Ren Yi 任一

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

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

Formation

EPFL août 2024 - juin 2028
Master ès sciences (MSc) en humanités digitales

Tsinghua University août 2022 - juil. 2024
Doctorat en systèmes de recommandation (retrait pour Master à l'EPFL),
GPA : 4.00/4.00

Tsinghua University août 2018 - juil. 2022
Licence, informatique et technologie, GPA : 3.89/4.00

Expériences de recherche

Quantifier l'invisible : Les femmes médecins dans les Guides Rosenwald août 2025 - fév. 2026
[Blog MEDIF] [Rapport de projet] [Prépublication annotation]
[Harnais d'annotation] [Benchmark annoté] [Harnais d'extraction]
[Enregistrements extraits]

- Collaboration avec des historien·ne·s du Laboratory for the History of Science and Technology (EPFL) et de l'Institut des humanités en médecine (Lausanne).
- Conception d'un cadre d'annotation « double triangle » combinant LLMs et annotation humaine pour une annotation scalable et fiable, réduisant de >50% l'effort d'annotation humaine tout en maintenant une précision >99%. **Prépublication : arXiv:2605.25781**.
- Construction d'un jeu de données de référence de 2 600+ entrées de médecins des Guides Rosenwald à l'aide du cadre proposé, pour évaluer l'extraction d'informations structurées à partir de documents historiques.
- Comparaison de quatre paradigmes d'extraction sur le benchmark et application de la meilleure approche (image + texte OCR) pour extraire 577 000+ entrées de médecins et d'officiels de santé des Guides Rosenwald (1887–1922).
- Découverte de 3 700+ entrées de femmes médecins (1887–1922), permettant de futures études sur la pratique médicale féminine.

Agent conversationnel médical embarqué pour les sages-femmes, *D-tree*, Zanzibar, Tanzanie fév. 2026 - juin 2026
[Code de l'application] [Rapport système] [Rapport benchmark] [Démo en ligne]
[Vidéo de démo]

- Développement d'un assistant médical RAG entièrement embarqué et hors ligne—un système de recherche EmbeddingGemma-300M et un générateur Google Gemma 4 E4B (int4) tournant sur Android grand public—répondant aux questions cliniques des sages-femmes avec des réponses sourcées, sans connexion et en conservant toutes les données patient sur l'appareil
- Constitution d'un corpus de directives pour le RAG de 87 documents de référence (OMS, NICE, MSF et directives nationales de Zanzibar/Tanzanie), découpé en 63 650 passages embarqués
- Conception et conduite d'une évaluation par couches du système, couvrant la qualité des réponses de bout en bout, le système de recherche et le générateur mesurés isolément, et la latence embarquée selon la longueur d'entrée et le backend CPU/GPU
- Développement d'un harnais d'évaluation ouvert sur deux benchmarks (mamabench et mamaretrieval), notés par un juge LLM validé contre les grilles de médecins de HealthBench

Articles

Double Triangle Annotation: A Scalable Human-in-the-Loop Framework for High-Precision Historical Document Annotation 2026
[Prépublication annotation] [Harnais d'annotation]
[Benchmark annoté]

Ren Yi (prépublication arXiv arXiv:2605.25781, 2026)

- Conception d'un cadre d'annotation autour d'un principe de consensus, appliqué récursivement sur deux niveaux—entre deux MLLMs indépendants (Claude, Qwen, Llama, Grok), puis entre deux systèmes human-in-the-loop construits sur ces modèles—n'escaladant que les désaccords vers un jury humain
- Word Error Rate de 0.003 atteint sur les Guides Rosenwald, avec >85% des champs auto-acceptés par consensus inter-modèles—réduisant la charge totale de relecture humaine de plus de 50% par rapport à une annotation manuelle complète par un seul annotateur
- Publication du premier benchmark d'extraction structurée pour les Guides Rosenwald (13 595 champs annotés sur 60 colonnes) pour soutenir les futurs travaux sur le traitement de documents historiques

MAM-AI: An On-Device Medical Retrieval-Augmented Generation System for Nurses and Midwives in Zanzibar [Rapport système] [Code de l'application] [Harnais d'évaluation] [Démo en ligne] [Vidéo de démo] 2026

- Un prompt système repensé est un levier majeur sur la qualité des réponses—réduisant les réponses évasives se limitant à « consulter un médecin senior » de 32,7% des questions à 3,2%, tout en doublant approximativement le rappel des faits clés de 0,139 à 0,279 et en n'augmentant les réponses potentiellement nuisibles que de 12,8% à 15,7%
- La recherche embarquée est performante, mais le classement n'est pas le levier : sur notre benchmark de recherche, l'embedder embarqué de 300M se classe parmi les systèmes de recherche cloud, mais un meilleur embedder ou un reranker affiné ne modifie pas les réponses de bout en bout—les leviers les plus importants pour la qualité des réponses sont le générateur ainsi que la couverture et la qualité du corpus
- Le générateur compte le plus, et la fidélité décide du déploiement : la qualité des réponses croît avec la taille du modèle avant de plafonner, tandis que la quantification et la recherche n'y changent presque rien ; parmi deux candidats embarqués de même taille, Gemma 4 E4B, plus fidèle mais moins serviable, est déployé plutôt que Gemma 3n E4B—un arbitrage entre sécurité et utilité

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

- Assemblage des deux benchmarks—mamabench (25 949 éléments de QA) et mamaretrieval (3 185 requêtes de recherche)—en cadrant et filtrant des sources rédigées par des experts (examens de licence, panels de médecins, directives) plutôt qu'en rédigeant de nouvelles questions cliniques
- Décomposition de la pertinence clinique en une grille graduée et multidimensionnelle (0–6)—un filtre thématique sur trois dimensions graduées : contenu pertinent, conseils actionnables et densité—plutôt qu'une étiquette binaire pertinent/non pertinent, distinguant une recherche solide d'une recherche simplement adéquate
- Mesure et divulgation des limites des étiquettes attribuées par LLM—via l'accord du classifieur de portée, une vérification par un modèle de pointe et un audit de complétude du pooling—plutôt que de présenter les benchmarks comme une vérité absolue

Stage

Sécuriser les interfaces homme-machine augmentées par LLM, *Digiinov SA*, Yverdon-les-Bains, Suisse août 2026 - janv. 2027

- Stage chez Digiinov, avec un encadrement académique du Center for Digital Trust (C4DT) de l'EPFL
- Développement d'agents conversationnels pour Digiinov HMI, une plateforme d'interface homme-machine (HMI) pour machines et instruments, offrant aux opérateurs un moyen interactif d'obtenir des informations et de déboguer
- Développement de la génération de séquences pour Digiinov Inspect, logiciel de contrôle pour équipements de tests et mesures, afin que les séquences d'examen d'un produit puissent être rédigées en langage naturel

Distinctions & prix

Diplômé·e remarquable du département d'informatique (Tsinghua University)	2022
Bourse nationale (top 2%)	2021
Excellent·e responsable étudiant·e (Tsinghua University)	2020
Bourse « Passion pour la lecture » (Tsinghua University)	2019, 2020

Engagement associatif

Représentant·e des étudiant·e·s et membre de la commission d'enseignement, section Humanités digitales, EPFL sept. 2024 - août 2025

- Plaidoyer pour la non-discrimination dans le processus d'admission pour les étudiant·e·s de certaines nationalités ; conduite d'une enquête auprès de plus de 800 répondant·e·s [Présentation]
- Collecte des retours des étudiant·e·s sur les cours, l'accompagnement carrière et l'expérience globale du programme ; transmission des résultats aux responsables de section pour améliorer la conception du programme

Compétences

IA/ML & Programmation : Large Language Models (application et évaluation), Retrieval Augmented Generation (RAG), déploiement de modèles de langage sur appareil, analyse de données ; Python (fort), C++ (fort), LaTeX (fort)

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

Langues : anglais (avancé), français (niveau professionnel), chinois (langue maternelle)