Rishi D. **Jha**

MASTER'S STUDENT · SECURITY RESEARCHER

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

"Master's student, research assistant, and former computer science and mathematics double major. Interests include the fundamentals of robustness, security, privacy, and fairness in ML. Motivated by hard problems."

Education

University of Washington — Seattle *Seattle, WA*

MS., COMPUTER SCIENCE *Sep. 2022 - Present*

- GPA: 4.0 / 4.0
- Graduate Research Assistant with Dr. Sewoong Oh

University of Washington — Seattle *Seattle, WA*

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

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

Selected Coursework

Machine Learning	Machine Learning[‡], Deep Learning Theory [†], Reinforcement Learning [†], NLP, Deep Learning
Other Computer Science	Randomized Algorithms[*], 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.
[*] Planned at the PhD level.

Publications

WORKSHOP PAPERS

[1] 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.

[2] **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)

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

Academic Research

Sewoong Lab — Theoretical Machine Learning and Robustness *Seattle, WA*

GRADUATE RESEARCH ASSISTANT *May 2021 – Present*

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

- Develop a novel Neural Tangent Kernel (NTK)-based backdoor attack that persists through the knowledge distillation process and infects networks with triggers they have never seen. The attack uses NTK-ized linear regression to find labels for a victim-controlled distillation set that minimize the squared loss on the attacker-controlled training set. At evaluation time, the triggers fool the victim network 70% of the time. Planned submission to **ICML 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** [1]. Worked with **Dr. Rajesh Rao** and **Dr. 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. Planned full conference submission in **Winter 2023**.
- 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** [2]. 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 **Dr. 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** [3].

- 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