

Tra My (Chiffon) Nguyen

San Francisco, CA, USA | chiffonng@uni.minerva.edu | <https://chiffonng.github.io/>

RESEARCH INTERESTS

Knowledge Extraction from Diverse Data, especially human-generated multimedia data and multimodal data
Continual Learning and Knowledge Editing in Language Models (LMs): understanding how LMs store and utilize knowledge; retrieval augmentation and updating parametric knowledge
Language as instruction: question answering, language agents, code generation

EDUCATION

Minerva University San Francisco, CA, USA
Bachelor of Science, Computational Science (GPA: 3.84 / 4.00) 09/2021 - Expected 05/2025
Relevant courses: Machine Learning, Statistical Modeling, Causal Inference, Software Engineering (React+Flask)
Self-Study: deeplearning.ai's [Machine Learning Specialization](#), [Natural Language Processing Specialization](#),
[CMU's Advanced NLP](#)

RESEARCH EXPERIENCE

Landshut University of Applied Sciences x BMW Group Landshut, Germany
Research Intern (AI & Mixed Reality Lab) 04/2024 - Present
Topic: Point cloud registration (different sensors) and 3D scene graph
Advisors: [Professor Eduard Kromer](#) and [Professor Sandra Eisenreich](#)
T-Hub Hyderabad, India
Student Researcher (Human Resources Department) 01/2023 - 04/2023
Topic: Mental health at workplace, qualitative + quantitative research
Supervisor: [Suvidha Jani](#)

TEACHING

Minerva University Remote
Teaching Assistant, Introduction to Computer Science & Statistics 01/2022 - 04/2023

- Graded 80 weekly quizzes, provided personalized feedback, and hosted office hours for 200+ freshmen
- Assisted reviewing assignments, leading to a 20% improvement in students' academic performance
- Collaborated with faculty to refine course materials, increasing student engagement by 25%

PROJECTS

Capstone/Thesis: Can Small Language Models Speak Coherent Vietnamese? 05/2024 - Present
Advisor: [Professor Patrick Watson](#)

- Constructing a quality dataset of Vietnamese stories (3GB), combining translated content from TinyStories with crawled human-generated content
- Training a 100M-parameter Vietnamese Small Language Model, and fine-tuning on this dataset

Q&A with My Second Brain (Python, LlamaIndex, HuggingFace) 10/2023 - 02/2024

- Implemented a retrieval augmented generation system to interact with personal knowledge base of PDFs and Markdown files, resulting in 50% more contextual and accurate answers, using *LlamaIndex*, *sentence-transformer*, and *HuggingFace's Mistral and OpenAI API*

SKILLS

Languages: Python, SQL, R, Javascript
Tools: Git, Docker, Excel/Google Sheets, Jupyter Notebook
Libraries: PyTorch, Hugging Face, scikit-learn, NumPy, Pandas, Matplotlib, PyMC, Scipy, LlamaIndex, LangChain
Frameworks: Flask, React, Jax/Flax

SERVICE & LEADERSHIP

- Vietcode (2019-2021)**: Led a team of 60 to organize tech webinars and robotics classes for 500+ students
- SheCodes Vietnam (2018-2020)**: Coordinated the first all-female Vietnamese hackathon for 200 participants

REFERENCES

Patrick Watson, Professor, Minerva University, pwatson@minerva.edu
Eduard Kromer, Professor, Landshut University of Applied Sciences, Eduard.Kromer@haw-landshut.de