

# YUEHAN ZHANG

✉ john.fohim@gmail.com · ☎ (+86) 173-629-97735 ·

## EDUCATION

---

**Wuhan University (WHU)**, Wuhan, China 2021 – Present

Master student in Data Science, GPA 3.45/4(**Average Score 88.21**), expected June 2024

**Wuhan University (WHU)**, Wuhan, China 2017 – 2021

Bachelor in Library Science, **GPA 3.81/4**

## PROJECT EXPERIENCE

---

On the Understanding of Argumentative Zoning of Scientific Documents Based on Deep Learning

*Bachelor Graduation Thesis*

Brief introduction: Long document summarization task. Constructed a structured full text dataset containing 370,000 scientific paper using PubMed-OA. Applied a divide and conquer framework for BART model fine-tuning and Rouge-L increased by 5 percentage points.

Construction of Data Platform and Deep Mining of Foreign Parliamentary Data

*MySQL, Python* National level Student Project

Brief introduction: Designed the database structure and implemented it with MySQL. Deployed the server. Data Platform is <http://parliament.whu.edu.cn>. Interviewed members of the Ministry of Foreign Affairs about the platform. Visualized the relationship between nations or organizations from the perspective of British Parliament using network analysis tools based on texts of parliaments.

## PUBLICATIONS

---

### Peer-reviewed journal and conference papers

- Wei Lu, Yan Ren, Yong Huang, Yi Bu, **Yuehan Zhang**. Scientific collaboration and career stages: An ego-centric perspective. *Journal of Informetrics*. DOI:10.1016/j.joi.2021.101207
- **ZHANG Yuehan**. Bibliometric Analysis of the Retracted Papers between 2012-2018 from China. *Library Tribune*, 2021,41(11):52-59.

### Ongoing

- Dying of detection: Is text generated by AI detectable?

## PRESENTATION

---

- Junling Wang, **Yuehan Zhang**, Jiani Huang, Jiayu Shen, Yiyang Wang, Jiamin Wang, Jiming Hu, and Wei Lu. Analyzing International Relations from British Parliamentary Debates. In *Proceedings of the ACM/IEEE Joint Conference on Digital Libraries in 2020 (JCDL'20)*. Association for Computing Machinery. (**Poster**)
- **Yuehan Zhang**. Bibliometric analysis of retracted publications between 2012-2018 in China. Workshop 7-1 Data-driven Smart Library in China, JCDL'20. Association for Computing Machinery. (**Oral Presentation**)

## SKILLS

---

- Programming Languages: Python > Java > VBA == SAS
- Platform: Linux(Ubuntu)
- HuggingFace, Pytorch, Cloud Services such as object storage, Git
- Crawlers and Browser Automation(Selenium), Task Automation on Windows(AutoIt)
- LLM related: Fine-tuning, Multi-card training(DeepSpeed), Instruction Tuning, Prompt Engineering

## HONORS AND AWARDS

---

<i>3<sup>rd</sup> Prize</i> , CAIL(Chinese Challenge of AI in Law) 2022, Legal News Summarization Track	Dec. 2022
National Encouragement Scholarship	2018-2019,2019-2020
<i>2<sup>nd</sup> Prize</i> , Scholarship Wuhan University	2018-2019,2019-2020
<i>3<sup>rd</sup> Prize</i> , Scholarship Wuhan University	2017-2018
<i>Merit Student</i> , Wuhan University	2017-2020
<i>3<sup>rd</sup> Prize</i> , SAS China Data Analysis Championship, SAS Greater China	2020
<i>2<sup>nd</sup> Prize</i> , 2019 NCIECC, OVERSEAS ENGLISH	Dec. 2019
<i>Excellence Award</i> of the 2 <sup>nd</sup> National University Data-Driven Innovation Competition	May. 2019
<i>1<sup>st</sup> Prize</i> of IEEE Xplore Search Contest	2018

## PROFESSIONAL EXPERIENCE

---

### Internship

*Research Assistant Intern* in Semantics Mining based on Deep Learning Group, The Wuhan Branch of the National Science Library, Chinese Academy of Sciences. Jun. 2020-Sep. 2020  
Brief introduction: Designed research protocol;Crawled medical news text for BERT fine-tuning.

### Voluntary Work

*Venue Volunteer* for the 7<sup>th</sup> CISM Military World Games Oct. 2019  
*Technical Volunteer* for the ACM/IEEE Joint Conference on Digital Libraries(JCDL'20) Aug. 2020  
*Member* of Wuhan University Youth Volunteers Association Oct. 2017-Oct. 2018

### Teaching

*Teaching Assistant*, Advanced English of Class of 2026 Feb. 2023-Jun. 2023

## MISCELLANEOUS

---

- Languages: English - **TOEFL 109**(*Listening30Reading29Speaking24Writing26*), Mandarin - Native speaker
- Interests: Playing the guitar, Basketball, Billiards, Table tennis, Soccer, Badminton, Video clip(Best Screen-play and 2<sup>nd</sup> Award of "Lab safety" Micro Video Solicitation, Wuhan University)
- A Live Editor & Live Announcer of the English Department of Wuhan University Broadcasting Station
- A tenor of the DrinkWater Choir
- 4<sup>th</sup> Place of Group one in 2018 Orienteering Competition of Wuhan University

# 张约翰

+86-173-6299-7735

yaohan\_trust@163.com

湖北武汉

## 教育经历

2021.09-现在	武汉大学	数据科学	硕士
2017.09-2021.6	武汉大学	图书馆学	本科 (GPA: 3.8/4.0)

- **荣誉:** 国家励志奖学金; 研究生二等奖学金; 乙等奖学金; 丙等奖学金;
- **相关课程:** 自然语言处理/机器学习导论/数据库技术与应用/SAS/VBA/数据结构/Java高级语言程序设计
- **竞赛:** 中国法律智能技术评测 (CAIL) 涉法舆情摘要赛道三等奖 (2022.12) 第二届全国高校数据驱动创新大赛优秀奖; 2018 "外研社杯" 全国英语阅读大赛校三等奖; 2018 IEEE Xplore Search Contest First Prize; 全国高校创新英语挑战赛二等奖; 第三届武汉大学学术搜索之星挑战赛三等奖

## 相关经验

LLM生成文本的可检测性	研究论文	2023年3月 - 2023年7月
• 构建Prompts并微调GPT-2模型开发实验数据集, 于多种检测器上进行评测, 独立完成论文撰写。		
中科院武汉文献情报中心基于深度学习的语义挖掘研究组	实习生	2020年6月 - 2020年9月
• 编写爬虫获取医学新闻数据、整理NSTL监测平台数据进行BERT微调, 支持科技文献AI引擎。		
国外议会数据平台建设 with 深度挖掘研究	国家级项目	2018年11月 - 2020年11月
• 议会数据集及网站已发布parliament.whu.edu.cn; 设计MySQL数据库结构及在WindowsServer部署网站。		
• 《多国议会数据集及平台建设研究》于《情报工程》; 英文海报《Analyzing International Relations from British Parliamentary Debates》, JCDL2020。		
生物医学科学文献的生成式结构化摘要构建	毕业论文	2021年4月
• 使用生物医学领域PubMed-OA全文数据, 构建一个37万篇结构化全文摘要配对数据集, 微调BART模型, 在Rouge和Bert-score指标上评测。		

## 语言能力

- 托福: 109分 (109/120)

## 相关技能

- 熟悉Linux系统常用操作, 熟练在Ubuntu环境中编程; 熟练使用对象存储等云服务
- 熟悉爬虫与自动化, 熟练编写Selenium爬虫, 能够绕过验证码识别并使用代理IP池; 结合AutoIt等工具可实现Windows平台任务自动化 (自动登录/上传文件等)
- 了解LLM相关技术, 了解多卡训练, 会使用DeepSpeed, 有指令微调经验, 熟悉Prompt Engineering
- 熟悉文本摘要技术, 熟练使用pandas等数据处理包
- 熟悉Git, 能够协同工作; 有Spark、Flask使用经验; 会使用MySQL、Redis, 能够编写SQL语句
- 熟悉Python, 使用Pytorch框架, 了解SPSS&SAS编程、Java开发
- 熟练使用Office三件套、Visio, 熟悉Excel, 能够编写VBA处理数据

## 校园经历

第七届世界军人运动会志愿者	2019年10月 - 2019年10月
• 作为羽毛球场馆信息技术处志愿者副组长, 累计服务时长60+, 受到执委会“优秀赛会志愿者”表彰。	
武汉大学广播台英语部直播编辑&直播播音员	2017年10月 - 2019年10月
• 主要负责New Trend节目录制, 每月一到两次进行英文稿件撰写与直播。	