Neel Jain

neelsjain.github.io | njain17@umd.edu

RESEARCH OVERVIEW

My research is dedicated to addressing the multifaceted aspects of LLMs, pushing the boundaries of its capabilities, and ultimately enhancing LLMs. My areas of interest include alignment, safety, evaluation, and others.

EDUCATION

Ph.D, Computer Science — 2021 - Present
University of Maryland, College Park — College Park, MD

- Advisor: Prof. Tom Goldstein

M.S, Computer Science — 2021 - 2023
University of Maryland, College Park — College Park, MD

- GPA: 3.87; Advisor: Prof. Tom Goldstein

B.A, Honors in Mathematics — 2015 - 2019
Williams College — Williamstown, MA

- Thesis: Expanding Zero-forcing to Multi-color Forcing in Graphs

PUBLICATIONS AND PAPERS

Hard Prompts Made Easy: Gradient-Based Discrete Optimization for Prompt Tuning and Discovery, *NeurIPS 2023* — December 2023
Y Wen, N Jain, J Kirchenbauer, M Goldblum, J Geiping, T Goldstein

NEFTune: Noisy Embeddings Improve Instruction Finetuning, *Under Review* — October 2023
N. Jain, P. Chiang, Y. Wen, J. Kirchenbauer, H. Chu, G. Somepalli, B. Bartoldson, B. Kailkhura, A. Schwarzschild, A. Saha, M. Goldblum, J. Geiping, T. Goldstein

Baseline Defenses for Adversarial Attacks Against Aligned Language Models, *Under Review* — September 2023
N. Jain, A. Schwarzschild, Y. Wen, G. Somepalli, J. Kirchenbauer, P. Chiang, M. Goldblum, A. Saha, J. Geiping, T. Goldstein

Bring Your Own Data! Self-Supervised Evaluation for Large Language Models, *Under Review* — June 2023
N Jain, K Saifullah, Y Wen, J Kirchenbauer, M Shu, A. Saha, M Goldblum, J Geiping, T Goldstein

How to Do a Vocab Swap? A Study of Embedding Replacement for Pretrained Transformers, *Under Review* — September 2022
N Jain, J Kirchenbauer, J Geiping, T Goldstein

Multi-color Forcing in Graphs, *Springer: Graphs and Combinatorics* — June 2020
C Bozeman, PE Harris, N Jain, B Young, T Yu *(As most math papers, authors are alphabetically order)*

OTHER RESEARCH EXPERIENCE

Thesis, Williams College — September 2018 - May 2019
Graph Theory, Advisor Pamela Harris — Williamstown, MA

Research Intern, Salk Institute For Biological Studies — May 2017 - August 2017
Computational Biology, Edward Stites Lab — San Diego, CA

EMPLOYMENT

Research Assistant, University of Maryland, College Park — June 2023 - Present
Professor Tom Goldstein — College Park, MD

Teaching Assistant, University of Maryland, College Park — January 2023 - May 2023
Advanced Numerical Optimization, Professor Tom Goldstein — College Park, MD

Teaching Assistant, University of Maryland, College Park — September 2022 - December 2022
Advanced Data Structures, Professor Micheal Marsh — College Park, MD

Research Assistant, University of Maryland, College Park — June 2022 - August 2022
Professor Tom Goldstein — College Park, MD

- Explored techniques on faster adaptation of existing large language models to new languages, creating new foundational models. This work is currently under review.

Teaching Assistant, University of Maryland, College Park — September 2021 - May 2022
Introduction to Data Science, Professor John Dickerson and Jose Calderon — College Park, MD

Summer Math Tutor, Hamilton College Consulting — June 2020 - August 2020

- Tutored students for SAT/ACT math and other broad math skills; these students saw an increase by 300 points for the SAT and 5 points on the ACT math section

Data Scientist Senior Consultant, Booz Allen Hamilton — July 2020 - April 2021
Strategic Innovation Group, Analytics — Washington, DC

- Created math models such as agent-based models and simulations like Monte Carlo in python and excel for various different analyses and studies including program evaluations for DoD OSD CAPE in a research oriented approach to the problems
- Built a webapp using Flask alongside HTML, CSS, and JS to display various analyses of a curated dataset

Data Scientist Consultant, Booz Allen Hamilton — July 2019 - July 2020
Strategic Innovation Group, Analytics — Washington, DC

- Built an end-to-end audio analysis pipeline for an app in Dart using Tensorflow in Python
- Helped build a data pipeline from google trends to a S3 bucket that pulls every hour via a cron job for COVID-19 data lake

Summer Games Internship, Booz Allen Hamilton — June 2018 - August 2018
Strategic Innovation Group, Analytics — Washington, DC

- Analyzed spatial data through QGIS's python script runner to create shapefiles for the RShiny front-end
- Used R to clean data and create a RShiny front-end

Teaching Assistant, Williams College — September 2016 - December 2016
Introduction to Mechanics, Professor William Wootters — Williams College, Williamstown, MA

Internship, Anokiwave — July 2016 - August 2016
Silicon IC, Numerical Simulations — San Diego, CA

RELEVANT COURSE RESEARCH PROJECTS

Studying Human Interactions with LLMs in QA Settings for Exploring Human Trust in LLMs — September 2022 - December 2022
Course: Human-AI Interaction — College Park, MD

Hallucinations in Closed Book Generative Question Answering — January 2022 - May 2022
Course: How and Why Artificial Intelligence Answers Questions — College Park, MD

Universal Adversarial Attacks on Meta-Learning Algorithms — September 2021 - December 2021
Course: Foundations of Deep Learning — College Park, MD

TALKS, LEADERSHIP, AND CERTIFICATIONS

Co-Lead Machine Learning Reading Group at UMD	June 2021
Outstanding Graduate Teaching Assistant Award Recipient	January 2021
Dean's and Chair's Fellowship	September 2021
Moderated Panel on the Math Community for Minorities & the Application of Math for Social Good, Williams College	September 2020
Quantum Algorithms for Cybersecurity, Chemistry, and Optimization Certificate, MIT xPRO	April 2020
Introduction of Quantum Computing Certificate, MIT xPRO	February 2020
Foundations of Natural Language Processing Certificate, NVIDIA	December 2019
Foundations of Computer Vision Certificate, NVIDIA	October 2019
Men's Varsity Squash Team, Williams College	2015-2019
Minority Student Athlete Advisory Committee, Gaius C. Bolin Chapter, Williams College	2018-2019

Student Athlete Advisory Committee, Williams College 2016-2017

SOFTWARE LANGUAGES AND TOOLS

Python; Pytorch; Transformers; Pandas; Numpy; Scikit-Learn; NLTK; Spacy; Tensorflow; Keras; Docker; Java