Joshua Caseiro DeOliveira

jcdeoliveira@wpi.edu | linkedin.com/in/joshua-caseiro-deoliveira | https://github.com/deoliveirajoshua

Education

Worcester Polytechnic Institute (WPI) | Worcester, MA

PhD, Data Science — In Progress

Advisor: Elke Rundensteiner, PhD

MS, Data Science — October 2023

Thesis: *Stabilizing GANs Under Limited Resources via Dynamic Machine Ordering*

BSc, Computer Science; BSc, Data Science — August 2022

Capstone (CS): *Building a Customizable GAN Package Tool in Python*

Capstone (DS): *Synthesizing Continuous and Discrete Mobile Sensor Data Using GAN models for Human Activity Recognition*

Graduation Honor: *High Distinction*

Research Experience

Research Intern - NLP, Adversarial Attacking — May – August 2023

Topologe LLC | Fairfax, VA

- Developed adversarial attacking algorithms against pre-trained, black-box, computer vision models.
- Developed product demo presented at Special Operations Command (SOCOM) Experimental Technologies Conference 2023
- Conducted exploratory research in novel feature engineering methods for identifying long-form LLM-generated text
- Presented the impact of this research at the Applied AI Challenge in LLMs hosted by the US Government's General Services Administration. Won the first-place award out of the 80+ companies and labs that entered.

Graduate Research Assistant — August 2022 – Present

Data Science Program, WPI | Worcester, MA

Advisor: Elke Rundensteiner, PhD

- Conducts research in the areas of generative modeling and deep learning optimization
- Member of the DAISY Lab and trainee on the NRT CEDAR Project
- Collaborates with fellow graduate researchers and faculty in data science-related research.
- Mentors undergraduate capstone projects in data science.

NSF-REU Summer Research Fellow — May – August 2021

Data Science Program, WPI | Worcester, MA

Advisors: Elke Rundensteiner, PhD; Walter Gerych, PhD Candidate

- Proposed a novel generative adversarial network (GAN) specifically designed for personalized, user-tailorable human activity recognition (HAR) data generation. This work allows for improving fairness and equity in downstream HAR tasks by up-sampling HAR datasets with generated data to better represent minority groups.

Undergrad Research Assistant: Open Set Library Classification in Embedded Firmware — January – May 2021

Department of Computer Science, WPI | Worcester, MA

Advisors: Robert Walls, PhD; Lorenzo DeCarli, PhD; Heshan Anupama, MS

- Investigated ways to analyze assembly instruction patterns to reverse engineer what libraries in a higher-level language were used in undocumented firmware.

Summer Research Assistant: Smile Detection — May – September 2019

Fuller Labs, Department of Computer Science, WPI | Worcester, MA

Advisors: Jacob Whitehill, Ph.D; Han Jiang, PhD Candidate

- Wrote scripts to scrape images from sit-coms and movies for downstream smile-detection tasks.
- Crowd-sourced the data labeling process with Amazon Mechanical Turk (MTURK).

Work Experience

Data Analyst November – December 2022
Strategic Initiatives and University Analytics, WPI | Worcester, MA

Data Science Intern - Perturbation Modeling May – August 2022
Topologe LLC | Fairfax, VA

- Developed adversarial attacking algorithms against pre-trained , white-box, computer vision models.
- Built a software-in-the-loop testbed to train and validate perturbation and adversarial attacking models for cyber physical systems in a simulated, 3D environment
- Presented a demo at the Special Operations Command (USASOC/SOCOM) Futures Forum 2022

Teaching and Mentoring Experience

Students Advised:

David Barsoum, Mass Academy	January - February 2023
Dillon McCarthy, BS, WPI	August 2022 - April 2023
Cindy Trac, BS, WPI	August 2022 - April 2023
Sirut Busai, BS, WPI	August 2022 - April 2023
Jason Dykstra, BS, WPI	August 2022 - April 2023
Aruzhan Koshkarova, BS, WPI	June - August 2022

Academic Outreach:

Accepted Student's Day Ambassador (Data Science @ WPI)	April 2022, 2023
Volunteer Judge (Data Science/Math), Mass Academy STEM Fair	Februrary 2023 - 2025
PyTorch: A Beginner's Workshop (held for students taking DS 3010)	December 2022
Open House Ambassador (Data Science @ WPI)	September 2022

Graduate Teaching Assistant August 2022 – Present
Data Science Program, WPI | Worcester, MA

MA 4635: Data Analytics and Statistical Learning
DS 502: Mathematical Foundations of Data Science
DS 3010: Computational Data Intelligence

Undergrad Peer Learning Assistant August 2020 – October 2021
Mathematical Sciences Department, WPI | Worcester, MA

MA 2621: Probability for Applications
MA 2612: Applied Statistics II
MA 2201: Discrete Math
MA 1024: Calculus IV (Vector Calculus)
MA 1021: Calculus I (Derivative Calculus)

Math Tutor August 2020 – October 2021
Mathematical Sciences Department, WPI | Worcester, MA

Teaching Assistant: Pre-Collegiate Outreach June 2019 – August 2020
Pre-Collegiate Outreach Program, WPI | Worcester, MA

Assisted in running a summer course for middle/high school students interested in computer science.

Honors and Awards

Applied AI Challenge in LLMs – First-place winner with Topologe LLC, General Service Administration	2023
Data Science Finalist, Graduate Research and Innovation Exchange @ WPI	2023
Best Overall Project (Data Science Division), Summer Research Showcase @ WPI	2022
Provost's Major Qualifying Project Award (Runner-Up), Computer Science Department @ WPI	2022

Outstanding Senior Award, Data Science Program @ WPI 2022
Best Poster Presentation, Undergraduate Data Science Research Symposium @ WPI 2021

Presentations

Applied AI Challenge in LLMs – Finalist Round 1^{st}-Place Winner May 2023
General Services Administration | Washington, DC

Co-presented with the CTO of Topologe LLC a product for identifying long-form LLM-generated text without architecture knowledge or API access. Black-box identification techniques allow for day-zero readiness to handle bad-actors that utilize generative models with potentially novel architectures from flooding digital forums with long-form fake news or spam.

Graduate Research Innovation Exchange April 2023
Worcester Polytechnic Institute | Worcester, MA

Presented to WPI faculty, staff, and external members my research in machine learning for improving mobile healthcare.

Poster on the Hill Presenter April 2022
Council of Undergraduate Research | Washington D.C

Invited to speak with the legislators Rep. Stephen Lynch (D) [MA-08] and Rep. Ayanna Pressley (D) [MA-07] and their staff about the applications of machine learning for improving mobile healthcare.

Technical Skills

Machine Learning: Generative modeling, adversarial attacking, deep learning theory, optimization
Mathematics: Linear algebra, vector calculus, group theory, causal inference
Computer Science: Algorithm analysis, high-performance computing
Programming: Python (PyTorch, PyTorch3D, AirSim), R, SQL, Java, C++
Software/Tools: LaTeX, Git, Slurm, Mathematica, Coq Proof Assistant, Unreal Engine

Professional Memberships

Upsilon Pi Epsilon (ΥΠΕ) - WPI Chapter
WPI Data Science Club

Publications

[1] **DeOliveira, J.**, Chakroborty, S., Gerych, W., Rundensteiner, E. *Towards Principled Methods for Data Attribution in Neural Nets.* Preparing for Submission.

[2] Hoffmann, D., VanNostrand, P., Ma, L., **DeOliveira, J.**, Cao, L., Rundensteiner, E. June 2025. *Agree to Disagree: Robust Anomaly Detection with Noisy Labels.* Oral Presentation, ACM SIGMOD, Berlin, Germany.

[3] **DeOliveira, J.**, Gerych, W., Rundensteiner, E., February 2025. *The Surprising Effectiveness of Infinite-Width NTKs for Characterizing and Improving Model Training.* Poster Presentation, Association of the Advancement of Artificial Intelligence (AAAI), Philadelphia, Pennsylvania.

[4] **DeOliveira, J.**, Gerych, W., Rundensteiner, E., December 2024. *GAN Stabilization Under Practical Training Assumptions.* Oral Presentation, IEEE International Conference on Big Data, Washington DC.

[5] **DeOliveira, J.**, Gerych, W., Koshkarova, A., Rundensteiner, E., Agu E., December 2022. *HAR-CTGAN: A Mobile Sensor Data Generation Tool for Human Activity Recognition.* Oral Presentation, IEEE International Conference on Big Data, 5th Special Session on HealthCare Data, Osaka, Japan.

[6] Gerych, W., Kim, H., **DeOliveira, J.**, Martin, M., Buquicchio, L., Chandrasekaran, K., Alajaji, A., Mansoor, H., Rundensteiner, E., Agu, E., December 2021. *GAN for Generating User-Specific Human Activity Data from an Incomplete Training Corpus.* Oral Presentation, IEEE International Conference on Big Data, 4th Special Session on HealthCare Data, Orlando, Florida.

[7] **DeOliveira, J.***, Kim, H.*, Martin, M.*, Gerych, W., Rundensteiner, E., October 2021. *Human Context Recognition: A Controllable GAN Approach.* Oral Presentation, MIT URTC, Cambridge, Massachusetts.

*Authors with * contributed equally.*