comparative Analysis of Deep RL Algorithms](#):** Reconstructed the Twin Delayed DDPG and Soft Actor-Critic algorithms using PyTorch and measured their average rewards on MuJoCo Walker2D "Run" and "Walk" tasks, writing benchmarking and visualization scripts for simulation testing
- **[GitHealth](#):** Launched a full-stack web app using pandas, Node.js, FastAPI, and the NetworkX Python library to assess engagement within the Autoware open-source community, visualizing metrics for 15 companies at individual and organizational levels
- **["PennOS" Operating System](#):** Built job scheduling, command line functionalities, and kernel-level management in C for a multi-threaded OS
- **[Football League Analytics](#):** Produced 7 dynamic web pages with React.js and wrote 10 SQL database queries for a web app built to analyze European football league statistics, comparing player ratings and match outcomes
- **[TinyML Gaze Tracking](#):** Engineered an ML application with Edge Impulse for real-time classification of eye movements using a mobile phone camera, collecting over 400 images across training and testing datasets
- **[Superscalar Processor](#):** Devised a 5-stage, pipelined superscalar processor in Verilog, handling instruction and program counter logic, ALU operations, dependency resolution, and data memory management

## Community Involvement

[Penn Masala](#) / South Asian A Cappella Group Vocalist

September 2020 – August 2024

- Toured 6 cities across India for 2 weeks in May 2023, sang at the White House in June 2023, and performed 2 shows at the Paris Olympics' inaugural India House in August 2024; performed in 7-8 collegiate and corporate shows across the US each semester
- Managed marketing, ticketing, and logistics for a 1250-person Spring Show concert and arranged 8 musical compositions

Wharton Undergraduate Healthcare Club Incubator Program / Pitch Competitor (GenoEHR)

January 2023 – April 2023

- Conceptualized a startup idea integrating genomic data analysis into cloud-based EHR systems, addressing unmet clinical genomics needs
- Generated market analyses, attended 6 biotech expert-led workshops, and presented a 21-slide [pitch deck](#), winning 2nd place in the Launch Phase pitch competition out of 15 teams in total