Dan DeGenaro

drd92@georgetown.edu | ddegenaro.github.io | daniel-degenaro | ddegenaro | bU6sD_0AAAAJ
2308476598 | Dan-Degenaro | 0009-0005-1850-1801

Education

Georgetown University, PhD in Computer Science — Aug 2025 – Present
Georgetown University, MS in Computational Linguistics — Aug 2023 – May 2025

- GPA: 3.967/4.0
- **Coursework:** NLP, Multilingual NLP, Machine Learning, Deep Learning, Hypothesis Testing, Corpus Linguistics, Historical Linguistics, Databases

University of Massachusetts Amherst, BS in Physics — Sep 2019 – May 2023

- GPA: 4.0/4.0
- **Coursework:** Computational Physics, Modern Physics, Thermodynamics, Classical Mechanics, Statistical Physics, Electricity and Magnetism, Quantum Mechanics I-II

University of Massachusetts Amherst, BS in Applied Mathematics — Sep 2019 – May 2023

- GPA: 4.0/4.0
- **Coursework:** Calculus III, Linear Algebra I-II, Probability and Statistics I, Differential Equations, Chaos Theory, Numerical Analysis, Linear Optimization

University of Massachusetts Amherst, BA in Linguistics — Sep 2019 – May 2023

- GPA: 4.0/4.0
- **Coursework:** Phonetics, Phonology, Syntax I-II, Semantics, Sociolinguistics, Computational Linguistics I-II, NLP
- **Minor in CS:** Algorithm Design and Analysis, Discrete Math, Data Structures
- **Minor in Russian:** Central Asian Politics, Putin's Russia (Smith College), Russian I-IV

Research Experience

Visiting Research Scholar, Johns Hopkins University – Baltimore, MD — Jun 2024 – Aug 2024

- Participated in the Human Language Technology Center of Excellence SCALE 2024 workshop.
- Contributed to a multimodal information retrieval system designed to retrieve relevant videos given text queries.
- Developed a novel technique using downstream retrieval systems to produce preference rankings. Fine-tuned LLM using reinforcement learning to produce more retrievable document summaries.

Undergraduate Researcher, University of Colorado, Colorado Springs – Colorado Springs, CO — May 2022 – Aug 2022

- Developed a novel technique for the distillation of a multilingual BERT model into a smaller model.
- Developed original dataset, trained and fine-tuned a series of neural networks using PyTorch.
- Wrote and presented a research paper documenting methodology and results.

Teaching Experience

Project Course Instructor, MIT MITES Semester Program – Remote — Jun 2025 – Aug 2025, Jun 2024 – Aug 2024

- Designed and implemented, from scratch, a machine learning and NLP course for sophisticated high school students.
- Taught advanced machine learning concepts including deep learning architectures such as transformers to 18 students for 7 weeks.
- Guided students to produce 5 group projects using cutting-edge machine learning techniques.
- Re-hired to teach again for Summer 2025.

Graduate Teaching Fellow, Georgetown University – Washington, DC — Sep 2023 – May 2025

- Served as a teaching fellow four times:

1. Fall 2023: COSC 5580 (Introduction to Data Analytics), graduate course taught by Prof. Lisa Singh
2. Spring 2024: COSC 5470 (Deep Learning for Comp. Vis.), graduate course taught by Prof. Sarah Bargal
3. Fall 2024: COSC 5455 (Deep Learning), graduate course taught by Prof. Sarah Bargal
4. Spring 2025: COSC 3470 (Deep Learning), undergraduate course taught by Prof. Sarah Bargal
5. Fall 2025: COSC 3470 once again

- Developed, debugged, tested, and graded assignments in the form of Python notebooks.
- Designed a group project using a generative audio model.
- Lectured on PyTorch and Google Colab fundamentals.

Publications

DeGenaro, **D.**, Yang, E., Etter, D., Carpenter, C., Sanders, K., Martin, A., Murray, K., & Kriz, R. (2025, August). FORTIFY: Generative Model Fine-tuning with ORPO for ReTrieval Expansion of InFormal NoisY Text. In R. Kriz & K. Murray (Eds.), *Proceedings of the 1st Workshop on Multimodal Augmented Generation via Multimodal Retrieval (MAGMaR 2025)* (pp. 100–115). Association for Computational Linguistics. https://doi.org/10.18653/v1/2025.magmar-1.13

Samuel, S., **DeGenaro**, **D.**, Guallar-Blasco, J., Sanders, K., Eisape, S., Reddy, A., Martin, A., Yates, A., Yang, E., Carpenter, C., Etter, D., Kayi, E., Wiesner, M., Murray, K., & Kriz, R. (2025). MMMORRF: Multimodal Multilingual MOdularized Reciprocal Rank Fusion. *Proceedings of the 48th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 4004–4009. https://doi.org/10.1145/3726302.3730157

DeGenaro, **D.**, & Lupicki, T. (2024). Experiments in Mamba Sequence Modeling and NLLB-200 Fine-Tuning for Low Resource Multilingual Machine Translation. *Proceedings of the 4th Workshop on Natural Language Processing for Indigenous Languages of the Americas (AmericasNLP 2024)*, 188–194. https://doi.org/10.18653/v1/2024.americasnlp-1.22

Non-archival Conference Talks

DeGenaro, **D.** (2023). *Acting different(-ly): Bringing derivational morphology into variationist linguistics* [New Ways of Analyzing Variation]. https://nwav51.org/test/call-for-abstracts/

Invited Talks and Guest Lectures

GUCS Research Presentation 30 Sep 2025

- Delivered a 15-minute talk about my work, FORTIFY (**DeGenaro** et al., 2025) to Georgetown's CS department.

Guest lecture, Georgetown COSC 3470 18 Sep 2025

- Gave a lecture to students taking COSC 3470: Deep Learning (Undergraduate) at Georgetown University.
- Demonstrated basic usage of PyTorch in the Google Colab computing environment.

Guest demo, Georgetown InLab 11 Apr 2025

- Gave a demonstration of my Whisper-UI software to fellow graduate students.
- This demo took place at a regular meeting of the Georgetown Linguistics Department Interaction Lab, a meeting of students and faculty focused on discourse analysis.

Virtual lecture delivered at Walter Payton College Preparatory High School 21 Mar 2025

- Spoke to a group of high school students interested in linguistics.
- Lectured on n-grams and provided empirical demonstrations in Python.

Guest lecture, Georgetown COSC 3470 6 Feb 2025

- Gave a lecture to students taking COSC 3470: Deep Learning (Undergraduate) at Georgetown University.
- Demonstrated the training process for a simple convolutional neural network in PyTorch.

Guest demo, Georgetown Variation Lab 4 Feb 2025

- Gave a demonstration of my Whisper-UI software to fellow graduate students.
- This demo took place at a regular meeting of the Georgetown Linguistics Department Variation Lab, a meeting of students and faculty focused on sociolinguistics.

Guest lecture, Georgetown COSC 3470 30 Jan 2025

- Same lecture as 18 Sep 2025.

GLSA Workshop on LaTeX 20 Sep 2024

- Led a workshop for linguistics graduate students at Georgetown University.
- Explained basic LaTeX usage, including some specialized content geared towards writing linguistics papers.

Guest lecture, Georgetown COSC 5470 18 Sep 2024

- Gave a lecture to students taking COSC 5470: Deep Learning (Graduate) at Georgetown University.
- Demonstrated basic usage of PyTorch in the Google Colab computing environment.

GLSA Workshop on LaTeX 29 Sep 2023

- Led a workshop for linguistics graduate students at Georgetown University.
- Explained basic LaTeX usage, including some specialized content geared towards writing linguistics papers.

Advising

Keven Amaya-Muñoz, Arko Barua, and Luan Hoang, To appear. 2025

- Student project I advised, to be presented at MIT's Undergraduate Research Technology Conference (URTC) as a poster.
- Benchmarked pretrained and from-scratch computer vision systems on a tumor identification task.
- Qualitatively compared explainability results and quantitatively compared strength of out-of-domain transfer.

Melissa Alfaro-Zeledon and Rida Karim, Identifying Key Factors for Femicide Prevention and Policy Development: Leveraging Supervised Machine Learning with Temporal and Geospatial Analysis 2024

- Student project I advised, presented at URTC as a lightning talk.
- Used XGBoost to detect serial offenses in a database of femicides in the United States.

Hubert Hsu and Bethany Ray, MT-MOE: Protein-Specific Drug Design Utilizing Mixture of Experts Transformers 2024

- Student project I advised, presented at URTC as a poster.
- Performance gains in drug design observed by replacing traditional transformer architecture with MoE.

Awards

Graduate Student Teaching Assistant Award (Sciences) 2024-2025

- Awarded by Georgetown University.
- This award recognizes excellence among graduate students serving as TAs. Awards are given to one student from each area: humanities, social science and science and an at-large award.

Summa Cum Laude 2023

- Awarded by University of Massachusetts Amherst.
- Awarded for academic excellence (GPA in top 5%).

Commonwealth Honors College, greatest distinction 2023

- Awarded by University of Massachusetts Amherst.
- Awarded for academic excellence (completed undergraduate thesis with GPA > 3.8).

LeRoy F. Cook, Jr. Memorial Scholarship 2022

- Awarded by Department of Physics, University of Massachusetts Amherst.
- Awarded for academic excellence in physics and for engaging in teaching/tutoring as an undergraduate.

UMass Chancellor's Award 2019 – 2023

- Awarded by Department of Physics, University of Massachusetts Amherst.
- Awarded for academic excellence in physics and for engaging in teaching/tutoring as an undergraduate.

Thomas J. Watson Memorial Scholarship 2019 – 2023

- Awarded by IBM Thomas J. Watson Foundation.
- Awarded for academic excellence.

Dean's List 2019 – 2023

- Awarded by University of Massachusetts Amherst.
- Awarded for academic excellence (GPA > 3.5).

Projects

Whisper-UI ddegenaro.github.io/whisper-ui

- GUI for OpenAI's open-source Whisper ASR systems.
- Drag-and-drop interface intended for use by linguists with little coding background.

TEX Table pypi.org/project/tex-table

- Simple Python library to automate the formatting of LATEX tables, including marking p-value significance.

Volunteering and Service

Activities Coordinator, Georgetown University Computation and Language Group (GUCL) 2025-2026

- Organized social events for graduate students/faculty in linguistics and computer science.

Treasurer, Georgetown Graduate Linguistics Student Association 2024-2025

- Organized student-led workshops and professional/networking events.
- Managed organization's sources of funding, reimbursements, catering/supplies for events, etc.

Reviewing

ACM Multimedia (ACMMM) 2025

Hard Skills

Programming languages (Ordered by proficiency): Python, R, MySQL, Java, MATLAB, HTML/CSS, JavaScript, C, C++

Libraries (Ordered by proficiency): PyTorch (including TorchVision and TorchAudio), Jupyter, Pandas, HuggingFace Transformers, SciKit-Learn, XGBoost, NumPy, SciPy, MatPlotLib, SentencePiece, Plotly, spaCy, StatsModels, Librosa, fastText, TRL, Seaborn, NLTK, Gensim, Keras, TensorFlow, Stanza, Diffusers, PyTerrier, OpenCV

Technologies (Ordered by proficiency): VSCode, Overleaf/LATEX, Zotero, Git/GitHub, Slurm, Windows, Linux, RStudio, Docker, MacOS, Praat, ELAN, AWS EC2

Instructional Software (Ordered by proficiency): Google Classroom, Canvas, Moodle, Blackboard

Languages (Ordered by proficiency): Spanish, French, Russian, German