

Min Cai |

✉ caimin2021@email.szu.edu.cn • 🌐 <https://henrycai11.github.io>

Research Interests

I have broad interests in ML and NLP, particularly in understanding the mechanisms behind neural language models (LMs), developing LLM agents capable of solving complex problems, and enhancing LLM reasoning abilities. **Currently, my primary focus is on inference-time algorithms for alignment and reasoning in LLMs.**

Recent research progress includes:

- **Interpreting and controlling LLM behaviors** for better alignment with human values (e.g., SELF-CONTROL).
- **LLM Agents** (e.g., STRATEGIST) capable of solving complex tasks, such as multi-agent social deduction games (e.g., AvalonBench).
- **Improving LLM reasoning abilities**, particularly by introducing advanced inference-time algorithms like Monte Carlo tree search (e.g., STRATEGIST), controlled text generation and representation engineering.

Education

Shenzhen University

M.S. in Computer Science

Shenzhen, China

2021 – 2024

Beijing Language and Culture University

B.A. in Translation (Linguistics)

Beijing, China

2016 – 2020

Publications

- SELF-CONTROL of LLM Behaviors by Compressing Suffix Gradient into Prefix Controller**
 - [Min Cai](#), Yuchen Zhang, Shichang Zhang, Fan Yin, Difan Zou, Yisong Yue, Ziniu Hu
 - ICML 2024 Workshop on Mechanistic Interpretability, Submitted to ICLR 2025
 - Code: github.com/HenryCai11/LLM-Control, [Website Demo](#) and [Arxiv Preprint](#)
- STRATEGIST: Learning Strategic Skills by LLMs via Bi-Level Tree Search**
 - Jonathan Light, [Min Cai](#), Weiqin Chen, Guanzhi Wang, Xiusi Chen, Wei Cheng, Yisong Yue, Ziniu Hu
 - ICML 2024 Workshop on AutoRL (ratings: 9,9,6), Submitted to ICLR 2025
 - Covered as highlight in State of AI Report 2024, by Air Street Capital.
 - Code: github.com/jonathanmli/Avalon-LLM, [Website Demo](#) and [Arxiv Preprint](#)
- DataSciBench: An LLM Agent Benchmark for Data Science**
 - Dan Zhang, Sining Zhoubian, [Min Cai](#), Fengzu Li, Lekang Yang, Wei Wang, Tianjiao Dong, Ziniu Hu, Jie Tang, Yisong Yue
 - Submitted to ICLR 2025
- AVALONBENCH: Evaluating LLMs Playing the Game of Avalon**
 - Jonathan Light*, [Min Cai](#)* (equal contribution), Sheng Shen, Ziniu Hu
 - NeurIPS 2023 workshop, Foundation Models for Decision Making
 - Code: github.com/jonathanmli/Avalon-LLM, [Website Demo](#) and [Arxiv Preprint](#)
- Self-Convinced Prompting: Few-Shot Question Answering with Repeated Introspection**
 - Haodi Zhang, [Min Cai](#) (first student author), Xinhe Zhang, Defu Lian, Rui Mao, Kaishun Wu
 - arXiv preprint: 2310.05035

Work and Research Experience

Zhipu AI

Advisor: Dan Zhang and Yuxiao Dong

Research Intern

Sep. 2024 – Present

- **Inference time LLM alignment**
 - We carried out study on LLM alignment (e.g., RLHF) at inference time using various methods such as controlled text generation, steering vectors and reward-guided tree search. Unlike reasoning, implementing inference-time methods for LLM alignment is more challenging due to the lack of ground truth answers as reward signals, and we endeavor to address problems as such in this project
 - **I'm leading the project and implemented a framework for inference-time methods**, including various sampling strategies, controlled text generation, steering vectors, and reward-guided tree search methods, e.g. beam search guided by reward model.
- **ReST-MCTS V2: Online RL with Process-Reward Guided Tree Search**
 - **Improving LLM reasoning with MCTS-augmented online training:** We proposed a self-training method using Monte Carlo tree search (MCTS) for online sampling, which aims at improving LLM reasoning ability, especially at theorem proving using lean, by sampling better trajectories using MCTS;
 - **I am implementing the training pipeline**, including online sampling with Monte Carlo tree search.

Zhipu AI

Advisor: Dan Zhang

Research Intern

Aug. 2024 – Sep. 2024

- We proposed DATASCI BENCH, a new benchmark on evaluating LLMs' ability on data science problems as scientific agents.
- We proposed a novel semi-automatic framework to tackle the difficulties of evaluating open-ended real-life data science problems. The framework automatically generates a set of (Task, Function, Code) triplets, allowing evaluation on problems that do not have gold answers.
- To elicit LLMs' ability to plan and solve problems as agents, we used MetaGPT and adapted DataInterpreter to our problem setup. **I implemented the code for the evaluation and the agent framework.**

University of Hong Kong

Advisor: Ziniu Hu, Shichang Zhang and Difan Zou

Research Assistant (Remote)

Jan. 2024 – Aug. 2024

- We proposed SELF-CONTROL, a novel method utilizing suffix gradients to control the behavior of large language models (LLMs) without explicit human annotations.
- Given a guideline expressed in suffix, SELF-CONTROL computes the gradient of this self-judgment with respect to the model's hidden states, directly influencing the auto-regressive generation towards desired behaviors.
- Further introduced SELF-CONTROL_{PREFIX} that use LoRA to compress self-collected <query, representation> pairs into a prefix controller for efficient inference-time control.
- **Led the whole project, implemented the code at Github: <https://github.com/HenryCai11/LLM-Control>**

University of California, Los Angeles

Advisor: Ziniu Hu

Research Assistant (Remote)

Jun. 2023 – May 2024

- We proposed AVALONBENCH, a game environment tailored for evaluating multi-agent LLM Agents.
- Collaboratively built the game engine for 'The Resistance: Avalon' and integrated it into AgentBench (2k stars on Github). Specifically, I implemented an asynchronous multi-agent module for AgentBench.
- We further proposed STRATEGIST which was featured in the **State of AI Report**, published by Air Street Capital as **one of this year's top contributions to the field**. STRATEGIST learns skills via bi-level tree search, and it has been integrated it into AVALONBENCH as an advanced agent. Specifically, I implemented the **dialogue search** module.

Shenzhen University

Advisor: Haodi Zhang

M.S. Student

Jun. 2021 – Jun. 2024

- We proposed Self-Convince, a self-refine framework that leverages self-generated signals, i.e. correctness, analysis to iteratively improve LLMs' reasoning ability.

Teaching Experience

- **Teaching Assistant for "Compliers" at Shenzhen University, Spring 2023.**
 - Hold a 40-student on-campus programming tutorial (2 hours per week), help students with their homework, go through and check their code.

Selected Honors and Awards

2021: Outstanding Student Scholarship, the second prize, Shenzhen University

2018: Outstanding Student Scholarship, the third prize, Beijing Language and Culture University

Core Skills

- Pytorch (Huggingface Transformers). Pytorch-based interpretability toolkits (e.g., TransformerLens)
- Parameter-efficient fine-tuning (Prefix-Prompt-Tuning, etc.)
- Asynchronous programming. Using Python's `asyncio` to build IO-bound systems, e.g. AgentBench.
- Efficient training (e.g., DeepSpeed) and inference (e.g., DeepSpeed and vLLM). For instance, using vLLM to speed up inference in DataSciBench.