YUANKAI LI

(+86) 156-0196-1787◇ yuankaili21@m.fudan.edu.cn◇ Website

EDUCATION

Fudan University *Sept. 2021 - present*
B.Eng. in Artificial Intelligence (Honor Class, Data Science Track)
GPA: 3.63/4.00 Rank: 4/26
Relevant coursework Method of Optimization(A), Stochastic Processes(A), Data Structure and Algorithm Design(A), Artificial Intelligence(Honor, A), Mathematical Statistics(Honor, A), Pattern Recognition and Machine Learning(A)

RESEARCH INTERESTS

Large Language Models(LLMs), with a specific focus on:

- Exploring and expanding the capability of LLMs to reason reliably via training, alignment and in-context learning
- Building helpful and reliable AI agents and exploring the use of LLMs in embodied AI
- Building computer systems that communicate through natural language and continuously improve through interaction

PUBLICATION

Dissecting Dissonance: Benchmarking Large Multimodal Models Against Self-contradictory Instructions
Jin Gao, Lei Gan*, **Yuankai Li***, Yixin Ye, Dequan Wang
ECCV 2024

LLMs as NPCs: Toward Human-Like and Interpretable Multi-Agent Driving Simulation
Second author, with additional authors omitted for anonymity.
Submitted to CoRL 2024

RESEARCH EXPERIENCE

Demonstration Selection for In-context Learning **Shanghai AI Laboratory**
Advisor: Prof. Dequan Wang and Prof. Zhijie Deng *Dec. 2023 - Feb. 2024*

- Proposed a training-free few-shot demonstration selection method for LLMs on knowledge-intensive QA tasks using sparse retrieval methods.
- Introduced a novel prompting paradigm that achieved 5% accuracy improvement on QA tasks including medicine and college mathematics together with the proposed demonstration selection method.
- Explored in depth what makes good demonstrations for domain-knowledge-intensive tasks.
- Contributed to a co-first-authored paper in preparation for COLM 2024.

Benchmarking LMMs against Self-contradictory Instruction **Shanghai AI Laboratory**
Advisor: Prof. Dequan Wang *Sept. 2023 - Nov. 2023*

- Introduced the idea of self-contradictory instructions in Large Multimodal Models(LMMs), emphasized its potential harm, and sought to benchmark and address this problem.

- Created a diverse benchmark dataset that aims to assess the capability of LMMs to handle self-contradictory instructions, covering both the language and the vision modality.
- Conducted thorough experiments on various LMMs and proved that current SOTA LMMs perform badly even when equipped with in-context learning.
- Contributed to a second-authored paper submitted to CVPR 2024.

Building LLMs Agent in Autonomous Driving Simulation **Shanghai AI Laboratory**
Advisor: Prof. Dequan Wang *June 2023 - Sept. 2023*

- Developed an autonomous driving simulation system using LLMs in the decision-making stage and devised a method to translate a series of LLM decisions into simulation trajectories.
- Proposed the idea that descriptive natural language can be used to generate rare corner case driving simulations, *e.g.* the vehicle executes an evasive maneuver when detecting an accident ahead.
- Through massive experiments, proved that LLMs can handle such decision-making tasks and respond accordingly to natural language.
- Contributed to a second-authored paper submitted to ICRA 2024.

PROJECT EXPERIENCE

Supervised Fine-tuning on Small Language Models **Fudan University**
Course Project of Pattern Recognition and Machine Learning *Dec. 2023 - Jan. 2024*

- Finetuned GPT2-small over the MOSS-002 SFT dataset and conducted various capability tests with the BigBench dataset.
- Explored the performances of models fine-tuned with different datasets, *i.e.* MOSS-002 SFT "helpful", "harmless" and "honest" dataset.
- Dealt with practical problems such as how to choose the most human-preferred model when facing over-fitting during fine-tuning.

Simple Machine Learning Framework **Fudan University**
Course Project of Artificial Intelligence(Honor) *Oct. 2023 - Dec. 2023*

- Redeveloped a simple machine learning framework in Python that implements a backpropagation algorithm, various neural network architectures (like MLP and CNN), manifold statistical learning models (like HMM and CRF), etc.
- Achieved accuracy over 95% in tasks like text recognition and named entity recognition(NER) classification.

AWARDS AND SCHOLARSHIPS

1^{st} *Prize*, East China Mathematical Contest in Modeling (Rank 4^{th} in East China) *June 2023*
2^{nd} *Prize*, Contemporary Undergraduate Mathematical Contest in Modeling (Shanghai) *Nov. 2023*

SKILLS

Programming Languages and skills C, Python, Matlab, R, LaTeX, Linux, Git
Frameworks Pytorch, Huggingface Transformers, Huggingface Datasets, Langchain
Languages English (TOEFL iBT 104), Chinese (native)