

NAISONG ZHOU

Bahnhofstrasse 3, 8953 Dietikon, Switzerland

☎ +41-779-885-563 ✉ naisong.zhou@epfl.ch

🌐 [naisong-zhou](https://www.linkedin.com/in/naisong-zhou) 📄 [ZnS77](https://github.com/ZnS77)

Education

École polytechnique fédérale de Lausanne

Master of Science in Computer Science 5.2/6.0

Sep. 2022 – June 2025 (expected)

Lausanne, Switzerland

Xi'an Jiaotong University

Bachelor of Engineering in Artificial Intelligence, 3.9/4.3

Sep. 2018 – June 2022

Xi'an, China

Skills

Programming Languages: *Experienced:* Python *Familiar:* Javascript C++ C# Java Scala SQL Swift

Software Development: Programming Paradigms Git Blender Unity Jekyll

Frameworks: Jupyter Matplotlib Numpy Pandas Scikit-learn HuggingFace Pytorch Tensorflow

Cloud Platforms: VS Code, Google Cloud Platform, Amazon Web Services, Kubernetes, Weight& Bias

Technologies/Frameworks: Linux, GitHub, JUnit, WordPress, Unity, Jekyll, Blender

Machine Learning: Recurrent Models, Transformer, Prompt Learning, SFT, RLHF, Low-rank Adaptation

Languages: Chinese(native), English(fluent), French(A1)

Experience

Oracle Softwares Zurich

Software Intern

Aug. 2024 – Feb. 2025 (Expected)

Zurich, Switzerland

- Developed a web-based LLM-powered chatbot to enhance MySQL Heatwave customer support by integrating GenAI features.
- Designed and implemented retrieval experiments, optimizing chunk detection and search criteria to improve data relevance and retrieval efficiency.

Bosch RBCD

Machine Learning Intern

July 2023 - Sep 2023

Wuxi, China

- Developed machine learning models for predictive analysis in automotive Engine Control Units, focusing on real-world applications.
- Implemented Gaussian process regression and recursive linear square methods for precise real-time vehicle mass and slope estimation.
- Optimized the models to achieve a mean squared error (MSE) of 10%, enhancing computational efficiency in ECUs.

Tecorigin

GPU Optimization Intern

Feb 2022 - Apr 2022

Remote

- Analyzed inefficient operators under MLPerf Datasets.
- Implemented block-based optimization strategies with c++, enhancing performance and reducing GPU memory usage with specific matrix operators.

Academic Projects

VLM-based agent for Web Navigation | *VLM prompt tuning* | *supervisor: Robert West*

Feb 2024 - June 2024

- Design and implement schema for web navigation tasks based on GPT-4-vision model.
- Navigation on Wikispeedia dataset, vary input styles, evaluate on navigation correctness and human-ness.
- Incorporate opensource VLM models in navigation agents.

Preference Aligned LLM Educational Assistant | *Instruction tuning*

Mar 2024 - June 2024

- Collect preference data for STEM exercises (open and MCQ questions) with GPT3.5 and human preference.
- Finetune on Mistral 7B with DPO trainer for preferred answers. Achieve 9% accuracy raise for MCQ questions and 50% raise in preference rewards accuracy.
- Model optimization towards quantization and RAG.

Physical optimization System for elastic sheets | *Python, C++ | Prof. Mark Pauly*

Sep 2023 - Jan 2024

- Developed incremental potential contact feature into newton optimization of elastic sheets under MeshFEM and IPC frameworks.
- Designed and implemented interactive mesh visualizations for contact simulation.
- Analyzed simulation system in case of correctness and efficiency.

Fine-Tuning and Prompt-Learning on Commonsense Causal Reasoning | *Machine Learning* Oct 2022 - Dec 2022

- Developed and compared machine learning models on the COPA dataset for CCR tasks, divided into classification and cause/effect generation.
- Implemented fine-tuning on BERT-based models and prompt-learning on GPT-3, evaluating their performance on the COPA dataset.
- Achieved over 90% accuracy with GPT-3 in prompt-learning, significantly outperforming fine-tuning methods, indicating the effectiveness of few-shot learning in model adaptation.

Bachelor Thesis: Causal Scene Graph Generation | *Python, PyTorch, GCN*

Oct. 2021 – Jun. 2022

- Developed a pipeline for visual feature extraction, unbiased causal training, and graph propagation.
- Reproduced Total Direct Effect results and evaluated unbiased training methods on Neural Motifs and VCTree, achieving an average ~10% improvement in meanRecall@10.
- Implemented causality-based relationship pair filtering to enhance scene graph generation accuracy.

Fun Projects

Part-time hobby projects.

AI Reading Companion Chrome Extension

Oct. 2024 – Present

- Identified the challenge of decreased reading patience and focus, especially in individuals with ADHD, due to fast-paced web browsing habits.
- Developed a solution using LLM to process page content and assist users with highlighted texts, summaries, and interactive chat functionality.
- Built the backend with Express.js and MongoDB, conducted prompt engineering for Gemini integration, and developed the frontend UI using Ant Design.
- Created as a submission for the Google Chrome Built-in AI Challenge.

Mod Auto-Update Feature in Stardew Valley Mod Manager (SMAPI)

Nov. 2024

- Addressed the issue of mods becoming unusable after game updates, despite updates not impacting mod functionality.
- Developed an auto-update feature leveraging detected mod metadata within SMAPI to streamline compatibility management for users and developers.