Shih-Chieh Dai

+1(512)705-9690 | shihchieh.dai@utah.edu | sjdai.github.io | Salt Lake City, UT, USA

Education

The University of Utah August 2024 – May 2029 (Expected)
Ph.D. in Computer Science *Salt Lake City, UT*

The University of Texas at Austin August 2021 – May 2023
M.S. in Information Science *Austin, TX*

National Chengchi University September 2016 – June 2020
B.S. in Management Information Systems *Taipei, Taiwan*

Skills

Programming Languages: Python, R, SQL, Java, JavaScript, Shell script
Machine Learning: PyTorch, Scikit-learn, Pandas, Numpy, Matplotlib, NLTK, GenSim, Spark
Cloud and DevOps: Docker, AWS (EC2), Azure, CI/CD
Web Development: Flask, React, HTML/CSS, Nginx, REST APIs
Database: MongoDB, Postgres, MySQL
Tools: Git, Vim, Linux

Research Experience

The University of Utah August 2024 – Present
Research Assistant *Salt Lake City, UT*

- **Code LLM Security**
  - ∗ Developed a method to enhance the functionality and security of LLM-generated code by identifying incorrect start tokens and enforcing correction to the appropriate tokens.
  - ∗ Proposed an evaluation metric that considers both the security and functionality of LLM-generated code.
  - ∗ Led an empirical study to critically analyze and improve evaluation methods for assessing the security and correctness of LLM-generated code.
  - ∗ Adapted existing secure code generation methods, including SVEN, SafeCoder, PromSec, and CodeGuardPlus, to five open-source LLMs—CodeLlama, Mistral, DeepSeek, Qwen, and StarCoder—using **Instruction Tuning** and **Prefix Tuning**.

Penn State University June 2022 – May 2024
Research Assistant *Remote*

- **Large Language Models for Thematic Analysis**
  - ∗ Developed a human-large language model interaction framework using **Instruction fine-tuning** and **Chain-of-Thought** for thematic analysis, achieved a high agreement score (**0.87**) between the human raters and the LLM.
- **Counterfactual Explanations for fake claims**
  - ∗ Proposed and implemented a framework (**QA-model**, **entailment model** and **transformer model**) using **PyTorch** for generating the explanation of the fact-checked results with **70%** correctness.
  - ∗ Conducted human-subject experiments with **2000+** participants recruited from **Amazon Mechanical Turk** and **Prolific**.

Projects

LLM's Susceptibility

- Studied the susceptibility of four LLMs—GPT-4, Claude Opus, Llama-3.1-8B, and DeepSeek-R1-Distilled-Llama-8B—to factually incorrect evidence.
- Identified factors that lead LLMs to follow fake evidence, such as the similarity between the question and evidence, linguistic features, and the quantity of evidence.

LLM for Unit Test Generation

- Conducting a study on existing work related to LLMs for unit test case generation.
- Constructed test cases for two existing security-related code benchmarks, SecCodePLT and CyberSecEval, including over 2,000 samples covering Python, C, and C++.
- Proposed an LLM agent framework to improve coverage rate and correctness by iterating the generation process between the generator and the LLM judge.

Selected Publications

[1] **Shih-Chieh Dai**, Aiping Xiong, Lun-Wei Ku, *"LLM-in-the-loop: Leveraging Large Language Model for Thematic Analysis"*, In: *Findings of the Conference on Empirical Methods in Natural Language Processing* **(EMNLP Findings 2023)**.

[2] **Shih-Chieh Dai***, Yi-Li Hsu*, Aiping Xiong, Lun-Wei Ku, *"Ask to Know More: Generating Counterfactual Explanations for Fake Claims"*, In: *ACM SIGKDD Conference on Knowledge Discovery and Data Mining.* **(KDD 2022)**

[3] Kuan-Chieh Lo, **Shih-Chieh Dai**, Aiping Xiong, Jing Jiang, Lun-Wei Ku, *"VICTOR: An Implicit Approach to Mitigate Misinformation via Continuous Verification Reading"*, In: *ACM The Web Conference.* **(WWW 2022)**

[4] **Shih-Chieh Dai**, Jun Xu, Guanhong Tao, *"Rethinking the Evaluation of Secure Code Generation"*, **(under review)**.

[5] **Shih-Chieh Dai**, Chien-Kun Huang, Lun-Wei Ku, *"How Susceptible Are Large Language Models to Factually Incorrect Evidence?"*, **(under review)**.

Full Publication List: https://scholar.google.com/citations?user=4ze3U6AAAAAJ&hl=en