

# Ren Yi

yi.ren@epfl.ch | +41766717387 | Lausanne, Suisse  
renyi.ch | github.com/nmrenyi | linkedin.com/in/renyich

## Formation

---

<b>EPFL</b> Master en humanités digitales	août 2024 - juin 2027
<b>Tsinghua University</b> Programme doctoral en systèmes d'information (retrait pour poursuivre un Master à l'EPFL) GPA : 4.0/4.0	août 2022 - juil. 2024
<b>Tsinghua University</b> Licence, informatique et technologie, GPA : 3.89/4.00	août 2018 - juil. 2022

## Compétences

---

**IA/ML** : Large Language Models (application et évaluation), Retrieval Augmented Generation (RAG), déploiement de modèles de langage sur appareil, systèmes de recommandation, analyse de données

**Programmation** : Python (fort), C++ (fort), Java, LaTeX (fort)

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

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

## Expériences de recherche

---

**Quantifier l'invisible : Les femmes médecins dans les Guides Rosenwald** août 2025 - fév. 2026  
[[Blog](#)] [[Rapport](#)] [[arXiv](#)] [[Annotation](#)] [[Extraction](#)] [[Benchmark](#)] [[Données](#)]

- 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 - présent  
[[Code](#)] [[Vidéo de démo](#)] [[Données](#)]

- Déploiement de [Google Gemma 4 E4B IT \(int4\)](#) sur des appareils embarqués pour fournir des informations médicales hors ligne aux sages-femmes et infirmier-ère-s dans des contextes à faibles ressources (connectivité Internet limitée)
- Application des techniques de Retrieval Augmented Generation (RAG) sur des supports officiels de formation des sages-femmes pour améliorer la précision et la pertinence locale des réponses
- Constitution d'un [jeu de données OBGYN ouvert](#) de 20 500+ questions en filtrant cinq benchmarks médicaux publics ([AfriMed-QA](#), [MedMCQA](#), [Kenya Clinical Vignettes](#), [MedQA-USMLE](#), [Women's Health Benchmark](#)) à l'aide des étiquettes existantes et de Gemini 3 Flash Preview pour la classification des questions
- Évaluation comparative du modèle embarqué face à des modèles de référence (ex. GPT-5) sur le jeu de données OBGYN afin d'identifier les écarts de performance
- Amélioration en cours de la qualité du RAG et préparation d'études d'évaluation à Zanzibar

## Projets de cours & stages

---

**Contagion des sentiments sur les réseaux sociaux : étude de cas sur Reddit, EPFL** [Rapport] mars 2025 - juin 2025

- Collaboration avec 3 membres ; analyse de 6,4 M de commentaires et 1 M de posts Reddit via classification de sentiment RoBERTa, validation manuelle et régression linéaire
- Confirmation d'une contagion émotionnelle statistiquement significative sur Reddit : le sentiment des posts influence significativement le sentiment des commentaires

**Prototype de design sur le « vrai coût » de l'alimentation (EPFL)** [Vidéo de démo] [Rapport] sept. 2024 - déc. 2024

- Collaboration avec 3 membres pour développer un prototype Figma proposant des recommandations de recettes à faible « vrai coût » alimentaire
- Promotion d'une culture alimentaire plus respectueuse de l'environnement ; meilleure note parmi tous les projets du parcours

**Stagiaire de recherche, Idle Fish (Alibaba Group)** mai 2024 - juin 2024

- Application de LLMs pour des agents conversationnels de service client, permettant la négociation automatisée acheteur-vendeur sur une plateforme de produits d'occasion
- Traitement de données réelles de conversations clients pour soutenir le développement

## Publications

---

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

Yi Ren (prépublication arXiv, 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

## 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** [Enquête] 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
- 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

**Assistant·e « Olympic Family », Jeux olympiques d'hiver & paralympiques Pékin 2022** janv. 2022 - avr. 2022

- Assistance au Directeur Général de l'Union internationale de patinage et au Président du Comité paralympique européen ; aide à la planification d'itinéraires, à la traduction et à l'accueil en Chine