

# HSIANG YAO

Hong Kong ▪ +86 18817750258 ▪ yaoxiangfd@gmail.com ▪ <https://scottyaohk.github.io>

## EDUCATION EXPERIENCE

### THE UNIVERSITY OF HONG KONG (QS21)

*Master of Finance (Financial Engineering)*

**HONG KONG**

*Aug. 2022 – May 2023*

### FUDAN UNIVERSITY (QS34)

*Bachelor of Economics (Mathematical Economics)*

**SHANGHAI**

*Sept. 2018 – June 2022*

- **GPA:** 3.49/4.0

- **Courses:** Python Programming (4.0), Mathematical Analysis (4.0), Differential Equation (4.0), Linear Algebra, Probability and Mathematical Statistics, Approach to Time Series Analysis, Operations Research, Quantitative Economics (4.0), Econometrics, Intermediate Microeconomics (4.0), Game Theory

### STANFORD UNIVERSITY

*Summer Workshops (Fundamentals of Data Science)*

**ONLINE**

*Aug. 2022*

- **Courses:** Mathematical Optimization, Intro to Machine Learning, Intro to Deep Learning, Intermediate Topics in Machine Learning and Deep Learning, Deep Learning for NLP, Intro to High Performance Computing

### COURSERA CERTIFICATES

**ONLINE**

- **Courses:** Discrete Mathematics Generality, Computer Organization, Operating Systems, Data Structures, Algorithmic Toolbox, C for Everyone: Programming Fundamentals, HTML, CSS, and Javascript for Web Developers

## PROFESSIONAL EXPERIENCES

### UNIVERSITY OF HONG KONG

**HONG KONG, SHEN ZHEN**

*Full-time Research Assistant at Artificial Intelligence Institute*

*Apr. 2023 – Now*

- Worked on the finetuning of LLMs. Developed chatbots with the capabilities of conducting conversations of different relationships, e.g., lovers, opponents. I curated and cleaned open-source datasets. Then finetuned GPT models, e.g., davinci and llama2, with conversation datasets of different relationships to produce different styles of chatbots. The finetuned chatbots can exhibit numerous features in the corresponding relationship type.
- Worked on unsupervised learning on image datasets. Applied various models like ResNet, VGG, VAE, etc. to extract low dimensional features of the images. Applied visualization algorithms to explore the patterns in the feature dataset. Successfully categorized the data into six types according to their semantic meanings.
- Designed and trained a CNN on coauthorship dataset, further worked on the explainability of CNN using Grad-CAM.
- Trained a Graph Neural Network for link prediction on coauthorship dataset, ultimately achieving an AUC score of 90%.
- Applying meanfield MARL to the problem of population mobility under crisis. (currently in progress)

## OTHER EXPERIENCE

### TALKVERSE (AI LANGUAGE LEARNING STARTUP)

**SHANGHAI**

*Co-founder (No longer participating in company matters)*

*2023*

- Our first product (烤鸭斩) is now available on Wechat.
- Undertook the responsibilities of a Product Manager, guiding the product from idea to launch. Conducted user research to understand user needs. Utilized user research insights to determine the prototype. Collaborated closely with UI designers and frontend developers, diving deep into specifics to ensure an excellent realization of product features, interface designs.
- Undertook the responsibilities of a CTO, solely responsible for backend development, including building the core logic, designing and implementing the system architecture, database designing, conducting high-concurrency tests, etc.
- Various other jobs in a startup, including connecting with angel investors, proactively reaching out to top talents, etc.

### PYTHON FREELANCE DEVELOPER

**WORK FROM HOME**

*Part-time.*

*Aug. 2022 – Mar. 2023*

Developed numerous software for different scenarios.

- Developed email marketing software that automates the tedious daily work of a salesperson. This software successfully addresses the pain point of my customer and it still generates subscription fee for me monthly.
- Developed pitch identification software for the Chinese string instrument Guzheng for a customer who want to use it for research purpose. The software realized an accuracy of 97% using Fourier transformation and peak detection algorithm.
- Other software I developed: color segmentation software, DNA sequence processing software, etc.

## SKILLS & INTERESTS

- Technical (ML-related): Python (Advanced), Pytorch (Advanced), Various huggingface libraries (Advanced)
- Technical (CS-related): Flask (Advanced), Linux (Advanced), Redis (Intermediate), MySQL (Intermediate), Nginx (Intermediate), Docker (Intermediate), C (Intermediate), Java (Intermediate), HTML / CSS / Javascript (Rudimentary)
- Language: Native in Mandarin, Fluent in English (TOEFL-112), Rudimentary in Cantonese
- Interests: Skiing, Diving (OW diver), Muay Thai, Travelling

# 姚想

电话: +86 18817750258 | 邮箱: yaoxiangfd@gmail.com | 个人网站: <https://scottyaohk.github.io>

## 教育经历

### 香港大学

金融硕士 (金融工程)

2022 年 8 月 – 2023 年 5 月

### 复旦大学

经济学学士 (数理经济学)

2018 年 9 月 – 2022 年 6 月

- 绩点: 3.49 / 4.0
- 课程: Python (4.0), 数学分析 (4.0), 微分方程 (4.0), 线性代数, 概率论与数理统计, 时间序列, 运筹学, 数量经济学 (4.0), 计量经济学, 中级微观经济学 (4.0), 博弈论

### 斯坦福大学

数据科学暑校

2022年8月

- 课程: 数学优化、机器学习导论、深度学习导论、机器学习和深度学习的中级课题、NLP、高性能计算导论

## Coursera 证书

- 课程: 离散数学概论, 计算机组成原理, 操作系统, 数据结构和算法, C 语言基础, HTML, CSS 以及 Javascript

## 职业经历

### 香港大学

人工智能研究院全职研究助理

香港、深圳

2023年4月 – 至今

- 对 LLM 进行微调, 开发了能进行不同关系对话的聊天机器人, 例如情侣关系、对手关系等等。收集并清洗了开源数据集, 然后使用不同关系下的对话数据集对 GPT 模型 (davinci 和 llama2) 进行了微调, 以产生不同风格的聊天机器人。微调后的机器人能够明显展现出对应关系的对话风格。(demo: <http://85.239.236.92:5000/>)
- 在图像数据集上进行无监督学习。运用 ResNet、VGG、VAE 等模型, 提取图像的低维特征, 并应用可视化算法来探索特征集中蕴含的模式, 结论能够把数据分成有明显语义区别的六大类。
- 针对目标需求, 设计训练 CNN 网络, 并解释 CNN 的结果。应用 Grad-CAM 以及可视化特征图等方法探索图像的各部分对最终分类的影响。
- 针对作者合作关系数据集, 清洗并整理数据, 训练 GNN 进行 link prediction, 最终 AUC 分数达到 90%。
- 运用 meanfield MARL 的方法探究灾难情况下人口流动的规律。(目前正在进行中)

## 其他经历

### 话界 (上海) 科技有限责任公司

联合创始人 (目前已不参与公司事务)

上海

2023年

- 我们的第一款产品目前已在微信小程序平台上线 (烤鸭斩, 雅思口语素材生成 APP)。
- 承担了产品经理的角色, 包括调研用户需求, 根据需求出产品原型图, 与 UI 设计师与前端工程师配合, 确定产品的具体逻辑和交互细节, 并协调各方完成产品的开发, 最终上线。
- 承担了技术负责人的角色, 独立负责后端开发, 包括后端代码业务逻辑的撰写, 设计并实现整体技术架构, 数据库设计, 压力测试, 等等。
- 其他各种工作, 比如主动接触潜在的优秀人才, 与天使投资人沟通, 等等。



## Python自由职业开发者

为各个背景的客户定制化开发软件。

2022年8月 – 2023年3月

- 开发了一款电子邮件营销软件, 自动化了销售人员繁琐的日常工作, 成功解决了客户的痛点, 目前该客户仍然每月为此软件支付订阅费。
- 为客户开发了一款用于识别古筝音高的软件。该软件使用傅立叶变换和峰值检测算法实现了 97% 的准确率。
- 开发的其他软件包括: 色彩分割软件, DNA 序列处理软件, 等等。

## 职业技能

- 机器学习相关: Python (熟练)、Pytorch (熟练)、各种 huggingface 库 (熟练)
- 计算机相关: Flask (熟练)、Linux (熟练)、Redis (掌握)、MySQL (掌握)、Nginx (掌握)、Docker (掌握)、C (掌握)、Java (掌握)、HTML / CSS / Javascript (初级)
- 语言: 普通话 (母语)、英语 (流利, 托福 112)、粤语 (初级)
- 爱好: 滑雪、潜水 (OW diver)、泰拳、旅游

## 附录（本科实习经历，可忽略）

### 腾讯CSIG

上海

投资分析实习生（toB/tech）

2021年9月 – 2022年3月

- 针对税务信息化进行 topdown 的研究(涵盖维度:金税工程、电子发票、这两者对财税信息化的影响、纳税流程实务、增值税发票 workflow 实务、市场规模测算、行业 mapping、中美对比; 研究的目标是寻找财税信息化行业有没有新的机会)
- 整理 toB/tech 领域的融资新闻, 已覆盖约 520 家国内企业, 194 家国外企业, 28 家 IPO(每一家企业覆盖的维度包括主要业务、商业模式、收入方式、历史投资者、运营状况、赛道基本面、竞品、投资价值等等; 覆盖赛道包括:SaaS、PaaS、AI、Infra、智慧产业、网络安全、VR/AR、自动驾驶等等; 整理结果每周在周会上向所有同事汇报)
- 现场参与与创业者的访谈, 做会议纪要(不需要借助录音, 现场速记加上会后整理, 能够快速交付结构化的会议纪要)
- 对细分行业和公司等进行研究, 包括 Loom、Webflow、Flexport 等

### 中金公司

上海

有色组研究助理

2021 年 6 月 – 2021 年 8 月

- 负责撰写铜管行业龙头海亮股份的深度报告(涉及铜加工、空调、水泥、猪肉等行业)
- 协助撰写碳中和相关的系列报告, 包括新能源车对铜需求的拉动、光伏对铜需求的拉动、风电对铜需求的拉动、储能对铜需求的拉动
- 负责铜行业的周报, 维护相关数据库, 负责公众号的编排和相关 ppt 的制作
- 关注其他类别的金属, 包括锡、锂、稀土等, 为研究工作做出相关的支持

### 中信证券

上海, 绍兴

投资银行部实习生

2021 年 1 月 – 2021 年 5 月

参与浙江某化工企业的 IPO 项目, 主要负责协助财务条线, 具体包括:

- 根据收入成本大表, 分析财务指标变动原因, 并汇报给项目组成员
- 撰写招股书管理层讨论与分析部分中的同行业可比公司部分
- 查找案例; 绘制 Pitchbook
- 负责核对内核意见答复

### 光大证券

台州

投资银行四部实习生

2020 年 6 月 – 2020 年 8 月

参与某包装企业 IPO 项目的半年报更新工作:

- 负责更新招股书中的数据及分析, 撰写上半年经营情况说明, 经过修改后递交给预审员
- 与客户沟通, 针对销售业务和采购业务进行穿行测试、截止性测试; 整理核查函证、核查第三方回款、盘点固定资产