Simon Dovan Nguyen

LinkedIn | Website | Github | simondn@uw.edu

EDUCATION

University of Washington, Seattle (expected) Jun. 2027
Ph.D. in Statistics. Advised by: Tyler H. McCormick

University of Michigan, Ann Arbor Apr. 2023
M.S. in Applied Statistics. Advised by: Ben B Hansen and Daniel Almirall

University of California, Irvine Jun. 2021
B.S. in Mathematics (Honors); B.A. in Economics (Honors); Minor in Statistics

CENTER FOR NAVAL ANALYSIS (Data Science Division) | Data Scientist Intern Jun. 2025 - Sep. 2025

Predictive Aircraft Readiness and Maintenance

- Developed ML models optimizing plane repair & delivery processes, increasing mission-capable F18 Super Hornets to 80%.
- Leveraged NLP to extract latent themes from maintenance logs, enabling earlier detection of supply risks by 12%
- Deployed predictive algorithms to identify early-warning signals for supply shortages, preventing naval operational shortfalls.

LLM-Powered Tactical Advisor for Specialized Wargaming

- Engineered an LLM advisor with a Retrieval Augmented Generation architecture to overcome wargaming expert shortages.
- Implemented RAG to dynamically align LLM recommendations with USN wargaming playbooks and established doctrine.
- Outperformed DoD human experts in complex wargaming simulations, demonstrating an 8% reduction in troop casualties.

RESEARCH EXPERIENCE

Reinforcement Learning-Driven Active Learning for Regression [Codebase] Jan. 2025 - Present

- Improved model performance by up to 8% over state-of-the-art active learning methods across 20 benchmark datasets.
- Utilized Soft Actor-Critic with Potential-Based Reward Shaping to autonomously learn the optimal data sampling policy.

Rashomon-Based Active Learning for Interpretability and Efficiency [OpenReview, NeurIPS] Jun. 2024 - Present

- Improved predictive accuracy by up to 20% with a novel active learning algorithm while reducing labeling cost.
- Designed adaptive sampling strategies to strategically collect data that minimize labeling cost while enhancing model training.
- Developed a selection metric accounting for multiple equally accurate explanations, ensuring robust & interpretable models.

Leveraging Large Language Models for Goals of Care Identification in Palliative Care Apr. 2024 - Present

- Deployed designed-based supervised learning techniques to optimize the evaluation of LLM predictions in healthcare.
- Applied a novel post-prediction inference correction, significantly refining downstream analyses for improved LLM accuracy.
- Quantified treatment effects in palliative care, addressing challenges of sparse positive outcomes in LLM predictions.

Optimal Full Matching Under a New Constraint on the Sharing of Controls Jun. 2021 - Apr. 2023

- Developed methods to estimate causal effects in non-randomized designs when clinical trials and A/B testing are infeasible.
- Implemented a novel innovation in propensity score matching, reducing the variance of the treatment effect estimate by 9%.

PRESENTATIONS

Neural Information Processing Systems (NeurIPS) [Abstract] Dec. 2024
Joint Statistical Meetings (JSM) Aug. 2023
International Chinese Statistical Association Applied Statistics Symposium (ICSA) Jun. 2023
Michigan Student Symposium for Interdisciplinary Statistical Sciences (MSSISS) Mar. 2023
Conference on Statistical Practice (CSP) [Abstract] Feb. 2023
International Conference on Health Policy Statistics (ICHPS) [Abstract] Jan. 2023

HONORS

Rackham Science Award Aug 2023
Rackham Merit Fellowship Aug. 2021
Departmental Honors in Mathematics and Economics Jun. 2021

SKILLS

R: `dplyr, tidyr, ggplot2, plotly, shiny, caret, randomForest, xgboost, glmnet, stats, lme4, MASS`
Python: `tensorflow, pytorch, pandas, numpy, matplotlib, seaborn, scikit-learn, xgboost, scipy, keras`
Additional Information: Active U.S. Security Clearance (Secret)