Rishi D. Jha

INCOMING PHD STUDENT · SECURITY RESEARCHER

425.677.4846 | rjha01@cs.uw.edu | rishijha.com | rjha18 | rishi-jha

Education

Cornell
Ithaca, NY

INCOMING PHD STUDENT, COMPUTER SCIENCE — Aug. 2023 - Present

- Details Forthcoming!

University of Washington — Seattle
Seattle, WA

MS., COMPUTER SCIENCE — Sep. 2022 - Jun. 2023

- Master's Thesis: *Label Poisoning is All You Need*
- Advisor: Prof. Sewoong Oh
- GPA: 3.88 (4.0 @ PhD level) / 4.0

University of Washington — Seattle
Seattle, WA

BS.BA., COMPUTER SCIENCE AND MATHEMATICS — PHILOSOPHY: *Cum Laude, Phi Beta Kappa* — Sep. 2018 - Mar. 2022

- Jun. 2022: Graduated Cum Laude with Phi Beta Kappa honors
- 2018-22: Dean's List (all eligible quarters)
- GPA: 3.84 / 4.0

Selected Coursework

Machine Learning	Machine Learning‡, Deep Learning Theory †, Reinforcement Learning †, NLP, Deep Learning
Other Computer Science	Cryptography‡, Human-Centered AI†, Algorithms, Databases
Mathematics	Real Analysis I & II, Probability and Statistics I, II, & III, Modern Algebra I & II, Linear Algebra
Philosophy	Neuroethics

‡Taken at both the undergraduate and PhD levels.
†Taken at the PhD level.

Publications

MASTER'S THESIS

[1] **Rishi Jha**. "Label Poisoning is All You Need". University of Washington, Seattle, 2023.

WORKSHOP PAPERS

[2] Dimitrios C. Gklezakos, **Rishi Jha**, and Rajesh P.N. Rao. "Hyper-Universal Policy Approximation: Learning to Generate Actions from a Single Image using Hypernets". In: *Neurovision 2022: A CVPR Workshop*. New Orleans, USA: Conference on Computer Vision and Pattern Recognition, June 2022.

[3] **Rishi Jha** and Kai Mihata. "On Geodesic Distances and Contextual Embedding Compression for Text Classification". In: *Proceedings of the Fifteenth Workshop on Graph-Based Methods for Natural Language Processing (TextGraphs-15)*. Mexico City, Mexico: Association for Computational Linguistics, June 2021, pp. 144–149.

PATENTS (PENDING)

[4] Nisha S. Hameed, **Rishi D. Jha**, and Evan Argyle. "Graph-Based Analysis of Security Incidents". U.S. pat. Microsoft.

Academic Research

Sewoong Lab — Foundations of Machine Learning
Seattle, WA

GRADUATE RESEARCH ASSISTANT — May 2021 – Aug 2023

Working with **Prof. Sewoong Oh** and **Jonathan Hayase** to:

- (Master's Thesis Project) Develop a novel trajectory-matching-based backdoor attack, FLIP, that corrupts (i.e., 'poisons') only the labels in a training set to create a backdoor with an arbitrary trigger. In particular, we show that with few-shot poisons (i.e., less than 1% of a dataset's training labels), FLIP can inject a backdoor with a 99.6% success rate while remaining undetected with less than a 1% degradation of clean accuracy. We also demonstrate FLIP's surprising robustness to dataset, trigger, and architecture. Thesis submitted in **June 2023**. Paper in review at **NeurIPS 2023**.
- (*Previously*) Create an open-source 'backdoor'-attack-benchmark platform and survey for robust machine learning algorithms. Code can be found **here**.

Center for Neurotechnology

Seattle, WA

UNDERGRADUATE ML RESEARCHER

Mar. 2020 – Aug. 2022

Paper accepted at **NeuroVision '22 at CVPR** [2]. Worked with **Prof. Rajesh Rao** and **Dimitrios Gklezakos** to:
- Develop a low-cost, 'personalized' hypernetwork for hierarchical and task-conditional RL called the Hyper-Universal Policy Approximator (HUPA). HUPAs are up to 35% more resilient to sparsity and have up to 25% better generalization than their traditional embedding alternatives.
- Construct an audio-visual hypernetwork for representation learning and classification on a massive dataset in which a video-controlled neural network controls the weights of an audio interpreter.
- Create a convolutional, manifold-learning based network to learn complex features in natural images in an unsupervised fashion using sparse coding. The system learns representational similarities between features and generalizes them.

Self-Directed

Seattle, WA

NLP RESEARCHER

Nov. 2020 – Jun. 2021

Paper accepted at **TextGraphs '21 at NAACL** [3]. Worked with **Kai Mihata** to:
- Investigate the downstream effects of compressing BERT embeddings using nonlinear dimensionality reduction techniques and geodesic estimations.
- Find that nonlinear compressions of the embeddings tend to work well in some data regimes, a feature that can be utilized in memory-constrained settings.

ICTD Lab

Seattle, WA

UNDERGRADUATE RESEARCHER

Nov. 2018 - May 2019

Worked with **Spencer Sevilla** to:
- Investigate the performance dynamics of different chat apps in poor network conditions.
- Implement a teaching solution for schoolchildren in rural Indonesia.

Research in Industry

Microsoft Defender Research

Redmond, WA

SOFTWARE ENGINEERING INTERN — DATA SCIENCE

Jul. 2022 - Sep. 2022

- Ideated, pitched, and implemented a low-cost, humanly interpretable meta-learning framework that exploits spectral similarities in existing classifier responses to drive robustness in the Defender product. The productionalized system was lightweight, had upwards of 97% precision and recall, and was humanly interpretable.
- The model is being pushed from pre-production to production and will start providing protection for billions of users by **Summer 2023**.

Microsoft Defender Research

Remote

SOFTWARE ENGINEERING INTERN — DATA SCIENCE

Jun. 2021 - Sep. 2021

Patent submitted in **Winter 2022** [4].
- Ideated and designed patent-pending approach to detect malicious Command-and-Control intrusions in corporate networks using spectral methods on graphs. The model achieved high precision and recall in finding Indicators of Compromise in historical data.
- The project has received significant investment from the team and Microsoft Research (MSR) since my departure with a goal of pushing an extension of the model to production in **Summer 2023**.

Teaching

University of Washington — Seattle

Seattle, WA

4X UNDERGRAD / GRAD MACHINE LEARNING TA

Mar. 2020 - Dec. 2021

During Spring 2020, Winter 2021, Spring 2021, Autumn 2021:
- Taught undergraduate and graduate students as an undergraduate through 25-person sections and biweekly office hours.
- Designed section materials for entire teaching staff, monitored discussion boards, and graded assignments.

University of Washington — Seattle

Seattle, WA

MACHINE LEARNING COURSE DESIGNER

Jun. 2021 - Sep. 2021

During Summer 2021, funded by **Dr. Sewoong Oh** to:
- Redesign the course's problem sets and homework infrastructure to keep up with a rapidly evolving course and field, and lower the barrier of entry to machine learning.
- Drive equitability by adding necessary data context, removing technical jargon, and constructing homework problems that required students to challenge algorithmic and implicit biases in machine learning.
- Create a new central grading system and TA codebase for future quarters and course staffs to use.

Other Work Experience

Microsoft   *Remote*
SOFTWARE ENGINEERING INTERN — DEFENDER SECURITY   *Jun. 2020 - Sep. 2020*
- Reduced related COGS by $100K - $1M by creating ML model to selectively download dangerous files for analysis. In production.
- Built infrastructure for safer ML model deployment. In production.
- Decreased researcher rule development time by 35%, by creating VSCode extension to natively test rules. In production.

Microsoft   *Redmond, WA*
EXPLORE INTERN — OFFICE.COM FRONT END   *Jun. 2019 – Aug. 2019*
- Designed, implemented, and released front end notes tool for the Office.com team using Typescript, Redux, and React internally.

Honors

2022	**Appointed**, Phi Beta Kappa	*Seattle, WA*
2022	**Appointed**, Cum Laude Scholar	*Seattle, WA*
2018-22	**Selected**, Dean's List (all eligible quarters)	*Seattle, WA*
2021-22	**Selected**, Varsity Climbing Team at UW	*Seattle, WA*
2020	**1ˢᵗ Place**, Rain City Send Bouldering Competition — Recreational Category	*Seattle, WA*
2019	**Finalist**, (Top 4 of 36 Teams) UW Foster CBDC: Consulting Challenge	*Seattle, WA*
2018	**Appointed**, National Merit Scholar	*Redmond, WA*
2017	**3ʳᵈ Place**, (1000+ Teams) Microsoft OneWeek Hackathon Consumer Category	*Redmond, WA*

Skills

Interests	Machine Learning, Robustness, Security, Privacy, Anomaly Detection, Graph Theory
Technical	Python, PyTorch, TensorFlow, JAX, C++, Java / C#,
Languages	English, Hindi, Spanish

Service

2023	**Reviewer**, ICML	*Remote*
2021	**Presenter**, High School Neuroscience Club @ The Overlake School	*Redmond, WA*