Fengyuan Liu

oxfengyuan@gmail.com

Education Background

University of Oxford, Oxford, England 2022-2023

- M.Sc. Advanced Computer Science
- Topics Covered: Safety of Foundation Model, Multimodal AI

University of Washington, Seattle, WA 2017-2021

- B.Sc. in Computer Science
- B.Sc. in Applied Computational Mathematical Science (Data Science & Statistics)
- GPA: 3.95/4.0 (around top 0.5%)
- Topics Covered: ML, DL, RL, NLP, Stochastic Process, Cryptography

Publications

[1] **Which Model Generated This Image? A Model-Agnostic Approach for Origin Attribution**
Fengyuan Liu, Haochen Luo, Yiming Li, Philip Torr, Jindong Gu
European Conference on Computer Vision ***(ECCV)***, 2024

[2] **An image is worth 1000 lies: Transferability of adversarial images across prompts on vision-language models**
Haochen Luo*, Jindong Gu*, **Fengyuan Liu**, Philip Torr
International Conference on Learning Representations ***(ICLR)***, 2024

[3] **DrugGPT: A Knowledge-Grounded Collaborative Large Language Model for Faithful and Evidence-based Drug Analysis**
Fenglin Liu*, Hongjian Zhou*, Wenjun Zhang, Guowei Huang, Lei Clifton, David Eyre, Haochen Luo, **Fengyuan Liu**, Kim Branson, Patrick Schwab, Xian Wu, Yefeng Zheng, Anshul Thakur, and David A. Clifton
Nature Biomedical Engineering ***(Nat. Biomed. Eng)***, 2024 (under review)

[4] **OpenFE: Automated Feature Generation beyond Expert-level Performance**
Tianping Zhang, Zheyu Zhang, Haoyan Luo, **Fengyuan Liu**, Wei Cao, Jian Li
International Conference on Machine Learning ***(ICML)***, 2023

Research Experience

Department of Engineering Science, University of Oxford Oxford, United Kindom
Research Intern, TVG, under the supervision of Dr. Jindong Gu and Prof. Philip Torr 05/2023–10/2023

Which Model Generated This Image? A Model-Agnostic Approach for Origin Attribution [1]

- Introduce a new important task, which aims to examine whether a given image is generated by a particular model.
- We formulate the introduced problem as a few-shot one-classification task. To address the task, we further propose a simple yet effective solution, named OCC-CLIP.
- We conduct extensive experiments on various visual generative models to verify the effectiveness of our OCC-CLIP. A further experiment is done to show its applicability in real-world commercial generation systems.

An Image is worth 1000 lies [2]

- Introduce cross-prompt adversarial transferability, an important perspective of adversarial transferability, contributing to the existing body of knowledge on VLMs' vulnerabilities.
- Propose a novel algorithm Cross-Prompt Attack (CroPA), designed to enhance cross-prompt adversarial transferability.
- Extensive experiments are conducted to verify the effectiveness of our approach on various

VLMs and tasks. Moreover, we provide further analysis to understand our approach.

Institute for Interdisciplinary Information Sciences, Tsinghua University Beijing, China
Research Intern, ADL Group, under the supervision of Prof. Jian Li 05/2022–09/2022

Automatic Feature Generation [4]

- Used GBDT to design a model called OpenFE to quickly and accurately measure the validity of new features
- Reproduced AutoCross, AutoFeat, SAFE and FCTree methods and compared them with OpenFE
- Did experiments and compared the prediction results with various kinds of databases

Smart beta based on multi-factor models

- Pre-processed raw factors in the tabular form about all stocks listed on the Shanghai and Shenzhen stock markets from 2017 to present
- Dealt with factors by filtering stocks, excluding extreme values, filling null values, doing industry neutral, and standardizing.
- Mainly employed Lightgbm to train and compare the prediction results with different labels (pct1, pct2, or pct5) with various factors combination
- Wrote a script to run once per day to forecast and prepare for practical application

Industry Experience

Tencent Shenzhen, China
Research Intern at Tencent AI lab & Robotics X 01/2024-present

Self-play for LLM-based Agent

- Investigate alignment through the lens of two-agent games, involving iterative interactions between an adversarial and a defensive agent.

Safety of Multi-LLM Systems

- Explore the weakness of Multi-agent Systems

Morgan Stanley
Part-time Assistant (PTA) of the Investment Analysis Project 01/2020-02/2020

- Analyzed the financial data in the annual and semi-annual reports of ION Geophysical Corporation
- Applied the Altman Z Score and SWOT model to analyze the basic situation, bankruptcy probability and acquisition risks & opportunities of the company

Activities & Talks & Service

Studies, Experiments, Applications Academy (organized by students from Keble College)
Honor Scholar of Mathematical Studies & Computer Science department 09/2022-03/2023

- Guided more teenagers to launch research, help them explore their interests and improve the society.

Oxford Fintech & Legaltech Society
Research Associate 01/2023-04/2023

- Explore the impact modern technology is having on financial institutions legal services, and regulation.

Academic Talks
A conversation with Fosun Group Global Partner Mr. Vincent Li - Oxford Said Business School 03/2023
Structural Deep Learning in Financial Asset Pricing by Jianqing Fan - Department of Statistics 10/2022

Conference Reviewer
NeurIPS 2023

Skills & Hobbies

Professional Qualification: CFA Exam Level I (August 2021): Pass
Computer Skills: Java, Python, C#, SQL, Java Script, MATLAB, R, LaTeX
Hobbies: Chinese Kung Fu, Piano, Swimming, Ancient Chinese Philosophy