PASAN DISSANAYAKE

✉ pasand@umd.edu   🔗 linkedin.com/in/pasandissanayake   🔗 pasandissanayake.github.io

RESEARCH INTERESTS

My research focuses on efficient and trustworthy machine learning, with an emphasis on knowledge distillation techniques to democratize reliable AI. I aim to extract and transfer the capabilities of large models into compact ones by utilizing the interpretability/explainability of the models and the transferring process, leveraging tools from optimization, information theory, and statistics.

EDUCATION

University of Maryland, College Park, MD   Started Aug. 2022
PhD in Electrical and Computer Engineering (CGPA: 3.96/4.0)
Advised by Prof. Sanghamitra Dutta
Keywords: Efficient & Trustworthy AI, Knowledge Distillation, Information Theory, Optimization

University of Moratuwa, Sri Lanka   Graduated Aug. 2021
BSc (Hons.) in Electronic and Telecommunication Engineering (GPA: 4.04/4.2)
Advised by Dr. Prathapasinghe Dharmawansa Dr. Ranga Rodrigo, Dr. Peshala Jayasekara
Thesis: Precise vehicle localization using fusion of multiple sensors for self-driving

PUBLICATIONS

- **P. Dissanayake**, F. Hamman, B. Halder, I. Sucholutsky, Q. Zhang, S. Dutta, "Quantifying Knowledge Distillation using Partial Information Decomposition," Artificial Intelligence and Statistics (AISTATS), 2025
- **P. Dissanayake** and S. Dutta, "Model Reconstruction Using Counterfactual Explanations: A Perspective From Polytope Theory", Neural Information Processing Systems (NeurIPS), 2024
- **P. Dissanayake**, P. Dharmawansa and Y. Chen, "Distribution of the Scaled Condition Number of Single-Spiked Complex Wishart Matrices," in IEEE Transactions on Information Theory, vol. 68, no. 10, pp. 6716-6737, 2022
- F. Hamman, **P. Dissanayake**, S. Mishra, F. Lecue, S. Dutta, "Quantifying Prediction Consistency Under Model Multiplicity in Tabular LLMs", International Conference on Machine Learning (ICML), 2025
- S. Meel, **P. Dissanayake**, M. Nomeir, S. Dutta, S. Ulukus, "Private Counterfactual Retrieval With Immutable Features", IEEE International Symposium on Information Theory (ISIT), 2025
- E. Noorani, **P. Dissanayake**, F. Hamman, S. Dutta, "Counterfactual Explanations for Model Ensembles Using Entropic Risk Measures" International Conference on Autonomous Agents and Multiagent Systems (AAMAS), 2025.
- B. Halder, F. Hamman, **P. Dissanayake**, Q. Zhang, I. Sucholutsky, S. Dutta, "Quantifying Spuriousness of Biased Datasets Using Partial Information Decomposition," Data-centric Machine Learning Research workshop at ICML, 2024
- P. Dharmawansa, **P. Dissanayake** and Y. Chen, "The Eigenvectors of Single-Spiked Complex Wishart Matrices: Finite and Asymptotic Analyses," in IEEE Transactions on Information Theory, vol. 68, no. 12, pp. 8092-8120, 2022
- M. Nomeir, **P. Dissanayake**, S. Meel, S. Dutta, S. Ulukus, "Private Counterfactual Retrieval" (under review)

RESEARCH PROJECTS

Task-aware distillation from multiple teacher LLMs *(On-going)*
The goal is to improve the performance of the student by leveraging the strengths of multiple teacher LLMs. The task-awareness is expected to improve robustness to adversarial/irrelevant knowledge in the teachers.

Extreme distillation: Distilling LLMs into MLPs *(On-going)*
Investigating the feasibility of distilling large language models (LLMs) into task-specific multi-layer perceptrons (MLPs) within a meta-learning setting. The goal is to yield the wold knowledge of the LLM for inferring parameters for classical ML models with only a few shots. Experiments are being conducted on tabular datasets using PyTorch.

Quantifying task-related knowledge in knowledge distillation
Introduced task-aware metrics to measure knowledge available to distill and already distilled information based on Partial Information Decomposition (PID). Conducted experiments on CIFAR-10/100 and CUB-200-2011 datasets using PyTorch (on WideResNet and ResNet models). [AISTATS 2025]

Counterfactual explanations and model extraction attacks
Investigated the use of counterfactual explanations for model extraction attacks under asymmetric query access (i.e., explanations only from one side of the decision boundary). Implemented attacks on tabular datasets using TensorFlow. Moreover, developed private retrieval mechanisms for counterfactuals within a PIR setting. [NeurIPS 2024, ISIT 2025, AAMAS 2025]

Quantifying prediction consistency of tabular LLMs under fine-tuning multiplicity
Developed a metric to efficiently quantify model multiplicity of LLMs fine-tuned on tabular datasets. Experiments were done with BigScience T0 and Google FLAN T5 models from Huggingface. LORA and T-Few methods were used as the fine-tuning techniques. [ICML 2025]

Single-spiked complex Wishart matrices
Studied the properties of two important statistics of the complex Wishart ensemble, with a focus on the centralized (zero-mean) case with a single-spiked covariance structure. Simulations were done using Mathematica. [IEEE Trans. Info. Theory, 2022]

SKILLS

Programming: Python, C++, Java, MATLAB, Mathematica
Data/ML Frameworks: Tensorflow, PyTorch, Keras, Scikit-Learn, Pandas, Weights&Biases
Other Tools: Git, Docker, ROS (Robot Operating System), LaTeX

AWARDS AND HONORS

- Outstanding Graduate Assistant (2024) - Graduate School, University of Maryland
- Outstanding Teaching Assistant (2025, 2024) - ECE, University of Maryland
- Dean's Fellowship (2022) - University of Maryland
- Dean's List (in all 8 semesters) - University of Moratuwa
- Google Hash Code (2019) Country rank: 3

CERTIFICATIONS

- Reinforcement Fine-Tuning LLMs With GRPO - Predibase on Deeplearning.ai
- Fundamentals of AI Agents Using RAG and LangChain - IBM on Coursera
- Machine Learning - Stanford University on Coursera
- Data Structures, Algorithmic Toolbox - University of California, San Diego on Coursera
- Certificate in Software Development in Java - National Institute of Business Management, SL
- Networking Essentials (NE-CS3032-IN-16-EN) - Offered by Cisco Networking Academy

RELEVANT COURSEWORK

- ENEE633 - Statistical Pattern Recognition (Spring 2025)
- CMSC858A - Concentration Inequalities for Randomized Algorithms and ML (Fall 2024)
- CMSC764 - Advanced Numerical Optimization (Spring 2024)
- CMSC742 - Algorithms in Machine Learning: Guarantees and Analyses (Fall 2023)
- ENEE627 - Information Theory (Spring 2023)
- ENEE620 - Random Processes in Communication and Control (Fall 2022)

EMPLOYMENT

Graduate Assistant — Aug. 2022 to Present
Department of Electrical and Computer Engineering,
University of Maryland, College Park, MD

Lecturer (on Contract) — Dec. 2021 to Aug. 2022
Department of Electronic and Telecommunication Engineering,
University of Moratuwa, Sri Lanka

Systems Engineer (Intern) — Jun. 2019 to Dec. 2019
Infrastructure Solutions Team
MillenniumIT ESP (Pvt) Ltd., Colombo 07, Sri Lanka

TEACHING EXPERIENCE

University of Maryland, College Park, MD ENEE621: Estimation and Detection Theory (Spring 2025); ENEE420: Communication Systems (Fall 2022); ENEE324: Engineering Probability (Spring 2023, Fall 2023, Spring 2025); ENEE436: Foundations of Machine Learning (Fall 2023)

University of Moratuwa, Sri Lanka EN4720: Security in Cyber-Physical Systems; EN2550: Fundamentals of Image Processing and Machine Vision; EN3053: Digital Communications - I

SERVICES AND EXTRA-CURRICULAR

- Reviewer - NeurIPS 2025; ICLR 2025; XAIFIN 2024; ISIT 2025, 2024; FAccT 2025, 2024; IEEE Wireless Communications Letters (*view on Publons*)
- Project Lead - UMD Trustworthy AI in Law and Society (TRAILS) AI Summer Camp 2024
- Member of the panel of judges - UMD Louis Stokes Alliance for Minority Participation (LSAMP) Fall Research Symposium 2023
- Mentor - Montgomery Blair High School Magnet Student Program 2023
- Secretary - Drama Society, University of Moratuwa for the year 2020/2021
- Co-chairperson - Organizing committee of "Abhina 2019", the annual talent show of the Dept. of Electronic and Telecommunication Engineering
- Volunteer - Soyuru Sathkaraya 2018, E-Care 2018, Grama Prabodhaya 2017: underprivileged school development projects organized by the student associations of University of Moratuwa