

# RAJEEV DATTA

1200 E. California Blvd, Pasadena, CA 91125

☎ 574-286-7126

✉ [rdatta@caltech.edu](mailto:rdatta@caltech.edu)

🌐 [linkedin.com/in/rajeev-datta-1795ba1b1](https://www.linkedin.com/in/rajeev-datta-1795ba1b1)

🐙 [github.com/rajeev-datta](https://github.com/rajeev-datta)

## Education

---

### Cornell University

Sep. 2024 – Present

*PhD Candidate, College of Computing and Information Science (Machine Learning Focus)*

### California Institute of Technology

Sep. 2020 – June 2024

*Bachelor of Science in Computer Science, Minor in Information and Data Sciences*

*GPA: 4.1/4.3*

## Research Experience

---

### California Institute of Technology, RSRG | Rigorous Systems Research Group

June 2022 – Dec. 2023

*Summer Undergraduate Research Fellowship Intern - Advised by Prof. Adam Wierman and Yisong Yue*

*Pasadena, CA*

- Created RL environments for simulating the control of battery storage systems in real-time market bidding, enabling researchers to monitor advancements and pinpoint challenges in a practical sustainability scenario.
- Created a single-agent environment using historical prices to simulate a real-time market, sourcing pertinent electricity price data from California Independent System Operator (CAISO) price dataset.
- Enhanced the single-agent environment by integrating carbon costs, including a carbon penalty determined from historical CAISO Marginal Operating Emissions Rate (MOER) data.
- Further improved environment realism by incorporating network congestion based on an IEEE test case transmission network and replacing historical price data with a convex optimization problem known as the market dispatch problem which utilizes historical CAISO data and participant bidding strategies to calculate electricity prices based on demand or injection of electricity across the grid.
- Examined three off-the-shelf RL algorithms (SAC, DQN, and PPO) and assessed their performance in both in-distribution (training and testing from the same data distribution) and out-of-distribution (training and testing from different data distributions) scenarios.
- Presented the key findings of trained agent performance in environments at the NeurIPS 2023 poster session, emphasizing the significance of robust RL algorithms and illustrating the intricacies of the electricity market bidding challenge.

### MIT, LIDS | Laboratory for Information & Decision Systems

June 2023 – Oct 2023

*Research Intern - Advised by Prof. Cathy Wu*

*Cambridge, MA*

- Created realistic, SUMO-based large-scale traffic simulation incorporating surrogate carbon emission models and realistic driving behavior, allowing for better estimations of eco-friendly driving's impact on emissions.
- Developed an RL algorithm in PyTorch to reduce emissions by over 20% when compared to the industry-standard.
- Developing a transfer learning training scheme to extract general-purpose, human-compatible driving policies.

### Stanford University, RSL | Radiological Sciences Laboratory

July 2021 – Sep. 2021

*Summer Undergraduate Research Fellowship Intern - Advised by Prof. Shreyas Vasanawala*

*Palo Alto, CA*

- Applied deep learning techniques for image reconstruction to accelerate MRI scans by mapping undersampled MRI images to their fully-sampled counterparts, maintaining diagnostic quality.
- Investigated the V-Net architecture and assessed the influence of spatial self-attention on performance using image metrics such as Structural Similarity Index (SSIM) and Peak Signal-to-Noise-Ratio (PSNR).
- Optimized the V-Net architecture with varying self-attention levels in the encoding and decoding phases using the NiftyNet platform.
- Created custom image-metric loss functions using SSIM and PSNR values between prediction and ground truth images. Additionally, designed a feature-based loss function by feeding predicted and ground truth images into a pre-trained VGG-19 model and comparing features at different layers.
- Improved SSIM and PSNR scores by 0.044% and 0.17% and presented key findings to over 70 members of Stanford Medical School's radiology department and at the Caltech Undergraduate Summer Symposium.

## Relevant Coursework

---

- |                    |                         |                        |                     |
|--------------------|-------------------------|------------------------|---------------------|
| • Machine Learning | • Computing Systems     | • Software Design      | • Compilers         |
| • Data Structures  | • Distributed Computing | • Computation Theory   | • Operating Systems |
| • Algorithms       | • Probability Models    | • Network Science      | • Quantum Computing |
| • Robotics         | • Linear Algebra        | • Discrete Mathematics | • GPU Programming   |

## Work Experience

---

### California Institute of Technology, CMS Department

Jan. 2023 – June 2023

CS 155: Machine Learning and Data Mining Teaching Assistant

Pasadena, CA

- Supported professor in managing a core ML course with 190+ undergraduates, including developing and grading assignments.
- Conducted two hours of weekly office hours for selected assignments covering matrix factorization techniques and hidden Markov models.
- Evaluated specific problems in course assignments for all students.
- Received a 5/5 on all metrics from students' teaching and quality feedback forms.

## Publications

---

1. C. Yeh, V. Li, **R. Datta**, J. Arroyo, N. Christianson, C. Zhang, Y. Chen, M. Hosseini, A. Golmohammadi, Y. Shi, Y. Yue, and A. Wierman, "SustainGym: A Benchmark Suite of Reinforcement Learning for Sustainability Applications," in Thirty-seventh Conference on Neural Information Processing Systems Datasets and Benchmarks Track, New Orleans, LA, USA, Dec. 2023. [Online]. Available: <https://openreview.net/forum?id=vZ9tA3o3hr>.
2. C. Yeh, V. Li, **R. Datta**, Y. Yue, and A. Wierman, "SustainGym: A Benchmark Suite of Reinforcement Learning for Sustainability Applications," in NeurIPS 2022 Workshop on Tackling Climate Change with Machine Learning, Dec. 2022. [Online]. Available: <https://www.climatechange.ai/papers/neurips2022/38>.

## Presentations & Awards

---

- 5/2023. Presented "Semi-Automatic Digitization of Single-Line Diagrams for Accelerated Deployment of Algorithmic Control on Real-World Power Systems." Meeting of the Minds
- 10/2022. Presented "SustainGym: A Benchmark Suite of Reinforcement Learning for Sustainability Applications." NeurIPS 2022 Workshop On Tackling Climate Change With Machine Learning
- 8/2022. Presented "SustainGym: A Benchmark Suite of Reinforcement Learning for Sustainability Applications." SURF Summer Symposium 2022 - California Institute of Technology
- 9/2021. Presented "Developing optimal U-Net variant to reconstruct undersampled Magnetic Resonance Imaging (MRI) images to improve image quality." SURF Summer Symposium 2021 - California Institute of Technology
- Valedictorian of Penn High School, class of 2020

## Technical Skills

---

**Languages:** Python, Java, C, C++, Rust, HTML/CSS, JavaScript, OCaml, Assembly, MatLab, Mathematica, SQL, R

**Developer Tools:** VS Code, Eclipse, AWS, Anaconda, RVIZ

**Packages/Frameworks:** Linux, Git, ROS 2, TensorFlow, PyTorch, Keras, Pandas, NetworkX, CVXPY, Numpy, SKLearn, Qiskit, Flask, Open AI Gym, Petting Zoo, Matplotlib, OpenCV, CUDA

## Projects

---

### GPU-accelerated Video Processing | C++, CUDA

May 2024 - May 2024

- Developed a CUDA-based application that applies canny edge detection to input video's frames and recombines frames.
- Parallelized canny edge detection algorithm over image pixels.

### Compiler | OCaml

September 2023 - December 2023

- Developed a compiler to convert full-fledged high-level programming language in the form of an abstract syntax tree (AST) to assembly language, which can then be compiled to machine language using standard tools.
- Programmed compiler handles scoping issues, conditionals, iterative loops, garbage collection, register allocation, and functions.

### Semi-Automatic Single-Line Diagram Annotation | Python

April 2023 - June 2023

- Developed precise annotations for single-line diagrams of electric grids to facilitate digital analysis and control algorithm application, especially for adapting grids to renewable energy sources.
- Implemented a template-matching algorithm with non-max suppression for accurate symbol recognition.
- Utilized the labeled dataset to automatically optimize algorithm hyperparameters, such as thresholds, by constructing precision-recall curves and selecting parameter with largest F1 score.
- Presented the resulting algorithms at Caltech's Meeting of the Minds, demonstrating their potential to reduce annotation times from days to seconds relevant to computer vision application.

## LEGO Stacking Robotic Arm | *Python, Linux*

January 2023 - March 2023

- Built a five-degree-of-freedom robotic arm using HEBI motors and 3D-printed components to stack large LEGO blocks from a given workspace onto a designated grid.
- Developed an automated ROS 2 program for this task, running it on a local Linux machine to control the motors.
- Showcased the project successfully in an open demo at the end of the Winter term.

## Rotating Tower of Hanoi Simulation | *Python*

September 2022 - December 2022

- Developed a ROS package for simulating a 6-DOF robotic arm and a rotating table with three equidistant poles.
- Programmed the arm to move a cube between poles as the table rotates periodically.
- Included an adjustable center point for the turntable to create singularities by moving it away from the arm.

## Chest X-Ray Competition | *Python*

April 2022 - June 2022

- Utilized custom Convolutional Neural Networks (CNNs) to predict pathologies based on Chest X-Ray images.

## Loan Data Kaggle Competition | *Python*

February 2022 - March 2022

- Collaborated with a team of four students to predict loan repayment based on applicant data and financial history.
- Processed tabular data and trained a custom Deep Neural Network (DNN) for improved accuracy on a public dataset.
- Achieved 13th place out of 57 teams in the Kaggle competition.

## Tank Trouble Game Implementation | *C*

April 2021 - June 2021

- Collaborated in a four-student team to recreate the Tank Trouble game using C.
- Developed power-up classes to enhance gameplay realism and excitement.
- Assisted teammates in implementing features such as mouse tracking and multiplayer functionality.
- Conducted comprehensive testing of essential game classes for physics simulation and gameplay.

## Leadership / Extracurriculars

---

### AfterMath

Fall 2021 – Present

*Travel Captain*

*California Institute of Technology*

- Joined the collegiate ultimate frisbee team during my Sophomore year of college.
- Took charge of travel logistics for the club as a part of the club's leadership team this year.
- Organized an overnight tournament in February 2023.

## Referees

---

### Prof. Adam Wierman

California Institute of Technology

*Professor, Dept. Computing + Mathematical Sciences (CMS)*

*Pasadena, California*

Director, Information Science and Technology

Email: [adamw@caltech.edu](mailto:adamw@caltech.edu)

### Prof. Yisong Yue

California Institute of Technology

*Professor, Dept. Computing + Mathematical Sciences (CMS)*

*Pasadena, California*

Principal Scientist at Latitude AI

Email: [yyue@caltech.edu](mailto:yyue@caltech.edu)

### Prof. Shreyas Vasanawala

Stanford University

*Professor, Dept. Pediatric Radiology*

*Palo Alto, California*

William R. Brody Professor of Pediatric Radiology and Child Health

Email: [vasanawala@stanford.edu](mailto:vasanawala@stanford.edu)