Tianyi Niu

tianyiniu.github.io · tianyin4@cs.unc.edu · Chapel Hill, North Carolina

EDUCATION

University of North Carolina at Chapel Hill, Chapel Hill, NC — August 2024 – May 2026

- Master of Science in Computer Science

University of North Carolina at Chapel Hill, Chapel Hill, NC — August 2020 – May 2024

- Bachelor of Science in Computer Science (Highest Honors) with a Second Major in Linguistics
- Thesis: *Priming: Multi-Stage Pretraining using Formal Languages with Ascending Complexity*
- Overall GPA: 3.96/4.0, Highest Distinction, Dean's List (all semesters offered), Phi Beta Kappa

WORKS IN PROGRESS

All drafts and manuscripts are available upon request via email and are also accessible at tianyiniu.github.io. In progress manuscripts are in the final stages of preparation and will be preprinted in the coming weeks.

- **Tianyi Niu**, Justin Chih-Yao Chen, Yue Zhang, et al., Elias Stengel-Eskin, Mohit Bansal. *CasCal: Consensus-Aware Skill Clustering and Aggregation for LLMs. In progress.*

PREPRINTS

- **Tianyi Niu**, Jaemin Cho, Elias Stengel-Eskin, Mohit Bansal. *RotBench: Evaluating Multimodal Large Language Models on Identifying Image Rotation*. arXiv preprint, 2025.
- Rajeev Bhatt Ambati*, **Tianyi Niu***, Aashu Singh, Shlok Mishra, Shashank Srivastava, Snigdha Chaturvedi. *Socratic Students: Teaching Language Models to Learn by Asking Questions*. arXiv preprint, 2025.
- Anika Sharma, **Tianyi Niu**, Emma Wrenn, Shashank Srivastava. *Adversarially Probing Cross-Family Sound Symbolism in 27 Languages*. arXiv preprint, 2025.

PUBLICATIONS

- Jane Xing, **Tianyi Niu**, Shashank Srivastava. *Chameleon LLMs: User Personas Influence Chatbot Personality Shifts*. Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing (EMNLP), 2025.
- Brandon Prickett*, **Tianyi Niu***, Katya Pertsova. *Probing Neural Network Generalization using Default Patterns*. Proceedings of the 22nd SIGMORPHON workshop on Computational Morphology, Phonology, and Phonetics (SIGMORPHON), 2025.
- Saif Khairat, **Tianyi Niu**, John Geracitano, Zhaoqiang Zhou. *Performance Evaluation of Popular Open-Source Large Language Models in Healthcare*. Proceedings of the 23rd International Conference on Informatics, Management, and Technology in Healthcare (ICIMTH), 2025.
- Saif Khairat, John Geracitano, **Tianyi Niu**, Kaushalya Mendis. *Acceptability of Academic Large Language Models for Patient Seeking Health Information*. Proceedings of the 23rd International Conference on Informatics, Management, and Technology in Healthcare (ICIMTH), 2025.
- **Tianyi Niu**. *Priming: Multi-Stage Pretraining using Formal Languages with Ascending Complexity*. Carolina Digital Repository, 2024 (*Undergraduate thesis*). https://doi.org/10.17615/nen2-9e70

INTERNSHIPS

Large Language Model Research and Development Intern May 2024 – August 2024
Lenovo

- Researched and developed a pipeline for automatically categorizing DPO data and selecting the most appropriate judge model for each category.
- Developed an automatic scraper to locate and format open-source images with corresponding text captions, creating a corpus of reliable multi-modal ground-truth training data.
- Compiled a detailed report on dataset compositions of open-source language models.
- Verified English-Chinese translations of DPO pairs to ensure consistency in bilingual datasets.

Software Defined Networking & Software Engineering Intern May 2023 – July 2023
Nokia Shanghai Bell

- Developed and tested automated redundancy switching tools for 7750 Service Router (SR7750) and Fortinet 3800, 3900E firewalls — fully adopted by the team to replace manual configuration editing.
- Developed an experimental machine learning pipeline to extract correlation patterns in failed system health checks.
- Gave a department-wide seminar titled *Understanding Large Language Models* to an 80-person non-technical audience, including department-level business executives.

RESEARCH EXPERIENCE

Research Assistant June 2025 – Present
Multimodal Understanding, Reasoning, and Generation Lab
Department of Computer Science, University of North Carolina at Chapel Hill
Principal Investigator: Mohit Bansal, PhD

- Designed a benchmark revealing that state-of-the-art MLLMs fail at identifying image rotation (particularly 90°/270°) even with auxiliary visual information, exposing fundamental gaps in spatial visual reasoning compared to humans.
- Developed an LLM routing pipeline using model consensus to predict accuracy without ground-truth data, enabling flexible model pools and adaptation to novel tasks through synthetic profiling data.
- Investigated methods of improving multi-agent aggregation under conflicting information.

Student Researcher May 2023 – May 2025
Learning from Language Lab
Department of Computer Science, University of North Carolina at Chapel Hill
Principal Investigator: Shashank Srivastava, PhD

- Co-authored EMNLP paper (Chameleon LLMs): Investigated how user personas induce personality shifts in chatbots. Parallelized primary experiment pipeline to obtain 10x speed-ups in model response collection.
- Designed and maintained a high-throughput LLM deployment via vLLM on lab servers (including Llama, Phi, Mistral, DeepSeek). Reduced storage usage and allowed greater accessibility for new lab members.
- Designed and executed student-teacher interaction experiments, showing student-led interaction paradigms yield strong performance improvements on complex reasoning tasks.
- Contributed to a cross-linguistic sound symbolism study by compiling ”big/small” word pairs across diverse languages and coordinating with linguistic experts for phonological transcriptions.
- Completed UNC Senior Honors thesis: Proposed and evaluated a three-stage pretraining approach using formal languages of ascending complexity, finding that Chomsky hierarchy complexity is not a suitable reference for strong inductive biases.

Research Assistant February 2025 – August 2025
Center for Virtual Care Value and Excellence
School of Nursing, University of North Carolina at Chapel Hill
Principal Investigator: Saif Khairat, PhD, MPH

- Benchmarked the viability of LLMs for healthcare; compiled a pilot patient survey to assess acceptability of LLMs for sensitive mental health information.
- Led development of an LLM-powered platform for supporting cancer patients, integrating RAG pipelines for grounded academic references, Fitbit data for personalization, and the Transtheoretical Model (TTM) to encourage positive behavior change.
- Implemented a chatbot to support natural language exploration of the Digital Health Index, greatly expediting research workflows.

Student Researcher August 2024 – January 2025
Department of Linguistics, University of North Carolina at Chapel Hill
Principal Investigator: Katya Pertsova, PhD

- Co-authored SIGMORPHON paper: Examined how attention-based neural networks and logistic regression learn minority default patterns in artificial phonological environments, demonstrating type-frequency exerts a stronger effect than distributional frequency.

TEACHING EXPERIENCE

Graduate Teaching Assistant August 2024 – December 2024
COMP 426: Modern Web Programming

Undergraduate Learning Assistant August 2022 – December 2022
COMP 210: Data Structures and Analysis

PROJECTS

BERT Propaganda Detection 2023

- Fine-tuned and analyzed a BERT-based propaganda detection model for news articles on UNC's *Longleaf* HPC cluster.
- Developed a program to expose and identify incorrect labels for specific propaganda techniques (e.g., "Repetition", "Appeal to Authority").

SMTP Client and Server Emulator 2023

- Developed a suite of command-line programs that emulate client and server communication using the SMTP protocol; capable of sending MIME emails to live SMTP servers.

SKILLS

- **Languages:** Python, Java, C, SQL, Bash
- **Systems/Tools:** Linux, Slurm
- **Libraries/Frameworks:** PyTorch, Keras, Matplotlib, Scikit-Learn, NumPy, NLTK, Spacy, HuggingFace

LANGUAGES

- **Native:** English, Mandarin
- **Intermediate:** Spanish
- **Elementary:** French