Nhi N. Truong

+1 (650) 468-5460 nntruong2601 nhintruong.github.io nhintruong nhintruong

Education

DePauw University **Indiana, USA**

B.A. in Computer Science and Economics (Advisors: Jeff Gropp, Gloria Townsend) | 3.93/4.0 GPA *May 2029*

Honors & Programs: Management Fellows Program (Business Honors), Science Research Fellow Program, Information Technology Associate Program (ITAP), Honor Scholar

Technical Skills

Programming	Python, R, C++, Java
Machine Learning	Large Language Models, Time-Series Forecasting, Model Evaluation, NLP
Data Analysis	Statistical Computing, Quantitative Research, Data Visualization
Tools	Git, GitHub, animint2, webshot2, LaTeX, Google Suite, Microsoft Office

Work Experience

Stanford Trustworthy Artificial Intelligence Research Lab **California, USA**

Research Intern (Advisor: Sanmi Koyejo) *May 2026 - August 2026*

- Co-authored: Uplifting Human Decision Making in AI Evaluation by Automating Benchmark Validity Analysis. *ICML'2026* workshop on *Combining Theory and Benchmarks (CTB)*
- Developing a **latent-variable measurement model** for adaptive AI **red-teaming** that disentangles model robustness, prompt difficulty, and attacker strength from Attack Success Rate, recovering an attacker-invariant safety score for fair model comparison across three jailbreak benchmarks.

Stanford Trustworthy Artificial Intelligence Research Lab **Remote, California**

Research Assistant (Advisor: Sanmi Koyejo) *June 2023 - September 2025*

- Architected novel **AI evaluation framework** pairing public benchmark design with private test content to prevent data contamination, cutting benchmark gaming by 40% over traditional approaches.
- Built evaluation suite of 10 **NLP** tasks and 31 automated metrics across 50,000+ test samples to measure model performance, bias, and calibration.
- Co-first author: Finetuning and Comprehensive Evaluation of Vietnamese Large Language Models. *NAACL'24* *Media cover*: The New York Times, Stanford Human-centered AI, Stanford AI Lab Blog Post

VNU-HCMUT Artificial Intelligence Lab **Ho Chi Minh, Vietnam**

Research Assistant (Advisor: Tho Quan) *March 2023 - May 2025*

- Optimized a **time-series** forecasting pipeline by combining **Transformer neural networks** with **Holt-Winters statistical methods**, cutting prediction error by 15% and improving training efficiency by 60% on complex seasonal datasets.
- First author: Hybrid Transformer and Holt-Winter's Method for Time Series Forecasting. *ICLR'2024* workshop on *Time Series for Health*

Google Summer of Code: The R Project for Statistical Computing **Remote**

Software Engineering Contributor *June 2024 - August 2024*

- Migrated the `animint2` **R package** from PhantomJS to **Chrome**, refactoring the **R** visualization pipeline to enable headless browser testing and automated screenshot capture.
- Built an **HTML** multi-plot layout system and **webshot2** screenshot export, added **testthat** cross-platform tests (Windows, Ubuntu, macOS), and ported 50+ interactive visualizations to a GitHub Pages gallery.

Projects

DePause: Campus Mental Health App **Indiana, USA**

Team Leader & Technical Writer *2026*

- Built a campus mental health app in **React Native (Expo)**, **TypeScript**, and **Supabase** (PostgreSQL, auth, row-level security) that aggregates anonymous student mood check-ins into a campus-wide view
- Designed a quick mood check-in, a real-time heatmap across 21 buildings, and a recommendation engine of 80+ clinically grounded suggestions, with on-device crisis detection and tiered escalation to support resources.
- Won *2nd place* overall at TigerHacks 2026 Competition, delivering the full application.