Rajeev Datta

🏠 rajeev-datta.github.io   in LinkedIn   ✉ rd635@cornell.edu   G Google Scholar   🪪 U.S. Citizen

Research Interests

3D/4D Reconstruction, Interpretable ML, Large Vision Foundational Models, Vision-Language Models

Education

Cornell University	**Sep. 2024 – Present**
PhD Student, College of Computing and Information Science (Machine Learning Focus)	*GPA: 3.9/4.0*
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*

Technical Skills

Languages: Python, Java, C, C++, Rust, OCaml, Assembly, MatLab, Mathematica, SQL, R, Julia
Developer Tools: VS Code, Eclipse, AWS, Anaconda, RVIZ
Packages/Frameworks: Linux, Git, ROS 2, TensorFlow, PyTorch, Keras, NetworkX, CVXPY, SKLearn, Qiskit, Flask, Open AI Gym, Petting Zoo, Matplotlib, OpenCV, CUDA, nerfstudio, SymbolicRegression, vLLM

Research Experience

Cornell University, Cornell Graphics Vision Group   **Sept. 2024 – Present**
Graduate Research Assistant - Advised by Prof. Kavita Bala and Prof. Bharath Hariharan   *Ithaca, NY*

- **Attribute Learning for Interpretable Zero-Shot Classification**
  - ∗ Devised highly interpretable zero-shot classification framework that predicts based on attributes.
  - ∗ Proposed training regime to fine-tune LlaVA for fine-grained attribute understanding.
  - ∗ Implemented automatic attribute annotation pipeline, reducing end-to-end time from days to hours.
  - ∗ **Keywords:** *Large Vision Language Models, Natural Language Processing, Zero-Shot Classification*

- **Data-Driven Uncertainty Quantification for Novel View Synthesis**
  - ∗ Developed first loss-based training scheme for uncertainty-aware Neural Radiance Fields (NeRFs).
  - ∗ Used learned uncertainty intervals to identify highly ambiguous regions of a scene.
  - ∗ Devised active perception regime to improve reconstruction quality for digital twins and robotics
  - ∗ **Keywords:** *3D Reconstruction, Neural Radiance Fields, Uncertainty Quantification*

California Institute of Technology, RSRG   **June 2022 – Dec. 2023**
Summer Research Fellow - Advised by Prof. Adam Wierman and Prof. Yisong Yue   *Pasadena, CA*

- Co-developed SustainGym, a RL benchmark suite released in NeurIPS 2023 Datasets & Benchmarks track.
- Evaluated state-of-the-art RL algorithms (SAC, DQN, and PPO) in out-of-distribution scenarios.
- Motivated future RL benchmarks (e.g., CommonPower and RL2Grid) and algorithms (e.g., C-MORL).
- **Keywords:** *Deep Reinforcement Learning, Sim2Real Gap, Robust Machine Learning*

Massachusetts Institute of Technology, LIDS   **June 2023 – Oct 2023**
Research Intern - Advised by Prof. Cathy Wu   *Cambridge, MA*

- Developed multi-task RL approach to learn traffic policies that cut emissions up to 20% over IDM.
- Built high-fidelity traffic simulation with surrogate carbon emission models and realistic driver behaviors.
- Explored transfer learning to allow trained agents to generalize across unseen traffic patterns.
- **Keywords:** *Deep Reinforcement Learning, Transfer Learning, Multi-Objective Reinforcement Learning*

Stanford University, RSL | Radiological Sciences Laboratory   **July 2021 – Sep. 2021**
Summer Research Fellow - Advised by Prof. Shreyas Vasanawala   *Palo Alto, CA (Virtual)*

- Developed method to construct images from undersampled MRI images, accelerating acquisition times.
- Investigated the influence of spatial self-attention on reconstructed image quality.
- **Keywords:** *Perceptual Loss, Spatial Self-Attention, Image Reconstruction*

Relevant Coursework

- Machine Learning
- Data Structures
- Algorithms
- Robotics
- Computing Systems
- Distributed Computing
- Probability Models
- Linear Algebra
- Software Design
- Computation Theory
- Responsible ML
- Discrete Mathematics
- Compilers
- Generative Models
- Quantum Computing
- GPU Programming

Teaching Experience

California Institute of Technology, CMS Department   **Jan. 2024 – June 2024**

CS 155: Machine Learning and Data Mining - Teaching Assistant with Prof. Yisong Yue   *Pasadena, CA*

- Assisted in teaching and managing a core Machine Learning course with 200+ undergraduates.
- Redesigned documentation and template code for deep learning assignment to improve clarity.
- Removed and replaced unclear conceptual questions from assignments to properly evaluate core ML principles.

California Institute of Technology, CMS Department   **Jan. 2023 – June 2023**

CS 155: Machine Learning and Data Mining - Teaching Assistant with Prof. Yisong Yue   *Pasadena, CA*

- Assisted in teaching and managing a core Machine Learning course with 190+ undergraduates.
- Redesigned documentation and template code for Hidden Markov Models assignment to improve clarity.
- Earned best TA rating among all metrics in student evaluation, reflecting effectiveness in instruction.

Publications

1. C. Kao, W. Zhao, S. Revankar, S. Speas, S. Bhagat, **R. Datta**, C. Phoo, U. Mall, C. Vondrick, K. Bala, and B. Hariharan, "Towards LLM Agents for Earth Observation," in ICML 2025 TerraBytes Workshop, June 2025. [Online]. Available: https://arxiv.org/pdf/2504.12110

2. 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.

3. 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

- 12/2023. Presented "SustainGym: A Benchmark Suite of Reinforcement Learning for Sustainability Applications." NeurIPS 2023 Poster Session [poster]

- 5/2023. Presented "Semi-Automatic Digitization of Single-Line Diagramsfor Accelerated Deployment of Algorithmic Control on Real-World Power Systems." Meeting of the Minds [details]

- 10/2022. Presented "SustainGym: A Benchmark Suite of Reinforcement Learning for Sustainability Applications." NeurIPS 2022 Workshop On Tackling Climate Change With Machine Learning [talk]

- 8/2022. Presented "SustainGym: A Benchmark Suite of Reinforcement Learning for Sustainability Applications." SURF Summer Symposium 2022 - California Institute of Technology [slides]

- 9/2021. Presented "Utilizing spatial self-attention and perceptual loss functions to better reconstruct undersampled Magnetic Resonance Imaging (MRI)." SURF Summer Symposium 2021 - California Institute of Technology [slides]

Honors

- **Summer Undergraduate Research Fellowship 2022**: Award to pursue research at Caltech.

- **Summer Undergraduate Research Fellowship 2021**: Award to pursue research at Stanford.

- **National Merit Scholarship**: Scholarship for top 0.5% of graduating high school seniors

Projects

Symbolic Decision Trees with LLM Priors | *Python, Julia*   **April 2025 - May 2025**
- Built novel framework for constructing decision trees with symbolic expressions of semantic features.
- Initialized banks of candidate symbolic expressions using "common sense" knowledge of LLMs.
- Learned "symbolic trees" increased accuracy by an average of 8.5% over traditional decision trees.
- **Keywords:** *Large Language Models, Symbolic Regression, Interpretable Machine Learning*

GPU-accelerated Video Processing | *C++, CUDA*   **April 2024 - May 2024**
- Parallelized canny edge detection algorithm on video using CUDA, enabling up to 64x speed up.
- **Keywords:** *CUDA, Parallel Computing, Image Processing*

Semi-Automatic Single-Line Diagram Annotation | *Python* [details]   **April 2023 - June 2023**
- Proposed method to automate annotation of electric grid diagrams.
- Employed efficient template matching algorithm with non-max suppression (NMS).
- Reduced time for electric grid diagram annotation from days to seconds.
- **Keywords:** *Template Matching, Non-Max Suppression, Efficient Data Labeling*

LEGO Stacking Robotic Arm | *Python, Linux* [details]   **January 2023 - March 2023**
- Built a five-degree-of-freedom robotic arm to stack LEGOs from a workspace onto a designated grid.
- Assembled arm using HEBI motors and 3D-printed components (e.g., arm links and sockets).
- Developed an automated ROS 2 program for this task, running it on a local Linux machine.
- **Keywords:** *Robotics, Computer Vision, 3D Printing*

Rotating Tower of Hanoi Simulation | *Python* [code] ⬡   **September 2022 - December 2022**
- Developed a ROS package for simulating a 6-DOF robotic arm and a rotating table with equidistant poles.
- Programmed arm to move cubes 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.
- **Keywords:** *Robotics, Motion Planning, Numerical Methods*

Interactive Deep Learning Tutorial | *Python* [code] ⬡   **October 2019 - June 2020**
- Developed a web interface to visualize how deep neural networks operate.
- Worked with two biology high school teachers to incorporate tutorial in lessons.
- **Keywords:** *Machine Learning, Flask, Education*

Leadership / Extracurriculars

AfterMath   **September 2021 – June 2024**
Travel Captain   *California Institute of Technology*
- Responsible for collegiate ultimate frisbee team's travel logistics as a part of club leadership.
- Helped organize an overnight tournament in February 2023 with six colleges.