

SON THE NGUYEN

Machine Learning Researcher — LLMs, Alignment & Optimization

Chicago, IL | snguye65@uic.edu | sonthenguyen.com | [linkedin.com/in/sonthenguyen](https://www.linkedin.com/in/sonthenguyen) | sonthenguyen/citations

SUMMARY

Machine learning researcher (PhD, December 2026) specializing in large language models and AI agents. Apply optimization and machine learning to improve training- and inference-time data quality, advancing model alignment, controllability, and reasoning. Seeking Applied Scientist / ML Engineer / Research Scientist roles.

TECHNICAL SKILLS

Languages: Python, SQL | **ML/DL:** PyTorch, Transformers, Sentence Transformers, Unsloth, vLLM, Scikit-learn

Focus areas: LLM alignment (RLHF), black-box & metaheuristic optimization, reward modeling, agentic AI, NLP

Data & infra: Pandas, Polars, FastAPI, Gradio, Docker, Git, Tailscale, Ngrok, Ray

EXPERIENCE

University of Illinois Chicago

2021 – Present

Graduate Research Assistant, LLM Alignment & Optimization

Chicago, IL

- Developed **MEMETRON**, a task-agnostic framework casting LLM decoding as discrete black-box optimization to find higher-reward responses without retraining or gradient access.
- Engineered **CURATRON**, an algorithm that recovers missing comparisons and corrects corrupted ones to complete and robustify LLM preference datasets for more reliable RLHF/DPO alignment.
- Developed and delivered a tutorial on training-free LLM alignment as lead instructor at the NAIRR 2026.
- Secured a 200,000 core-hour **NSF ACCESS** compute grant as project lead, powering the lab's agentic AI and LLM research.
- Advanced a \$25K **CTRI**-funded effort to translate the lab's LLM alignment and inference-optimization technology (MEMETRON, CURATRON) toward licensing and commercialization.

Xpressive.AI

Summer 2024

AI/Machine Learning Intern, AI Avatars

Remote

- Accelerated real-time AI avatar generation 4×, cutting latency for smoother, more responsive interaction.
- Improved avatar realism and engagement by refining the underlying machine-learning models.
- Shipped user-facing functionality to create and customize personal AI avatars, boosting personalization and engagement.

SELECTED PUBLICATIONS

Nguyen, Son The and Theja Tulabandhula (2026). “MEMETRON: Memetic Response Optimizer for Reward-Guided Post-Decoding Optimization of Large Language Models”. In: *Transactions on Machine Learning Research*. Preliminary version presented at:

- The Workshop on Information Technologies and Systems (WITS 2025).
- National Artificial Intelligence Research Resource Annual Meeting (NAIRR 2026). ISSN: 2835-8856. URL: <https://openreview.net/forum?id=QRW80Gn3vb>.

Nguyen, Son The, Niranjana Uma Naresh, and Theja Tulabandhula (2024). “CURATRON: Complete and Robust Preference Data for Rigorous Alignment of Large Language Models”. In: *Proceedings of the Fifth Workshop on Data Science with Human-in-the-Loop (DaSH 2024)*. Preliminary version presented at:

- Data Centric AI Workshop (DCAI 2024).
- Bi-directionality in Human-AI Collaborative Systems (BHAICS 2024).

URL: <https://arxiv.org/abs/2403.02745>.

EDUCATION

Ph.D., Management Information Systems, University of Illinois Chicago

December 2026

M.S., Business Analytics, University of Illinois Chicago

2021

B.S., Information Systems & Operations Management, George Mason University

2020

AWARDS, GRANTS & TEACHING

- NSF ACCESS CI Explore Grant with 200,000 core-hours, project lead (2024–2027).
- WITS 2022 Best Paper Nomination | NASA University Student Research Challenge (2021).
- UIC Graduate College Doctoral Fellowship & University Doctoral Scholarship (2021–2026).
- Instructor & TA, graduate ML courses at UIC: Deep Learning, Machine Learning, Text Analytics, ML Deployment.