Will Schwarzer

wschwarzer@umass.edu willschwarzer.com willschwarzer willschwarzer

Education

University of Massachusetts Amherst *Amherst, MA*
MS/PhD in Computer Science *September 2021 - Present*
GPA: 4.00/4.00

Carleton College *Northfield, MN*
B.A. in Computer Science and Mathematics, minors in Cognitive Science and Music Performance *September 2016 - June 2020*
GPA: 4.00/4.00 (Summa cum Laude)

Research Experience

Research Assistant *UMass, supported by Dolby*
Mentors: Xiaoyu Liu, Philip S. Thomas *February 2024 - Present*

- Studying the adversarial robustness of speech enhancement models, preparing for submission to ICLR or ICML 2025
- Will subsequently research either fairness and safety guarantees or reinforcement learning for generative audio models

Research Intern *Adobe, Inc.*
Mentors: Nikos Vlassis, Jennifer Healey *May 2023 - August 2023*

- Created and studied a novel method for training LLMs using arbitrary textual feedback, especially for use in RLAIF
- Demoed method at company-wide event and presented to research leadership
- Currently in preparation for submission

Research Assistant *UMass, supported by NSF*
Mentors: Philip S. Thomas, Scott Niekum, Bruno Castro da Silva *September 2021 - May 2023*

- Researched methods to allow the use of suboptimal demonstrations in inverse reinforcement learning
- Preparing for submission to AAAI or ICLR 2025

Computer Vision Engineer *Coros, Corp.*
Mentor: Stephen Hahn *January 2021 - August 2021*

- Optimized YOLO object detection models in PyTorch for automated parcel barcode scanning

Research Intern *Stanford University*
Mentors: Jesse Mu, Noah Goodman *June 2019 - August 2019*

- In PyTorch, developed a few-shot "concept captioning" network to output a natural language description of a set of images

Class Projects

2023	**Multimodal Robustness** Conducted survey on adversarial attacks against multimodal LLMs	*UMass Amherst*
2022	**Algorithms with Predictions** Developed supervised reward inference with worst-case guarantees	*UMass Amherst*
2022	**RL Baselines** Hand-implemented and evaluated foundational RL algorithms and environments	*UMass Amherst*
2021	**Constrained Optimization** Implemented constrained gradient descent using the KKT conditions	*UMass Amherst*

Prizes & Scholarships

2020	**Reeve Prize** Awarded to distinguished members of the senior class based on GPA	*Carleton College*
2019	**Goldwater Scholarship** National scholarship awarded each year to approximately 500 of the most promising STEM researchers nationwide; only 62 math/CS scholarships awarded in 2019	*US Government*
2019	**Phi Beta Kappa Second Year Prize** Awarded to the top student of the sophomore class	*Carleton College*
2019	**Damon Scholarship** Awarded to 10 juniors with strong academic profiles and moral character	*Carleton College*
2018	**Phi Beta Kappa First Year Prize** Awarded to the top student of the freshman class	*Carleton College*
2018	**Mortar Board Prize** Awarded to freshmen with a distinguished GPA (approx. top 5%)	*Carleton College*

Honors & Awards

2020	**CRA Outstanding Undergraduate Researcher Award (Honorable Mention)** Awarded to students who show outstanding potential in computing research	*Comp. Res. Assoc.*
2020	**Distinction in Computer Science** Awarded based on CS GPA and distinction in the senior project	*Carleton College*
2020	**Honors in Music Performance** Awarded for exceptional contribution to music at Carleton	*Carleton College*
2019	**Sigma Xi Membership** Offered to students having demonstrated aptitude for research	*Carleton College*
2019	**Phi Beta Kappa Membership** Inducted as a junior	*Carleton College*
2018	**Exemplary Writing Portfolio** Awarded unanimously by both readers; represents top 6-9%	*Carleton College*
2016-19	**Dean's List** Awarded to top 10% of each class by GPA (not awarded to seniors)	*Carleton College*
2015	**PSME Achievement Award** Nominated for and awarded simultaneously by two math professors	*Foothill College*

Skills

Technical skills Python (PyTorch, Jax, HuggingFace, various reinforcement learning libraries); HPC usage (Slurm); technical writing and presentations

Areas and Methods AI fairness and safety; adversarial robustness; speech enhancement; LLMs and RLHF; reinforcement learning; reward design, inverse reinforcement learning, and imitation learning

Natural Languages Fluent in Mandarin