

Leyao Wang

Nashville, TN || leyao.wang@vanderbilt.edu || 949-508-6809

[Website](#) || [Google Scholar](#) || [Linkedin](#)

EDUCATION

Vanderbilt University

Aug 2022 – May 2025

Bachelor of Science, Computer Science & Mathematics | Minor: Data Science

GPA: 3.99 / 4.00

Relevant Coursework: Machine Learning (Python), Artificial Intelligence (Python), Data Structure (C++), Operating System (C / Linux), Data Science (R), Database Management (SQL), Algorithms, Probability, Ordinary Differential Equations

University of California, Irvine

Sep 2021 – June 2022

Bachelor of Science, Computer Science

GPA: 4.00 / 4.00

Relevant Coursework: Software Design (Python), Linear Algebra, Statistics, Calculus

RESEARCH EXPERIENCE

Chatterjee Lab, Duke University (Remote)

May 2024 - Current

Undergraduate Researcher (Supervisor: Professor Pranam Chatterjee)

- Deployed Masked Language Model tasks on ChemBERTa to encode modified SMILES sequence of peptides
- Fine-tuned the model with Weight-Decomposed Low-Rank Adaptation (DoRA), improving the Top 10 accuracy by 41.3% in peptide-protein interaction prediction
- Evaluated fine-tuned model embeddings using PAMPA benchmarks with MLP regression, highlighting DoRA's advantages

Network and Data Science Lab, Vanderbilt University

Jan 2024 - Current

Undergraduate Researcher (Supervisor: Professor Tyler Derr)

- Leveraged LLaMA-3 to augment minority classes in Text-Attributed Graphs (TAG), significantly alleviating class imbalance in node classification across various datasets
- Developed and pre-trained a text-based link predictor to seamlessly integrate synthetic nodes, maintaining graph topology
- Conducted comprehensive ablation studies, sensitivity analysis, and baseline comparisons, demonstrating the model's robustness against class imbalance with an improvement of 23.7% in node classification accuracy.

Computer Science and Artificial Intelligence Laboratory, MIT (Remote)

Mar 2024 - May 2024

Research Assistant (Supervisor: Dr. Wengiang Chen)

- Supported fine-tuning of LLaMA-2 with wearable sensor data, advancing Sensor2Text for Q&A on human action prediction
- Assist in the evaluation of Sensor2Text with captioning scores, facilitating the project progression by 15%

SPHERE Lab, Vanderbilt University

Apr 2023 - Apr 2024

Undergraduate Researcher (Supervisor: Professor Zhijun Yin)

- Integrated health-related responses from online forums into GPT-4 and evaluated the accuracy of the model's responses by comparing them to official expert-provided answers.
- Led a systematic review on the utilization of Large Language Models (LLMs) in healthcare, filtering 65 pertinent papers from 820 search results across PubMed, ACM, and IEEE digital libraries; Highlighted the imperative for examining the reliability of LLMs

Yale University School of Medicine (Remote)

Aug 2021 - Jan 2022

Research Assistant (Supervisor: Dr. Apurv Hirsh Shekhar)

- First-authored a 20-page review regarding the efficacy of neural networks (GAN, CNN) in COVID-19 diagnosis, comparing the performance of various models in abnormal X-ray image detection and cough sound classification

MIT CEE Department (Remote)

May 2020 - Oct 2020

Research Assistant (Supervisor: Professor Otto X. Cordero)

- Simulated viral transmission using differential equations with loaded data from WHO and JHU (total population > 100,000)
- Conducted statistical analysis on the simulated data to propose effective measures for controlling pandemics

PUBLICATION

JOURNAL (Submitted)

1. **Leyao Wang**, Rishab Pulugurta, Pranay Vure, Yinuo Zhang, Aastha Pal, Pranam Chatterjee. "PepDoRA: A Unified Peptide Language Model via Weight-Decomposed Low-Rank Adaptation." Nature Methods.

CONFERENCE (Submitted)

1. **Leyao Wang**, Yu Wang, Bo Ni, Yuying Zhao, Tyler Derr. "Large Language Model-based Augmentation for Imbalanced Node Classification on Text-Attributed Graphs". 39th AAAI Conference on Artificial Intelligence (AAAI-25).

JOURNAL (Accepted)

1. ***Leyao Wang,** *Zhiyu Wan, Congning Ni, Qingyuan Song, Yang Li, Ellen W Clayton, Bradley Malin, Zhijun Yin. “*Application and Concerns of ChatGPT and Conversational Large Language Models in Healthcare: A Systematic Review*”. Journal of Medical Internet Research (JMIR).

CONFERENCE (Accepted)

1. Wenqiang Chen*, Jason Cheng*, **Leyao Wang,** Wei Zhao, Wojciech Matusik. “*Sensor2Text: Enabling Natural Language Interactions for Daily Activity Tracking Using Wearable Sensors*”. ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp/ISWC 2024).

WORKSHOP (Accepted)

1. **Leyao Wang,** Rishab Pulugurta, Pranay Vure, Aastha Pal, Yinuo Zhang, Pranam Chatterjee. “*PepDoRA: A Unified Peptide Language Model via Weight-Decomposed Low-Rank Adaptation*”. NeurIPS 2024 Workshop on AI for New Drug Modalities (AIDrugX @ NeurIPS24).

WORK IN PROGRESS

1. **Leyao Wang,** Yuying Zhao, Xutao Mao, Haoran Ma, Tyler Derr. “*Handling Missing data in Relational Database using Large Language Model -based Feature Imputation*”

WORK EXPERIENCE

Institute for Software Integrated Systems, Vanderbilt University **Apr 2024 - June 2024**
Intern (Supervisor: Professor Gautam Biswas)

- Transcribed noisy educational audio using Deepgram and WhisperX. Evaluated their performance based on the accuracy of speaker identification and transcription
- Employed GPT-4o to analyze the sentiments of transcripts and detect valence-arousal or learning-centered emotions

Vanderbilt Math Department **Aug 2023 - Dec 2023**

Teaching Assistant (Supervisor: Professor Henry Chan)

- Hosted bi-weekly office hours for Maths 2810 Probability and Statistics, supporting 4 professors in their teaching
- Addressed inquiries from 132 students across 5 sessions about probability and statistics analysis, both online and offline

Metissa.AI (Remote)

Research and Development Intern **Apr 2023 - Aug 2023**

- Managed the design of AI crash courses regarding chatbot customization, collaborating with Rice and Princeton University Ph.Ds
- Created an OpenAI-powered chatbot using GPT-3.5 to handle AP Exam questions, boosting student productivity by 40%
- Implemented Retrieval-Augmented Generation (RAG) with Chroma vector databases, fostering LLM knowledge retrieval by 20%

AWARDS & SCHOLARSHIPS

Dean’s Lists, All Semesters **Aug 2021 - Current**
Apr 2024

- For obtaining a semester GPA of at least 3.5

Vanderbilt University Summer Research Program, \$ 6000

- Funds exceptional Vanderbilt undergraduates to conduct summer research

Data Science Institute Research Program, \$ 6000 (declined due to accepting VUSRP) **Apr 2024**

- Provides stipends for outstanding Vanderbilt undergraduates for summer research relating to Data Science

Institute for Software Integrated System Summer Internship Program, \$ 8000 (declined due to accepting VUSRP) **Apr 2024**

- Grants awards to distinguished Vanderbilt undergraduates to conduct computational research

VOLUNTEER & ACTIVITIES

Education Without Barrier (Remote) **May 2021 - Jan 2023**

Web Developer & Math Teacher

- Developed and maintained online platforms for the delivery of free remote courses to underserved students
- Volunteered weekly math lessons for children with critical illnesses, providing personalized support and instruction

Associated Students at UCI **Jan 2022 - June 2022**

Student Government Senator

- Organized weekly meetings to propose and develop academic initiatives, such as a High School Outreach program and a campus Resource Database.
- Led a Mentorship Program for undergraduates, collecting contact information from 30+ graduates to invite them as mentors.

TECHNICAL SKILLS & INTERESTS

Programming Language: C++, Python, Java, HTML/CSS, JavaScript (React.js.), MySQL, Matlab, R, Racket, Linux, GitHub/Git

Framework & Libraries: PyTorch, Scikit-Learn, Transformers, Langchain, OpenAI, Pandas, Numpy, Matplotlib

Research Interests: Natural Language Processing (NLP), Large Language Models (LLM), Machine Learning, Deep Learning