Wonkwon Lee

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

Research Interests

My research interests focus on building robust, safe, and interpretable AI systems, with a particular emphasis on their applications in healthcare. I am passionate about exploring out-of-distribution robustness in vision models, ensuring that AI systems can maintain performance in unpredictable environments. My past work includes differentially private synthetic data generation and benchmarking AI systems for privacy, utility, and reproducibility. Additionally, I am interested in applying causal inference and reasoning to AI systems to address alignment challenges, ensuring AI can understand human intentions and provide reliable, human-aligned decision-making.

Education

New York University, Courant Institute **09/2021 - 05/2023**
Computer Science MS

Advisor: Professor **Julia Stoyanovich**
Affiliate: Center for Responsible AI
Publication: *Epistemic Parity: Reproducibility as an Evaluation Metric for Differential Privacy*
Coursework: Computer Vision, Natural Language Processing, Responsible Data Science, Data Science for Healthcare

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

Advisor: **Eva M. Navarro-Lopez**
Publication: *Models of Neurons and Neuronal Networks*
Coursework: Convex Optimization, Linear Algebra, Partial Differential Equations, Real and Complex Analysis, Image Processing, Cryptography

Research Experience

Center for Responsible AI, NYU **September 2022 - May 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, NYU **October 2021 - February 2022**
Graduate Research Assistant New York, NY

- Performed **predictive diagnostic** research under Professor John T. McDevitt and conducted statistical and meta-analysis.
- Preprocessed and visualized biomarker data from microfluidic chip sensors using R, Matplotlib, and Pandas.

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.

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.
- Monitored and maintained network systems to ensure optimal uptime and reliability

Stealth Project (EPLIA) — **01/2023 – 12/2023**
Co-founder / Software Engineer — San Francisco, CA

- Co-founded a startup developing a telemedicine app that connects patients with doctors fluent in their native language, enhancing accessibility and patient care.
- Designed and implemented core application features using **Next.js** for the front-end, **AWS** for cloud services, and **WebRTC** for real-time communication.

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

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

Projects

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.

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.

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**

Landslide Prediction Modeling — **04/2021 – 08/2021**

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

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

KMA 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

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
Native	Fluent	Fluent

References

Julia Stoyanovich
Associate Professor, Computer Science and Engineering
https://stoyanovich.org

New York University
370 Jay Street
Brooklyn, NY 11201
stoyanovich@nyu.edu

Martin Lotz
Associate Professor, Mathematics
https://homepages.warwick.ac.uk/staff/Martin.Lotz

Zeeman Building
Coventry CV4 7AL
United Kingdom
martin.lotz@warwick.ac.uk

Rizos Sakellariou
Professor, Department of Computer Science
http://www.cs.man.ac.uk/~rizos/

University of Manchester
Oxford Road
Manchester M13 9PL
U.K.
rizos@manchester.ac.uk