# Kieu Dang

NY, USA vdang@albany.edu medium.com/@kirudang kirudang.github.io

## RESEARCH INTERESTS

Trustworthy AI, with a focus on privacy, security, and robustness in LLMs. My work explores adversarial robustness, differential privacy, and watermarking to ensure reliable and ethical deployment of LLMs in real-world applications.

## EDUCATION

**State University of New York at Albany (SUNY Albany)**, NY, USA. Aug 2024 – Present
PhD Student in Information Science Cumulative GPA: 4.00/4.00
Research topics: Trustworthy Machine Learning/AI through the Lens of Privacy and Security.

**Northeastern University**, MA, USA. Aug 2021 – July 2023
Master in Analytics, Applied Machine Intelligence. Cumulative GPA: 4.00/4.00
Thesis: Improving Safety through the Integration of Multi-sensor Fusion and Deep learning-based Object Detection.

## PUBLICATIONS

- **Kieu Dang**, Phung Lai. Navigating Trustworthiness in LLMs: An Examination of Privacy, Security, and Robustness. In *Proceedings of Computational Data and Social Networks* **(CSoNet 2024)**.
- Dylan Tarace, Phung Lai, **Kieu Dang**, Unal Tatar. AI-Powered Assessment of Wazuh for Obfuscated Threat Detection. In *Proceedings of the IEEE Systems and Information Engineering Design Symposium* **(SIEDS 2025)**.

**Forthcoming**

- **Kieu Dang**, Phung Lai, NhatHai Phan, Yelong Shen, Ruoming Jin. SafeSeal: Certifiable Watermarking for LLM Deployments. In *ACM Conference on Computer and Communications Security (CCS 2025)* - **Submitted**.
- **Kieu Dang**, Phung Lai, NhatHai Phan, Yelong Shen, Ruoming Jin, Abdallah Khreishah, My Thai. SoK: Are Watermarks in LLMs Ready for Deployment? In *IEEE Symposium on Security and Privacy (S&P 2025)* - **Submitted**.
- **Kieu Dang**, Phung Lai, NhatHai Phan, Yelong Shen, Ruoming Jin, Abdallah Khreishah. $\delta$-Steal: LLM Stealing Attack with LDP. In *European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML PKDD 2025)* - **Submitted**.

## PATENTS

- **Kieu Dang**, Phung Lai, NhatHai Phan. SafeSeal: Certifiable Watermarking for LLM Deployments. Filed non-provisional US patent on April 15, 2025.

## RESEARCH EXPERIENCE

**Responsible AI Lab - SUNY Albany**, NY, USA. Jan 2024 – Present
***Research Assistant***

- Cooperate with Microsoft, NJIT, KSU, UF and HBKU (Qatar) to prove theories, implement, and conduct experiments for Trustworthiness in Machine Learning and Deep Learning using Python.
- Design and evaluate adversarial robustness strategies for large language models, focusing on watermarking and privacy-preserving techniques.

## TEACHING & STUDENT MENTORSHIP EXPERIENCE

**SUNY Albany**, NY, USA. Jan 2024 – Present
***Intern Mentor:*** Mentor undergraduate interns on the research project *Transforming Large Language Model Alignment: Automating Reference Data Generation through Explainable AI*, while also supporting their development of technical skills.

**Northeastern University**, ON, Canada. Aug 2021 – Dec 2021
***Teaching Assistant:*** ALY 6010 - Introduction to Statistics and Probability.

## PROFESSIONAL EXPERIENCE

**A Medium Corporation**, CA, USA. Nov 2022 – Present
***Machine Learning Content Engineer (Freelancer – Partner Program)***

Translate complex machine learning topics, especially in NLP, into accessible and insightful content for a diverse audience.

**Hitachi Vantara Corporation**, CA, USA Oct 2023 – May 2024
***Senior Data Scientist (Remote Full Time)***

- Spearheaded NLP initiatives to enhance financial document analysis, improving entity recognition and sentiment extraction models by 18%, leading to faster client reporting workflows.
- Collaborated with a cross-functional team of 6 engineers and domain experts to deploy deep learning models on healthcare and manufacturing text datasets, achieving a 12% uplift in operational prediction accuracy.
- Built and validated causal language modeling pipelines for retail demand forecasting, reducing supply chain prediction errors by 15% across 3 major client projects.

**Definity Financial**, ON, Canada. Jan 2023 – May 2023
***Data Scientist and Modeling (Co-op)***

- Prepared image, text, and tabular data on car accidents for predictive modeling.
- Built and deployed four models for underwriting, actuary, and claims with reproducible pipelines.

**Alibaba Group – Lazada E-commerce**, HCMC, Vietnam. Sep 2020 – Aug 2021
***Senior Manager, Category Management***

- Collaborated with the Data Science team to evaluate search exposure and sales drivers (e.g., free shipping, vouchers) using text mining and A/B testing techniques.
- Conducted hypothesis-driven ad-hoc analysis to address business-critical operational queries.

## HONORS & AWARDS

- Second Prize, Best Poster at NTIR 2025 *Apr 2025*
- Valedictorian, Master in Analytics – Fall 2021 Cohort, Northeastern University *Jul 2023*
- Hackathon Winner – OCR and Language Model for Form Autofill, Definity Financial *Feb 2023*
- Top 20 Finalist – VinUniversity Global Case Competition (VGCC), VinUniversity *Dec 2021*
- Merit-based Scholarship, Foreign Trade University *2013 & 2014*
- Talent Incubation Scholarship (Top 5 students), Coca-Cola Vietnam & Thanh Nien News *Nov 2013*
- Ambassador – Exchange Participant, AIESEC Indonesia *Jun – Aug 2013*

## PROFESSIONAL SERVICES

- **Organizer committee**: Where Innovation Meets Information NTIR Conference 2025 *April 2025*
- **Journal reviewer**: Journal of Combinatorial Optimization - JOCO 2025 *Jan - May 2025*
- **Conference reviewer**: Computational Data and Social Networks - CSoNet 2024 *Oct - Dec 2024*