

ZHIYU (ZOE) XIE

650-391-7531 • Stanford, CA • zhiyuxie@stanford.edu • zhiyuxie.top

Education

Stanford University

M.S. Computer Science, GPA: 4.21/4.30

Sep. 2023 – Jun. 2025

Stanford, California

- Randomized Algorithms(A+), Deep Generative Models(A+), Distributed System(A+), Computer Vision(A+) ...

Tsinghua University

B.ENG. Computer Science and Technology, GPA: 3.97/4.00 (Rank: 3/202)

Aug. 2019 – Jun. 2023

Beijing, China

- Data Structures(A), Theory of Computation(A), Operating Systems(A), Theory of Computer Network(A) ...

Experience

Two Sigma

Software Engineering Intern

Jun. 2024 – Aug. 2024

New York City

- Developed **LLM Security Application Reviewer** to streamline the security assessment process for Windows Apps.
- Implemented a **Flask**-based web service with **React** frontend and **Postgres** database.
- Integrated security check APIs (e.g. VirusTotal, Signify) to analyze application executables.
- Utilized **LLM** to extract security-related information from software documentations, enhancing performance with Retrieval-Augmented Generation (**RAG**), semantic chunking, and a second-pass LLM filter.
- Deployed the service to **production**, reducing review time from days to **4 minutes**, with an estimated annual savings of **1.25 FTE**.

UCLA PlusLab

Undergraduate Research Assistant, Advisor: Nanyun Peng

Jun. 2022 – May. 2023

Los Angeles, California

- Incorporated useful external information, Abstract Meaning Representation (AMR) Graph into generative models for event argument extraction task. Published as **co-first author** in *ACL 2023*.
- Proposed **AMPERE**, which generates AMR-aware prefixes for every layer of the model, and introduced an adjusted copy mechanism. AMPERE achieves 4% – 10% F1 score improvements, and is particularly effective in low-resource settings.

Google TensorFlow Lite Team

Jul. 2021 – Sep. 2021

Beijing, China

STEP Intern, Mentor: Tian Lin

- Introduced **model distillation** method for model compression in *TensorFlow Lite Model Maker library*.
- Enabled users to create a fine-tuned end-to-end model on a customized dataset for image, audio and text classification tasks in just **6 lines of code**. Achieved a **90% reduction in parameters** while maintaining a competitive accuracy.

Publications

Test-time Adaptation in Non-stationary Environments via Adaptive Representation Alignment

- Zhen-Yu Zhang, **Zhiyu Xie**, Huaxiu Yao, Masashi Sugiyama. In *NeurIPs2024*.

TextEE: Benchmark, Reevaluation, Reflections, and Future Challenges in Event Extraction

- Kuan-Hao Huang, I-Hung Hsu, Tanmay Parekh, **Zhiyu Xie**, ..., Heng Ji. In *ACL Findings 2024*.

AMPERE: AMR-Aware Prefix for Generation-Based Event Argument Extraction Model

- I-Hung Hsu*, **Zhiyu Xie***, Kuan-Hao Huang, Prem Natarajan, Nanyun Peng. In *ACL Main 2023*.

Manual Evaluation Matters: Reviewing Test Protocols of Distantly Supervised Relation Extraction

- Tianyu Gao, Xu Han, Keyue Qiu, Yuzhuo Bai, **Zhiyu Xie**, ..., Jie Zhou. In *ACL Findings 2021*.

Honors and Awards

- Siebel Scholar (1 of 78 selected globally) 2024
- Google Women Techmakers Scholarship (34 winners in China) 2020
- The Bronze Medal in National Olympiad in Informatics (**NOI**) 2017
- The Bronze Medal in the 11th Asia-Pacific Informatics Olympiad (**APIO**) 2017

Projects

- **TinyDFS** : Designed and implemented a fault-tolerant **Distributed File System** from scratch in Go that supports diverse user-friendly APIs and optimizes for sequential read/write bandwidth. Evaluated on GCP cluster.
- **RISC-V CPU**: Designed a 5-stage pipelined RISC-V CPU in Verilog. Enabled interrupt and exception handling, virtual memory system, TLB, branch prediction, and VGA interface.
- **Ray Tracing**: Implemented efficient ray tracing and progressive photon mapping algorithm that supports soft shadow, anti-aliasing, depth of field, and texture.

Skills

Programming Languages: C++, Python, Java, Go, HTML/CSS, SQL

Technologies/Frameworks: PyTorch, TensorFlow, Numpy, Git, Linux, AWS, GCP, Docker, Flask, React, Android