Wonkwon Raymond Lee

111 Sylvan Ave, Englewood Cliffs, NJ 07632 646-469-7805 wonkwon.lee@nyu.edu wonkwonlee.github.io

Research Interests

My research interests focus on developing secure, robust, interpretable, and fair AI systems, with a particular emphasis on privacy-preserving techniques, out-of-distribution resilience, and interpretability. In my previous work, I have specialized in differentially private synthetic data generation and evaluating AI systems for privacy, utility, and reproducibility. I am also passionate about addressing uncertainty and alignment challenges, aiming to integrate AI with human values to create safe and trustworthy applications, especially in critical sectors such as healthcare, finance, and criminal justice.

Education

New York University **09/2021 - 05/2023**
Computer Science Master of Science

Advisor: **Julia Stoyanovich**
Coursework: Computer Vision, Natural Language Processing, Responsible Data Science, Data Science for Healthcare, Big Data, Advanced Database Systems

University of Manchester **09/2015 - 06/2018**
Computer Science and Mathematics Bachelor of Science

Advisor: **Eva M. Navarro-Lopez, Martin Lotz**
Thesis: *Models of Neurons and Neuronal Networks*
Coursework: Machine Learning, Convex Optimization, Linear Algebra, Partial Differential Equations, Complex Analysis, Image Processing, Cryptography, Algebraic Structures

Research Experience

Center for Responsible AI, New York University **09/2022 - 05/2023**
Graduate Research Assistant New York, NY

- Conducted research under Professor Julia Stoyanovich on evaluating **differentially private (DP)** synthetic data generation methods.
- Developed "Epistemic Parity," an evaluation metric based on the likelihood of reproducibility of quantitative claims in social science research.
- Created **SynRD**, an open-source DP synthetic data benchmarking Python package that organizes the Epistemic Parity workflow, existing papers, and datasets.

McDevitt Lab, New York University **10/2021 - 02/2022**
Graduate Research Assistant New York, NY

- Performed diagnostic prediction modeling research for the Colgate Project under Professor John T. McDevitt, utilizing machine learning and statistical methods for data analysis.
- Preprocessed and visualized complex unstructured biomarker data from microfluidic sensors using SQL, Stata, R, Pandas, and Seaborn.
- Conducted a meta-analysis to combine and analyze data from multiple sources by extracting semantics.

Independent Research, University of Manchester **09/2017 – 06/2018**
Undergraduate Research Assistant Manchester, UK

- Implemented a spiking neural network simulator using Python, QtPy5, Brian2, and Neurodynex to investigate neuromorphic computing paradigms inspired by biological neural systems.
- Simulated and analyzed dynamical synchronization patterns influenced by network topology and external stimuli.
- Conducted research under Dr. Eva Lopez, culminating in the thesis "Models of Neurons and Neuronal Networks," which received the Project Excellence Award.

Wireless Intelligence at Network Edge Lab, Korea University **06/2016 – 08/2016**
Undergraduate Research Intern Seoul, South Korea

- Worked on an IoT Drone project under the supervision of Professor Hwangnam Kim as a Summer Undergraduate Research Intern.
- Developed and implemented new functionalities in MATLAB to optimize real-time simulation of networked drone fleets.

Industry Experience

LG CNS America — **04/2024 - Present**
System Engineer — Englewood Cliffs, NJ

- Designed and implemented network infrastructure to enhance system performance and security, collaborating with cross-functional teams to troubleshoot and resolve complex networking issues.
- Implemented automated network monitoring and reporting systems to ensure optimal uptime and reliability using **Python**, **Netmiko**, and **PRTG API**.
- Created comprehensive network documentation, including diagrams, operational procedures, and troubleshooting guides, to facilitate knowledge sharing and system maintenance.

Stealth Project (EPLIA) — **01/2023 – 01/2024**
Co-founder / CTO — San Francisco, CA

- Co-founded a healthcare startup aimed at improving accessibility by addressing language barriers in telemedicine.
- Led the design and development of a web application using **Next.js**, **AWS** cloud infrastructure, and **WebRTC** for real-time communication.
- Managed cross-functional collaboration to deliver a scalable, reliable platform tailored to the unique needs of diverse users.

Pricewaterhouse Coopers — **06/2022 - 08/2022**
Data Scientist Intern — New York, NY

- Implemented and fine-tuned a BERT model to classify semantic relationships between entities using **PyTorch**.
- Designed data annotation protocols and ML pipelines from data, training, to deployment; deployed the models to **AWS** for scalable production use.

Publications

Epistemic Parity: Reproducibility as an Evaluation Metric for Differential Privacy — **05/2024**

Rosenblatt, L., Herman, B., Holovenko, A., **Lee, W.**, Loftus, J., McKinnie, E., ... & Stoyanovich, J. (2024). Epistemic Parity: Reproducibility as an Evaluation Metric for Differential Privacy. *ACM SIGMOD Record*, *53*(1), 65-74.

Out of distribution performance of state of art vision model — **01/2023**

Rahman, S., & **Lee, W**. (2023). Out of distribution performance of state of art vision model. *arXiv preprint arXiv:2301.10750*.

Epistemic Parity: Reproducibility as an Evaluation Metric for Differential Privacy — **08/2023**

Rosenblatt, L., Herman, B., Holovenko, A., **Lee, W.**, Loftus, J., McKinnie, E., ... & Stoyanovich, J. (2023). Epistemic Parity: Reproducibility as an Evaluation Metric for Differential Privacy. *Proceedings of the VLDB Endowment*, *16*(11), 3178-3191.

Awards

SIGMOD Research Highlight Awards — **06/2024**
ACM SIGMOD 2024

Best Experiment, Analysis & Benchmark Paper Runner-up — **08/2023**
VLDB 2023

Wasserman Center Career Grant — **11/2021**
New York University

Landslide Prediction Modeling Contest — **09/2021**
Korea Meteorological Administration

Computer Science Final-Year Project Award — **07/2018**
University of Manchester

International Mathematical Excellence Scholarship — **09/2015 - 09/2017**
University of Manchester

Projects

Time-series Medical Image Classification 01/2023 – 05/2023

- Developed a time-series classification model to predict disease progression from multi-image chest X-rays by fine-tuning pre-trained **DenseNet121** and **Vision Transformer** models on the MS-CXR-T dataset.

Out-of-Distribution Robustness Evaluation Of Vision Models 09/2022 - 01/2023

- Conducted **out-of-distribution robustness** comparison of **58** computer vision models, including ViT, convolution, attention-convolution hybrid, sequence-, and network-based, using OOD benchmark datasets to assess performance under distribution shifts.

Privacy-Preserving Synthetic Data Generation 03/2022 – 05/2022

- Generated privacy-preserving synthetic datasets using DataSynthesizer and MST, analyzing statistical accuracy, KL-divergence, and mutual information across varying epsilon values.

Model Explanations with SHAP 03/2022 – 05/2022

- Used SHAP to analyze classifier predictions on the 20 Newsgroups dataset, proposing a feature selection strategy that improved classifier accuracy by 2%.

Collaborative-Filter Based Recommender System 02/2022 – 05/2022

- Implemented a collaborative-filter-based movie recommender system using **PySpark**'s alternating least square method and achieved mean average precision of **0.066**

Fairness-Enhancing Interventions in ML Pipelines 01/2022 – 03/2022

- Implemented fairness-enhancing algorithms (pre-, in-, and post-processing) with AIF360, evaluating fairness-performance trade-offs using Folktables ACSIncome dataset.

Landslide Prediction Modeling 04/2021 – 08/2021

- Preprocessed **GIS** and **time-series** climate data and implemented XGBoost and LightGBM models using **TensorFlow** and won 6th out of 150 teams in a national data science competition. **(top 4%)**

Spiking Neural Network Simulatation 09/2017 – 05/2018

- Developed a Python GUI simulator to analyze spiking neural network dynamics, firing patterns, and synchronization properties using libraries like Brian2 and PyDSTool.

Skills

Programming Languages

Python, Java, C/C++, R, MATLAB, JavaScript, SQL

Frameworks and Libraries

PyTorch, TensorFlow, scikit-learn, Pandas, Numpy, SciPy, Flask, Django, Apache Spark

Tools and Methodologies

Jupyter Notebooks, Git/GitHub, Docker, AWS, IBM Cloud, LaTeX

Languages

Korean, English, Japanese, Spanish

References

Julia Stoyanovich
Associate Professor, New York University
stoyanovich@nyu.edu

Martin Lotz
Associate Professor, University of Warwick
martin.lotz@warwick.ac.uk

Ian Pratt-Hartmann
Senior Lecturer, University of Manchester
ipratt@cs.man.ac.uk