

Siyang (Suyo) Quan

Phone: +44 7422515353 / +1 3439890281 / +86 18319536710

Email: siyang.quan.2023@alumni.ucl.ac.uk / 1810108407@pku.edu.cn

Personal website: <https://suyotsuy.com/> / <https://suyoatucl.github.io/>

EDUCATION

Department of Information Studies, University College London **London, United Kingdom**

Master of Arts in Digital Humanities

Sep 2023 – Sep 2024

Weighted Average: 74.71/100 (Distinction)

GPA: 4.00/4.00, Rank: 1/31

Key Courses: NLP and Text Analysis, Data Visualization and GIS, Programming and Scripting (JS, Java), Internet Technologies (HTML, CSS), Digitisation

Thesis Topic: Digital Tools for Linguistic Empowerment: A Decentralised Framework for Real-Time Language Documentation and Automatic Classification

Department of History, Peking University

Beijing, China

Bachelor of History

Aug 2019 – Jul 2023

Weighted Average: 85.43/100

GPA: 3.57/4.00, Rank: 14/83

Key Courses: Digital Methods of History, Historical Linguistics, Python and Artificial Intelligence in the Humanities

Thesis Topic: A Comparative Study of the Construction of "Fascist" Discourse in the Chinese and British Mainstream Media in 1978: A Critical Discourse Analysis with the *People's Daily* and *The Times* as Examples

Health Science Center, Peking University

Beijing, China

Bachelor of Science

Aug 2018 – Jul 2023

Weighted Average: 82.54/100

GPA: 3.08/4.00, Rank: 17/90

Key Courses: Biostatistics, Advanced Mathematics, Python for Web Scraping, Clinical Big Data and Applications

Thesis Topic: Study on the Group of Male Nursing Staff in Beijing during the Early Period of the People's Republic of China

PUBLICATIONS

Quan, S. (2022). Historical Evolution of Anti-epidemic Measures in Republican Guangzhou: A Textual Analysis Based on Historical Resources. In Shelley, M., Akerson, V. & Sahin, I. (Eds.), *Proceedings of International Conference on Social and Education Sciences 2022* (pp. 251-281). ISTES Organization.

INDEPENDENT WORK

GooGan: A Decentralised Framework for Real-Time Language Documentation and Automatic Classification

<https://suyoatucl.github.io/GooGan/>

React Web Application

- **This project supports the decentralisation of language data storage, empowering non-linguists to participate in language preservation, aligning with ethical guidelines for linguistic research.**
- Developed a React-based web application as a demo of the master's thesis topic, enabling real-time language documentation, phoneme transcription, and automatic classification for low-resource languages, with cross-platform functionality.
- Implemented decentralised database technologies (OrbitDB and IPFS) alongside Google Cloud to support large-scale public data management, peer-to-peer resource sharing, and decentralised language data storage, achieving storage of 10 GB with a retrieval time of 5000 ms.
- Designed language comparison tools, including automatic cognate detection, linguistic distance calculation, and phylogenetic tree generation, using LingPy and Phylotree.js.
- Integrated Azure Speech Studio and Allosaurus for phoneme transcription and real-time processing.
- Conducted usability testing to enhance interface design and functionality, optimising user experience and data processing workflows.

XiaoChaGuan: AI-Powered Multilingual Fact-Checking Browser Extension

<https://github.com/suyoatucl/XiaoChaGuan/>

Chrome/Edge Extension (Ongoing)

- **Developing an intelligent Chrome/Edge extension enabling users in censorship-restricted environments to perform cross-lingual fact verification on social media content, addressing the issue of misrepresentation.**

- Key Technologies: TypeScript, Svelte, Plasmio Framework, LLM APIs (Claude/GPT-4), RAG architecture, IndexedDB, Shadowsocks protocol, sentence-transformers, jieba NLP.
- Core Features: Automated claim detection with NLP-based real-time scanning across Chinese social media platforms; multi-layered network solution with Shadowsocks/VMess/Tor integration achieving 95%+ uptime; RAG-powered cross-lingual verification pipeline combining semantic search, multi-source retrieval, and LLM analysis for 10+ languages; intelligent IndexedDB-based offline-first caching architecture reducing API calls.

Pollution is Colonialism (Game)

<https://suyoatuecl.github.io/PIC/> ; <https://yumisuyoworkshop.itch.io/pic>

JavaScript-based Interactive Fiction Game for the Web

- **This project explores the intersection of digital storytelling and humanities themes, serving as an innovative model for digital humanities research and pedagogy.**
- Adapted from: Liboiron, M. (2021). *Pollution is colonialism*. Duke University Press.
- Initially designed as an educational tool for digital humanities coursework, aiming to engage players with environmental issues from a humanities perspective, garnering 1,000+ users.
- Created a web-based game using JavaScript and Ruby, guiding players through four perspectives (settlers, Indigenous people, scientists, and government) to explore human-land relationships and perceptions of pollution.

Qióngyá Fāngyán (Online Dictionary of Hainan Min)

<https://c.c1nd.cn/HW6Bw>

WeChat Mini Program

- **An audible dictionary specifically designed for Hainanese speakers unfamiliar with the International Phonetic Alphabet.**
- Used WeChat open platform, Bmob development framework, LayUI, WXML, JavaScript and other front-end development technologies for the development of the system that consisted of a small program body for dictionary users and a management end of the backend maintainer.
- Built a Hainanese speech database on the backend using Speech Studio and Allosaurus.

ACADEMIC ACTIVITIES

Big Data Analysis Platform for Classical Literature Project

Sep 2021 – Mar 2022

Member of Professor Wang Jun's Team at the Center for Digital Humanities, Peking University

- Project: Big Data Analysis Platform for Classical Documents (<https://reuse.pkudh.org/>)
- Responsible for backend development with Django and frontend development with React and Vue.js.
- Developed visualizations of intertextuality and rhetorical devices in the *Analects of Confucius*, enabling users to explore textual relationships.

Collection and Digitization Project of Oral History Materials and Dialect Archives in Lingbei Township, Suixi County

Mar 2020 – Jun 2020

Student of Department of History, Peking University

- Conducted field research, collected oral history data of several families since the late 19th century.
- Digitised audio and text records onto a React-based platform to support long-term archival storage.

Fieldwork and Resource Collection for Jizhao Language

Jul 2020 – Aug 2020

Member of the Study Group from Dept. of Minority Languages and Literatures, Minzu University of China

- Conducted fieldwork and interviews with native speakers of Jizhao, a critically endangered Kra-Dai language spoken by few elderly individuals.
- Collected Swadesh's 200 core vocabulary items using professional recording equipment; data were submitted to the China Language Resources Preservation Project (<https://zhongguoyuyan.cn/point/62868>).

Information Collection and Sharing Online Platform Project

Sep 2020 – Nov 2020

Member of Professor Wang Guoping's Team, School of Information Science and Technology, Peking University

- Collected information from major websites and compiled them into one navigational site.
- Participated in the front and back-end development of websites using Python tools such as Beautiful Soup, Ajax, and MySQL databases.

PROFESSIONAL EXPERIENCES

Journal of Open Humanities Data, Ubiquity Press, London, United Kingdom

Aug 2024 –

Junior Editor, Editorial Assistant

- Participated in the “JOHD in China” project, contributing to meetings and project tasks as required.

- Provide backend operations and database maintenance services for the JOHD website, designing corresponding repository API options for data paper submitters across different disciplines and regions.

MiHoYo Co., Ltd., Shanghai, China

May 2025 – Aug 2025

NLP Research Intern (Fixed-Term Intern)

- Explore cutting-edge research in NLP, including data construction, model training, intelligent agents, dialogue systems.

Journal of Open Humanities Data, Ubiquity Press, London, United Kingdom

May 2024 – Jun 2024

Junior Data Analyst (Placement Offered by Dr Barbara McGillivray, King's College London)

- Crawled the journal's website to extract citations and usage metrics.
- Built an algorithm to identify the most popular JOHD articles, and compiled lists of top articles based on various metrics.
- Analysed citation and usage data with R and Python scripts to identify trends by geographic region, discipline, and over time, producing visualisations and a comprehensive report.

Appen Limited, Beijing, China

Oct 2021 – Sep 2022

AI Translation Specialist (Remote)

- Evaluated AI translations from Cantonese/Mandarin/Japanese/Kazakh to English, by assessing text and assigning semantic similarity scores.
- Quantified the accuracy, fluency, and overall quality of online messages that have been translated by machines.

Baidu Cloud Computing Technology (Beijing) Co., Beijing, China

Mar 2022 – Jun 2022

NLP (Natural Language Processing) Intern

- Developed visual modeling components and operators for Baidu's full-featured AI development platform (BML).
- Improved the entity analysis operator, lexical analysis operator, and Chinese DNN language modeling operator to enhance the system for better language understanding, resulting in an average reduction in processing time of more than 70%.

National Library of China, Beijing, China

Jan 2022 – Mar 2022

Document Digitization Intern, Center for Preservation and Protection of Internet Information Resources

- Project: Resource Base of Ancient Chinese Books (<http://read.nlc.cn/thematDataSearch/toGujiIndex>)
- Designed and optimized the information retrieval system for literature resources.
- Used Scrapy, a web-crawling tool, to extract online literature materials and established text indexing and the search engine framework through PyLucene, reducing search response time by 50%.
- Optimized the search response time of the repository and improved the comprehensiveness and accuracy when searching with keywords in classical or early modern Chinese.

EXTRACURRICULAR ACTIVITIES

- Editor for the "Linguistic Geography" section of the website Gu Yin Xiao Jing (kaom.net) (ID: MISAKA)
- Content creator of Zhihu, a Q&A Platform, to regularly update historical linguistic and archaeographic articles on Min Chinese and ancient Sino-Tibetan languages, with up to 1,000 followers (<https://www.zhihu.com/people/triehho>)
- Person in Charge, Beijing, "Investigation Mission of School of Basic Medical Sciences to Beijing, Shanxi, Sichuan, Guangdong and Guangxi to resume Work and Production" (Won the 1st Prize of Summer Social Practice of Peking University Health Science Center)
- Head of New Media Agency, Publicity Department of the Student Union, Peking University Health Science Center (2019-2020)
- Participant of the International Seminar on Digital Humanities of East Asian Classics (online) held by the Institute of Chinese Language Education, Dankook University, Korea (05.2021)

CERTIFICATIONS, AWARDS & HONORS

- | | |
|---|---------|
| ▪ Oracle AI Vector Search Certified Professional | 10/2025 |
| ▪ Oracle Cloud Infrastructure 2025 Certified Generative AI Professional | 10/2025 |
| ▪ Faculty of Arts & Humanities Dean's List (2024), University College London | 02/2025 |
| ▪ Luye Scholarship in Peking University Health Science Center for the 2021-2022 Academic Year | 09/2022 |
| ▪ Peking University Academic Excellence Award for the 2021-2022 Academic Year | 09/2022 |
| ▪ Outstanding Camper, Summer Camp for Outstanding Undergraduate held by School of Humanities, Tsinghua University (Rank: 14/62) | 08/2022 |

- Outstanding Participant, Summer Camp for Chinese History held by School of Humanities, Shanghai Jiao Tong University (Rank: 8/23) 08/2022
- **Honorable Mention, American College Mathematical Modeling Competition (MCM/ICM)** 05/2022
- Third Prize, the 30th Peking University Challenge Cup (2021-2022): Research on the Historical Evolution of Anti-epidemic Measures in Republican Guangzhou: A Quantitative Analysis 05/2022
- **Third Prize, 11th Asia and Pacific Mathematical Contest in Modeling (APMCM)** 02/2022
- Third Prize, the 29th Peking University Challenge Cup (2020-2021): Knowledge Mapping Analysis based on Text Mining 05/2021
- Peking University Social Work Award for the 2019-2020 Academic Year 09/2020
- First Prize, 2020 Summer Social Practice Competition of Peking University Health Science Center: "Resumption of Work and Production of Enterprises in Beijing, Shanxi, Sichuan, Guangdong and Guangxi: A Survey from the Field-Study Group" 08/2020
- First Prize, Peking University Medical School Student Innovation Experimental Project (2019-2020) & 2020 National Student Innovation and Entrepreneurship Training Program 06/2020

LANGUAGES SPOKEN

Highly Proficient: English (TOFEL 112, IELTS 7.5), Chinese (Hokkien, Hakka, Cantonese, Mandarin)

Proficient: Japanese (JLPT N2), Kazakh

Basic: German (TDN 3), French (DELFI A2), Ancient Greek, Latin, Ancient Egyptian

TECHNICAL SKILLS AND OTHERS

Programming & Mark-up: Python, R, JavaScript/TypeScript, Java, Ruby, C#, C++, MATLAB, PHP, SQL, XML, HTML, CSS; Regex

Data Science: Pandas, NumPy, SciPy, scikit-learn, PyTorch, TensorFlow, Keras

Information Retrieval: Lucene/PyLucene; classical IR (BM25, query expansion); dense retrieval (bi-encoders); reranking (cross-encoders); Oracle AI Vector Search

NLP & Language Models: Hugging Face Transformers, BERT, NLTK, spaCy, Word2Vec, SentencePiece

LLM Training & Optimisation: OCI Generative AI, Ollama, DeepSpeed, PEFT (LoRA/QLoRA), bitsandbytes, Accelerate, vLLM

LLM Evaluation & Design: Ragas, prompt design, automated evaluation (nDCG, MRR, faithfulness checks)

Data Acquisition & Web: BeautifulSoup, Scrapy; REST APIs, FastAPI

Databases & Storage: MongoDB, MySQL, PostgreSQL; OrbitDB/IPFS; object storage & artefact versioning

Cloud & MLOps: Docker; CI-driven evaluation (GitHub Actions); Oracle Cloud Infrastructure; remote training/evaluation on cloud VMs (Azure, GCP, AWS)

Visualisation & Front-end: React, D3/Phyloree.js, Plotly, Matplotlib, Seaborn

Geospatial & Research Analytics: ArcGIS, QGIS, Gephi; NVivo; SPSS, Stata, SAS

Tooling & Workflow: Git/GitHub, Linux, Unix shell, Powershell, LaTeX/Overleaf, Jupyter/Notebooks, Microsoft Office (Excel VBA, Word, PowerPoint)

Design & Interactive Media: Adobe Creative Suite (Photoshop, After Effects, Audition, Premiere Pro), AutoCAD, 3ds Max, Unity, Unreal Engine, Figma

Diving Skill: PADI OW (Open Water Diver)

More skills being learnt ...