

Sara Vargas

NYC, NY 10009 • saravargasmar@gmail.com • +1 3324006110 • [LinkedIn](#) • [GitHub](#) • [Portfolio Web](#)

Education

New York University - Courant Institute School of Mathematics, Computing, and Data Science NY, NY
MSc in Computer Science May 2027

- **Relevant Coursework:** Natural Language Processing (Prof. Eunsol Choi), Deep Learning (Profs. Alfredo Canziani and Yann LeCun), Deep Learning, Machine Learning (Joan Bruna)

New York University - Courant Institute of Mathematics New York, NY
B.A. Honors Economics and Computer Science, Minor in Mathematics May 2025

- **Awards:** Dean's Undergraduate Research Fund Grant, The College of Arts and Sciences Service Grant, Dean's List

Technical Skills

- **Programming Languages:** Python, JavaScript, TypeScript, Java, C, R, Bash, HTML, CSS
- **ML/AI:** PyTorch, scikit-learn, NumPy, Matplotlib scikit-learn, NumPy, Matplotlib, transformer architectures, CNN/RNN models, LLM Evaluation, Prompt Engineering.
- **Web & APIs:** RESTful APIs, Node.js, React, Express, Vite.
- **Tools & Platforms:** Docker, Git, MongoDB, SQLAlchemy, HPC systems (Greene cluster), Kubernetes, CI/CD

Projects: [Portfolio](#)

Paper Trail: RAG System for Dataset Discovery (*NLP Course Project*) - Built a Retrieval-Augmented Generation system to enhance automated dataset description generation by extracting context from citing papers. Implemented Related Work Profile module to capture real-world dataset usage, limitations, and applications from scientific literature. Extended AutoDDG pipeline with literature-grounded context, improving description quality over baseline LLM approaches as measured by ROUGE-L, BERTScore, a reference free coverage score and LLM-as-judge evaluation.

Fine-tuning Language Models (BERT & T5) - Fine-tuned BERT for IMDB sentiment classification achieving **91%+ test accuracy** on NYU's Greene HPC cluster. Improved out-of-distribution robustness through data augmentation techniques. Fine-tuned T5-small for text-to-SQL translation on flight booking dataset, implementing end-to-end training and evaluation pipelines in PyTorch.

Experience

NYU Center for Data Science New York, NY
Machine Learning Research Assistant - Demining Group October 2025 – Present

- Lead research on AI-powered landmine risk prediction under Professor Mateo Dulce Rubio, designing and prototyping deep learning architectures in PyTorch to achieve robust out-of-distribution generalization when scaling across 1,103 Colombian municipalities
- Built end-to-end ML pipeline from automated data retrieval (ArcGIS Pro APIs) to model deployment, scaling RELand Project coverage from 15 municipalities to 1,103 while maintaining predictive accuracy on geographically diverse regions.
- Weekly team presentations on research progress translating technical model outputs into actionable policy recommendations for humanitarian demining organizations.

Octavate – Software Developer May 2025 – Aug 2025, New York, NY

- Built interactive UI features for an AI-powered music recommendation chatbot using React. Implemented interactive visuals for company website using TouchDesigner and p5.js

New York University - [Econometrics Research Honors Thesis](#) New York, NY Nov 2024 – May 2025

- Implemented an event study with linear and logistic regression models in R; performed hypothesis testing and robustness checks across demographic subgroups to study the effects of parental paid leaves on different demographic subgroups.
- Found larger post-policy increases in short-run labor-force attachment for mothers of color and unmarried mothers (about 5–20% relative gains, depending on specification).

Emerging Leaders in Technology and Engineering - Education Fellow Sep 2023 – Sep 2024, New York, NY

- Instructed weekly 5-hour sessions on Python programming for high school students from underrepresented backgrounds.
- Pedagogical strategies to enhance curriculum comprehension using p5 library to emphasize build visual creative portfolios.

NYU Gallatin - Student Technician Jan 2024 – Sep 2024, New York, NY

- Supported faculty and staff with AV system troubleshooting managing hardware/software requests. Collaborated with IT team on tech enhancement projects.

Barcelona Supercomputing Center Barcelona, Spain
Student Researcher Jan 2021 – Jun 2021

- Built Python-based models for large-scale computational analysis on HPC systems using Jupyter Notebooks

Leadership & Activities

Research+ Fellow at NYU CAS May 2024 – Sep 2024

- Selected for a competitive summer research fellowship in computational research at NYU.

NYU Undergraduate Research Conference Sep 2024 – May 2025

- Presenting undergraduate thesis on paid parental leave policy impacts.

Interests: Classical music training in harp and singing