Leyao Wang

New Haven, CT || leyao.wang.lw855@yale.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 **Honor:** summa cum laude (*Top 5%*)

Relevant Coursework: *Machine Learning (Python), Artificial Intelligence (Python), Data Structure (C++), Operating System (C / Linux), Data Science (R), Database Management (MySQL), 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), enhancing peptide encoding with an improvement of 41.3% in the Top 10 accuracy of peptide-protein interaction prediction

Network and Data Science Lab, Vanderbilt University **Jan 2024 - Current**

Undergraduate Researcher ***(Supervisor: Professor Tyler Derr)***

- Devised a metric to assess knowledge overlap between LLMs and knowledge graphs by analyzing confidence scores
- Leveraged Llama-3 to augment minority classes in Text-Attributed Graphs (TAG), significantly alleviating class imbalance in node classification across various datasets with an improvement of 23% in node classification accuracy
- Aligned LLMs with user preference to impute missing Amazon reviews; 20% boost in link prediction and recommendation

Computer Science and Artificial Intelligence Laboratory, MIT *(Remote)* **Mar 2024 - May 2024**

Research Assistant ***(Supervisor: Dr. Wenqiang Chen)***

- Supported fine-tuning of LLaMA-2 with wearable sensor data, advancing Sensor2Text for Q&A on human action prediction
- Assisted 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

PREPRINT

1. Bo Ni, Zheyuan Liu†, **Leyao Wang**†, Yongjia Lei†, Yuying Zhao, Xueqi Cheng, Qingkai Zeng, Luna Dong, Yinglong Xia, Krishnaram Kenthapadi, Ryan Rossi, Franck Dernoncourt, Md Mehrab Tanjim, Nesreen Ahmed, Xiaorui Liu, Wenqi Fan, Erik Blasch, Yu Wang*, Meng Jiang*, Tyler Derr*. *"Towards Trustworthy Retrieval Augmented Generation for Large Language Models: A Survey"*
2. **Leyao Wang**, Yu Wan*g, Bo Ni, Yuying Zhao, Hanyu Wang, Yao Ma, Tyler Derr. *"SaVe-TAG: Semantic-aware Vicinal Risk Minimization for Long-Tailed Text-Attributed Graphs"*.

CONFERENCE

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/IMWUT 2024).

WORKSHOP

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"*. NeurIPS 2024 Workshop on AI for New Drug Modalities (AIDrugX @ NeurIPS24).

JOURNAL

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).
2. Haoran Zhang*, Shixuan Zhuo*, Danni Song, **Leyao Wang**, Junyi Gu, Junyan Ma, Yang Gu, Minghui Ji, Meijuan Chen, Yuanyuan Guo.*"Icariin Inhibits Intestinal Inflammation of DSS-Induced Colitis Mice Through Modulating Intestinal Flora Abundance and Modulating p-p65/p65 Molecule"*

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, measuring the performance based on speaker identification
- 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 PhDs
- 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 Award for Outstanding Scholarship, Vanderbilt School of Engineering **May 2025**

- Recognized for achieving *summa cum laude* distionction (GPA ≥ top 5% of the previous three years' graduating seniors).

Honorable Mention, CRA Outstanding Undergraduate Award **Jan 2025**

- For undergraduate students in North America who show outstanding potential in an area of computing research.

Vanderbilt University Summer Research Program, **$ 6000** **Apr 2024**

- 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 develop academic initiatives, such as a High School Outreach and 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: Python, C++, Java, HTML/CSS, JavaScript (React.js.), MySQL, R, Racket, Linux, GitHub/Git
Research Interests: Natural Language Processing (NLP), Large Language Models (LLM), Machine Learning, Deep Learning