Ren Yi

yi.ren@epfl.ch | +41766717387 | Lausanne

renyi.ch | github.com/nmrenyi | linkedin.com/in/renyich

Education

EPFL Aug 2024 - Jun 2027
Master in Digital Humanities

Tsinghua University Aug 2022 - Jul 2024
PhD Program in Information Systems (withdrew to pursue MSc at EPFL), GPA: 4.00/4.00

Tsinghua University Aug 2018 - Jul 2022
Bachelor's Degree, Computer Science and Technology, GPA: 3.89/4.00

Skills

AI/ML: Large Language Models (application and evaluation), Retrieval Augmented Generation (RAG), on-device language model deployment, recommender systems, data analysis

Programming: Python (strong), C++ (strong), Java, LaTeX (strong)

Tools: Figma, Adobe Photoshop, Adobe Premiere Pro, MS Office

Languages: English (advanced), French (working proficiency), Chinese (native)

Research Experiences

Quantifying the Invisible: Women Doctors in Rosenwald Guides Aug 2025 - Feb 2026
[Blog] [Report] [Preprint] [Annotation] [Extraction] [Benchmark] [Dataset]

- Collaborated with historians from the Laboratory for the History of Science and Technology (EPFL) and the Institut des humanités en médecine (Lausanne).
- Designed a double-triangular annotation framework combining LLMs and human labeling for scalable and reliable annotation, achieving >50% reduction in human annotation effort while maintaining >99% accuracy. **Preprint: arXiv:2605.25781**.
- Built a benchmark dataset of 2,600+ doctor entries from the Rosenwald Guides using the proposed framework to support evaluation of structured information extraction from historical documents.
- Compared four extraction paradigms on the benchmark and applied the best-performing approach (image + OCR text input) to extract 577,000+ doctor and health official records from the Rosenwald Guides (1887–1922).
- Discovered 3,700+ records of female doctors (1887–1922), enabling future study of women's medical practice.

On-device medical chatbot for nurse-midwives, *D-tree*, Zanzibar, Tanzania Feb 2026 - Present
[Code] [Demo Video] [Dataset]

- Deployed Google Gemma 4 E4B IT (int4) on edge devices to provide offline medical information for midwives and nurses in low-resource settings (limited Internet connectivity)
- Used Retrieval Augmented Generation (RAG) techniques on official midwife training materials to enhance the system's ability to provide accurate and locally feasible information
- Curated an open OBGYN QA dataset of 20,500+ questions by filtering five public medical benchmarks (AfriMed-QA, MedMCQA, Kenya Clinical Vignettes, MedQA-USMLE, Women's Health Benchmark) using existing tags and Gemini 3 Flash Preview for question classification
- Benchmarked the on-device model against leading models (e.g. GPT-5) on the curated OBGYN dataset to evaluate performance gaps and identify areas for improvement
- Improving RAG quality and preparing user evaluation studies with nurse-midwives in Zanzibar

Course Projects & Internships

Sentiment Contagion on Social Media: A Case Study on Reddit, EPFL [Report] Mar 2025 - Jun 2025

- Collaborated with 3 members to analyze over 6.4 million comments and 1 million posts from Reddit's r/unpopularopinion using RoBERTa-based sentiment classification with manual validation and linear regression
- Confirmed statistically significant emotional contagion on Reddit: post sentiment significantly influences comment sentiment

Design prototype for true cost of food, EPFL [Demo Video][Report] Sept 2024 - Dec 2024

- Collaborated with 3 members to develop a design prototype with Figma, offering recipe-based food recommendation with low true cost of food to users
- Promoted environmentally friendly food culture, and received the highest grade among all designs in the track

Research Intern, Idle Fish (Alibaba Group) May 2024 - Jun 2024

- Applied LLMs to conversational agents for customer service, enabling automated buyer-seller bargaining on a used product transaction platform
- Processed real-world customer conversation data to support model development and evaluation

Publication

Double Triangle Annotation: A Scalable Human-in-the-Loop Framework for High-Precision Historical Document Annotation [Preprint] [Annotation] [Benchmark] 2026

Yi Ren (arXiv preprint, 2026)

- Designed an annotation framework around a consensus principle, applied recursively at two layers—between two independent MLLMs (Claude, Qwen, Llama, Grok), then between two human-in-the-loop systems built from them—escalating only disagreements to a human jury
- Achieved a Word Error Rate of 0.003 on the Rosenwald Guides, with >85% of fields auto-accepted by model consensus—cutting total human review workload by more than 50% compared to a full manual annotation by one annotator
- Released the first structured-extraction benchmark for the Rosenwald Guides (13,595 annotated fields across 60 columns) to support future work on historical document processing

Honors & Awards

Outstanding Graduate of Computer Science and Technology Department at Tsinghua University	2022
National Scholarship (Top 2%)	2021
Excellent Student Leader at Tsinghua University	2020
Passion For Reading Scholarship at Tsinghua University	2019, 2020

Social Work

Student Representative & Member of Teaching Committee, Digital Humanities, EPFL [Survey] Sept 2024 - Aug 2025

- Advocated for anti-discrimination principles and equal opportunities regarding concerns over security screenings in the admissions process for applicants of certain nationalities. Led efforts to gather student perspectives through a survey that reached over 800 respondents
- Collected student feedback on courses, career support, and overall program experiences; communicated findings to the section staff to help improve program design

Olympic Family Assistant, Beijing 2022 Winter Olympics & Paralympics Jan 2022 - Apr 2022

- Assisted Director General of International Skating Union and President of European Paralympic Committee
- Created a welcoming environment for clients to work in China; assisted with itinerary planning and translation