Date of birth：11/29/1998          Phone：+86-17362997735

Email：yaohan_trust@163.com          Postal Code:430072

## EDUCATION

- Wuhan University, Wuhan, China (2021.9-Present), Master of Data Science
- Wuhan University, Wuhan, China (2017.9-2021.6), Bachelor of Library Science (GPA:3.81/4.0)

Selected Course: NLP/ML Introduction/Databases and Application/SAS/VBA/Data Structure/Java

## RESEARCH EXPERIENCE

**Project:** Construction of Data Platform and Deep Mining of Foreign Parliamentary Data

**Description**: Undergraduate Training Program for Innovation and Entrepreneurship (National level). Participated in constructing the original database and website of Parliament Big Data Platform and used text mining techniques to study international relations based on texts of parliaments.

**Project:** Theme Distribution and Evolution of Innovation and Entrepreneurship training projects for college students.

**Description:** Undergraduate Training Program for Innovation and Entrepreneurship (College level). Processed data and used network analysis tools to detect the distribution features and evolution path.

**Graduation Thesis**: On the Understanding of Argumentative Zoning of Scientific Documents Based on Deep Learning.

**Description:** Construct a structured dataset for scientific document summarization using Pubmed-OA, and apply several baselines and fine-tune BART model to achieve better performance.

## PUBLICATIONS

**Peer-reviewed journal and conference papers:**

1. 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

2. ZHANG Yuehan. Bibliometric Analysis of the Retracted Papers between 2012-2018 from China. Library Tribune, 2021,41(11):52-59.

**Ongoing:**

1. **Dying of detection: Is text generated by AI detectable?**

## PRESENTATION

1. 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**)

2. 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＆RELATED EXPERIENCE

1.Experience with Pretrained Models/HuggingFace/

2.English Language: **TOEFL109** with Listening30 Reading29 Speaking24 Writing26

3.Familiar with common operations of Linux system, proficient in programming in Ubuntu environment; proficient in using cloud services such as object storage

4.Familiar with crawlers and automation, skilled in writing Selenium crawlers, able to bypass verification code identification and use proxy IP pool; Windows platform task automation combined with tools such as AutoIt.

5.LLM related technologies, multi-card training, DeepSpeed, Instruction tuning experience, familiar with Prompt Engineering

6.Familiar with Git, able to work together;

7.Familiar with Python, use Pytorch framework, understand Java development

## WORKING EXPERIENCE

- VOLUNTARY WORK

    1.Venue Volunteer for the 7th CISM Military World Games, 2019.10-2019.10

    2.Technical Volunteer for the ACM/IEEE Joint Conference on Digital Libraries '20, 2020.8-2020.8

- STUEDENT WORK

    1.Commissary in charge of studies of class 2017 in Library Science, Wuhan University, China (2020.9-2021.6)

    2.Monitor of class 2021 in Data Science, Wuhan University, China (2021.9-Present)

    3. Member of Wuhan University Youth Volunteers Association, Wuhan University, China (2017-2018)

    4. Live Editor & Live Announcer of the English Department of Wuhan University Broadcasting Station, Wuhan University, China (2017-2019)

- INTERSHIP

    Research assistant intern. Semantics Mining based on Deep Learning Group, THE WUHAN BRANCH OF THE NATIONAL SCIENCE LIBRARY, Chinese Academy of Sciences. Wuhan, China. (2020.6-2020.9)

## HONORARY  AWARDS ＆ COMPETITIONS

- CAIL(Challenge of AI in Law) 2022, Legal News Summarization Track, Third Prize, 2022.12
- "National Encouragement Scholarship", Wuhan University, 2018-2019, 2019-2020
- "The Second Prize Scholarship", Wuhan University, 2018-2019, 2019-2020
- "The Third Prize Scholarship", Wuhan University, 2017-2018
- "Merit Student", Wuhan University, 2017-2018,2018-2019, 2019-2020
- Third Prize of Top10 Stars of Scholar Search Contest, Wuhan University, 2021
- Third Prize of 2020 SAS China Data Analysis Championship, SAS Greater China, 2020
- Second Prize of 2019 NCIECC, OVERSEAS ENGLISH, 2019.12
- Third Prize of 3rd Star of Scholar Search Contest, Wuhan University, 2019.11
- Third Award in the 4th Sea Knowledge Contest of Wuhan University, 2019.5
- Excellence Award of the 2nd National University Data-Driven Innovation Competition, 2019.5
- IEEE Xplore Search Contest First Prize, 2018
- Third Prize Award in the Preliminary Round of the 2018'FLTRP Cup' English Reading Contest, 2018

## STUDENT ACTIVITIES ＆ INTERESTS

- STUEDENT ACTIVITIES

1.A tenor of the DrinkWater Choir (Choir of School of Information Management), 2017-2019

2.Fourth Place of Group one in 2018 Student Orientation Competition of Wuhan University, 2018.9

3. Summer Holiday Social Practices in Beijing: Interview with members of the Ministry of Foreign Affairs, 2019

4. Summer Holiday Social Practices in Taicang: Taicang "Colorful Summer" Summer School, 2018

- INTERESTS

1.Video clip. Best Screenplay and Second Award of "Lab safety" Micro Video Solicitation, Wuhan University, 2018.11

2.Playing the guitar.

3.Basketball, Billiards (Member of Billiards Association of Wuhan University), Table tennis, Soccer, Badminton.

# 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与深度挖掘研究** | **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。