

Siyang (Suyo) Quan

71 Alie Street, Aldgate, London E1 8NF | +44 7422515353 | siyang.quan.2023@alumni.ucl.ac.uk | <https://suyotsuy.com/>

SUMMARY OF QUALIFICATIONS

A digital expert in the fields of humanities and medical science. Advanced Mandarin/English/Cantonese/Japanese/Kazakh fluency with experience in evaluating AI translations and analyzing data. Proven attention to detail through frontend development, backend operations and database maintenance. Skilled in Python and R for data analysis. Demonstrated ability to concentrate on complex tasks, optimizing information-retrieval pipelines and developing AI models. Aspiring to contribute to the digitalization of languages and heritage, as well as the construction of multimodal knowledge bases.

KEY SKILLS

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

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

Information Retrieval & Web: Lucene/PyLucene; classical IR (BM25, query expansion); dense retrieval; Beautiful Soup, Scrapy; REST APIs, FastAPI

NLP & LLM Ecosystem: Transformers, BERT, NLTK, spaCy, Word2Vec, SentencePiece, PEFT (LoRA/QLoRA), Accelerate, vLLM, Ragas

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

Databases & Storage: MongoDB, MySQL, PostgreSQL, IPFS

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

Languages Spoken: English (fluent); Mandarin/Cantonese/Hokkien Chinese, Kazakh (native); Japanese, German, French (advanced)

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 IConSES 2022* (pp. 251-281). ISTES Organization.

PROFESSIONAL EXPERIENCES

Junior Editor & Editorial Assistant, Journal of Open Humanities Data, Ubiquity Press, London, United Kingdom May 2024 –

- 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.
- Analysed citation and usage data with R and Python scripts to identify trends by geographic region, discipline, and over time, producing visualisations and comprehensive reports.

AI Translation Specialist (Remote), Appen Limited, Beijing, China

Oct 2021 – Sep 2022

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

NLP Analyst Intern, Baidu Cloud Computing Technology Co., Beijing, China

Mar 2022 – Jun 2022

- 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%.

Document Digitization Intern, Center for Preservation and Protection of Internet Information Resources, National Library of China, Beijing, China

Jan 2022 – Mar 2022

- Optimised an information-retrieval pipeline for literature resources, building web crawling with Scrapy and text indexing/search with PyLucene.
- Reduced search latency by ~50% and improved keyword search coverage and accuracy, particularly for classical/early-modern Chinese, via index tuning and ranking adjustments.

PROJECT EXPERIENCES

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

React Web Application | <https://suyoatocl.github.io/GooGan/>

- React demo of my master's thesis enabling real-time language documentation, phoneme transcription and automatic classification for low-resource languages, with community-centred and decentralised design.
- Implemented OrbitDB+IPFS with GCP for peer-to-peer storage and public data management; shipped comparison tools (cognate detection, distance, phylogenetic trees); handled ~10 GB with ~5 s retrieval.

Pollution is Colonialism (Game)

JavaScript-based Interactive Game for the Web | <https://suyoatocl.github.io/PIC/> ; <https://yuminisuyoworkshop.itch.io/pic>

- Created a web-based game using JavaScript and Ruby, simulating the science-nature-indigenous intersection on pollution issues.
- 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.

Big Data Analysis Platform for Classical Literature Project

Member of Professor Wang Jun's Team at the Center for Digital Humanities, Peking University | <https://reuse.pkudh.org/>

- Responsible for backend development with Django and frontend development with React and Vue.js.
- Developed user-centered visualizations of intertextuality and rhetorical devices in the *Analects of Confucius*.

KEY CERTIFICATIONS, AWARDS & HONORS

Oracle AI Vector Search Certified Professional & Oracle Cloud Infrastructure 2025 Certified Generative AI Professional

10/2025

Faculty of Arts & Humanities Dean's List (2024), University College London

02/2025

Honorable Mention, American College Mathematical Modeling Competition (MCM/ICM)

05/2022

Third Prize, 11th Asia and Pacific Mathematical Contest in Modeling (APMCM)

02/2022

EDUCATION

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

Sep 2023 – Sep 2024

Master of Arts in Digital Humanities

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

Department of History & Health Science Center, Peking University, Beijing, China

Aug 2018 – Jul 2023

Bachelor of History in World History & Bachelor of Science in Health Science (Concurrent Dual Degree) GPA:3.57/4.00, Rank: 14/83

Key Courses: Digital Methods of History, Historical Linguistics, Python and Artificial Intelligence, Biostatistics, Advanced Mathematics, Python for Web Scraping, Clinical Big Data and Applications

Click [here](#) to view my complete resume!