

YITONG GU

Location: La Jolla, CA

Master of Science

University of California at San Diego, US

+1(858)396-4128

✉ yitongg3@gmail.com

🌐 LinkedIn Profile

EDUCATION

•University of California at San Diego

Master of Science in Electrical and Computer Engineering

San Diego, CA

2024 - Expected 2026

•University of Illinois Urbana-Champaign

Bachelor of Science in Computer Engineering

Urbana-Champaign, IL

2020 - 2024

•Zhejiang University

Bachelor of Engineering in Electrical and Computer Engineering

Zhejiang, China

2020 - 2024

TECHNICAL SKILLS

Languages: C/C++, Python, x86 assembly, Javascript, Java, HTML+CSS, Verilog, SystemVerilog

Dev Tools: Docker, VScode, Git, Github, Nodejs, GDB/CMake, latex

Databases: MongoDB, Relational Database(mysql)

Relevant Skills: Algorithms, Data Structures, Operating Systems, Computer Architecture, Software Engineering, Distributed System, Numerical Analysis

INTERNSHIP EXPERIENCE

•Machine Learning Engineer Internship

Boston Scientific at Shanghai, China. Mentor: AD. Feng Kevin Xu

Shanghai, China

June 2024 - Aug. 2024

- Designed and implemented a Large Language Model powered business intelligent system used for data analysis on a daily updating dataset with an irregularly updating set of specified featured questions.
- Proposed, designed and implemented CLAPPO: Classified pLanning Agent with extendable Predefined Problem list & tOolkit, an autonomous reasoning framework optimizing structured decision-making processes.
- CLAPPO achieved an accuracy of **92%**, surpassing the **baseline 49%** yielded by existing agents, CoT, ReAct, PlanAndAct under the same settings, improving multi-step planning and context-aware generation.
- Participated in developing an Ollama-based secure environment for protected data flow between user, local Large Language Model and the private database, via a docker container supported GUI for prompted user interaction.

SELECTED RESEARCH EXPERIENCES

•Large Language Model Reasoning

QLab at University of California San Diego. Advisor: Prof. Lianhui Qin

San Diego, CA

Sep. 2024 - Present

- Participating in Flow of Reasoning: Training LLMs for Divergent Problem Solving with Minimal Examples.
- Testing and analyzing behavior features on XoT: Defying the Law of Penrose Triangle for Thought Generation.
- Conducted tests and analysis of mechanisms in hypothesis search, an inductive reasoning with language models.
- Engaged in research on ARC1-D, focusing on advancing methodologies in AI reasoning exploring frameworks that improve models' ability to handle abstract reasoning tasks.

•Fast co-bottleneck Detection among Flows from Different Sources

University of Illinois Urbana-Champaign. Advisor: Prof. Yih-chun Hu

Urbana-Champaign, IL

June 2023 - Dec. 2023

- Participated in FlowBot, a novel model-based passive co-bottleneck detector for deployment on a video server.
- Accelerating clustering algorithm by implementing integrated algorithms and decentralizing the process.
- Designed and implemented a parallelized DBSCAN algorithm in C++ with multi-threading to efficiently handle uneven distance distributions, improving clustering performance in varying density regions.
- Enhanced the algorithm's clustering performance by adjusting the neighborhood search criteria, ensuring accurate identification of clusters despite variations in data density and improving overall robustness.
- The accelerated implementation is **32% faster** than the original design, Chinese Whispers clustering method.

SELECTED PROJECTS

•Autonomous Fishing Rod with Fish Species Identification

Senior Design Project

Zhejiang, China

Feb. 2024 - June 2024

- Designed and implemented an Arduino-powered autonomous fishing rod with an integrated bite-detection system.
- Trained a lightweight neural network for real-time fish species identification, achieving 91.45% accuracy.
- Optimized an embedded AI pipeline by deploying the trained model on a Raspberry Pi 4, enabling low-power, real-time image classification with a camera interface and a fisher-friendly user interface.

•Branch Prediction for CPU Performance Optimization

Computer Architecture Course Project

San Diego, CA

Feb. 2025

- Designed and implemented a 32KByte branch predictor, researching on indexing methods, tested over gshare and tournament baselines on test sets and hidden set.
- Enhanced branch prediction accuracy by fine-tuning branch history depth and pattern matching, reducing mis-prediction rates by 14.20% (testset) and 13.87% (hidden testset).