Ren Yi 任一

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

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

Education

EPFL Aug 2024 - Jun 2028
Master of Science (MSc) in Digital Humanities

Tsinghua University Aug 2022 - Jul 2024
PhD in Recommender Systems (withdrew for 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

Research Experiences

Quantifying the Invisible: Women Doctors in Rosenwald Guides Aug 2025 - Feb 2026
[MEDIF Blog] [Project Report] [Annotation Preprint] [Annotation Harness]
[Annotated Benchmark] [Extraction Harness] [Extracted Records]

- 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from the Rosenwald Guides using the proposed framework to support evaluation of structured information extraction from historical documents.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 - June 2026
[App Code] [System Report] [Benchmark Report] [Live Demo] [Video Demo]

- Built a fully offline, on-device medical RAG assistant—an EmbeddingGemma-300M retriever and a Google Gemma 4 E4B (int4) generator running on commodity Android—that answers nurse-midwives' clinical questions with cited guidance, requiring no connectivity and keeping all patient data on the device
- Curated a RAG guideline corpus of 87 authoritative documents (WHO, NICE, MSF, and Zanzibar/Tanzania national guidelines), chunked into 63,650 on-device passages
- Designed and ran a layered evaluation of the system, covering end-to-end answer quality, the retriever and the generator each measured in isolation, and on-device latency across input lengths and CPU/GPU backends
- Built an open evaluation harness over two benchmarks (mamabench and mamaretrieval), scored by an LLM judge validated against physician rubrics from HealthBench

Papers

Double Triangle Annotation: A Scalable Human-in-the-Loop Framework for High-Precision Historical Document Annotation [Annotation Preprint] [Annotation Harness] [Annotated Benchmark] 2026

Ren Yi (arXiv preprint arXiv:2605.25781, 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

MAM-AI: An On-Device Medical Retrieval-Augmented Generation System for Nurses and Midwives in Zanzibar [System Report] [App Code] [Eval Harness] [Live Demo] [Video Demo] 2026

- A redesigned system prompt is a major lever on answer quality—cutting deflective answers that only said "refer to a senior doctor" from 32.7% of questions to 3.2%, while roughly doubling key-fact recall from 0.139 to 0.279 and raising potentially-harmful answers only from 12.8% to 15.7%
- On-device retrieval is strong, but ranking is not the lever: on our retrieval benchmark the 300M on-device embedder ranks among the cloud retrievers, yet a better embedder or a fine-tuned reranker leaves end-to-end answers unchanged—the larger levers on answer quality are the generator and corpus coverage and quality
- The generator matters most, and faithfulness decides deployment: answer quality rises with model size before plateauing, while quantization and retrieval barely move it; among two same-size on-device candidates, the more faithful but less helpful Gemma 4 E4B is deployed over Gemma 3n E4B—a safety-versus-usefulness choice

mamabench and mamaretrieval: Benchmarks for Evaluating Medical Retrieval-Augmented Generation in Maternal, Neonatal, and Reproductive Health [Benchmark Report] [mamabench] [mamaretrieval] 2026

- Assembled both benchmarks—mamabench (25,949 QA items) and mamaretrieval (3,185 retrieval queries)—by scoping and filtering expert-authored sources (licensing exams, physician panels, guidelines) rather than authoring new clinical questions
- Decomposed clinical relevance into a graded, multi-dimensional rubric (0–6)—a topical gate over three graded dimensions: meaningful content, actionable guidance, and density—rather than a binary relevant/not-relevant label, separating strong retrieval from merely adequate retrieval
- Measured and disclosed the limits of the LLM-assigned labels—via scope-classifier agreement, a frontier-judge check, and a pooling-completeness audit—rather than presenting the benchmarks as an oracle

Internship

Securing LLM-enhanced human-machine interfaces, *Digiinov SA*, Yverdon-les-Bains, Switzerland Aug 2026 - Jan 2027

- Internship at Digiinov, with academic guidance from the EPFL Center for Digital Trust (C4DT)
- Building conversational agents for Digiinov HMI, a human-machine interface (HMI) platform for machines and instruments, giving operators an interactive way to obtain information and debug
- Developing sequence generation for Digiinov Inspect, control software for test and measurement equipment, so that examining sequences for a product can be written in plain language

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 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 [Presentation]
- Collected student feedback on courses, career support, and overall program experiences; communicated findings to the section staff to help improve program design

Skills

AI/ML & Programming: Large Language Models (application and evaluation), Retrieval Augmented Generation (RAG), on-device language model deployment, data analysis; Python (strong), C++ (strong), LaTeX (strong)

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

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