RENQING CUOMAO

+41 76 262 80 40 ◇ renqing.cuomao@epfl.ch ◇ 10 February 2000
Chemin des Berges 12, Chavannes-près-Renens - Vaud, Switzerland

EDUCATION

Master in Computer Science, EPFL, Switzerland — Sept 2023 - present

Bachelor in Computer Science, Minzu University of China (MUC), Beijing, China — Sept 2017 - Jun 2021

SKILLS

Coding	Python, C/C++, Scala, SQL
ML	PyTorch, pandas, NumPy, transformers, scikit-learn, seaborn, matplotlib, UMAP, FastText, neural networks, High-Performance Computing (Todi, Clariden clusters)
Web Dev	HTML, CSS, JavaScript, Node.js, React.js
Relevant Coursework	Machine Learning, Modern Natural Language Processing, Data Visualization, Distributed Information Systems, etc
Languages	Tibetan (Native), Chinese (Fluent), English (Fluent)

ACADEMIC PROJECTS (SELECTED)

Optimizing Multilingual LLM Pretraining with Model-based Data Filtering — Sept 2024 - Jan 2025
Research Project on HORIZONTAL: LLMs @MLO, EPFL (under SwissAI initiative)

- Developed a model-based filtering framework for multilingual datasets using Transformer-, FastText- and LLM-based classifiers. Conducted ablation studies on FineWeb-2, demonstrating the potential for efficient LLM pretraining with only 15% of tokens while maintaining MMLU performance. Extended the framework to five languages and contributed to the curation of refined pre-training datasets.

Evaluating Textual Coherence Classification for Meditron LLM — Sept 2024 - Dec 2024
Machine Learning course, CS-433, EPFL

- Designed and evaluated classifiers to filter low-quality text for Meditron, a medical-domain LLM. Trained a coherence classifier achieving 87% accuracy with an 8% false positive rate, ensuring high-quality pretraining data and automated large-scale dataset processing to enhance model performance.

Alpaca-Tutor: An LLM-based Chatbot for Assisting STEM Students — Mar 2024 - Jun 2024
Modern natural language processing course, CS-552, EPFL

- Developed an LLM-powered chatbot for EPFL students, fine-tuning Llama-3 8B using LoRA and Direct Preference Optimization (DPO). Generated a synthetic dataset of preference pairs with OpenAI GPT 4-o and Llama 3 70B incorporating chain-of-thought prompting examples.

Document Retrieval and Recommender Systems — Sept 2024 – Dec 2024
Distributed Information Systems course, CS-423, EPFL

- Developed a multilingual document retrieval system using TF-IDF and BM25s from scratch, achieving 80.2% Recall@10 in validation and retrieving 2,000 queries in under 4 minutes on CPU. Built and benchmarked collaborative filtering, matrix factorization, and neural models (MLP, GMF, NeuMF) for book rating prediction; selected MLP for best RMSE (0.797) and efficiency.

WORK EXPERIENCE

Hainan Tibetan Information Technology Research Center — Jun 2024 - Aug 2024
Software Engineer intern for the Yongzin Search Engine, the Tibetan *Google*

- Developed Tibetan-language text embeddings and trained fastText/BERT-based classifiers, improving query understanding and search relevance by 25%, particularly for ambiguous and long-tail queries.
- Built a high-quality Tibetan text corpus for training NLP models, enhancing multilingual and cross-lingual information retrieval, significantly improving user experience for 30,000+ daily users.

MUC Innovation and Entrepreneurship Center — Dec 2018 - Sept 2020
Machine Learning & Computer Vision Engineer

- Curated a high-quality dataset by collecting and preprocessing 60,000 handwritten Tibetan characters, and fine-tuned CNN models with various architectures using TensorFlow + Keras to optimize recognition accuracy.
- Developed an interactive web interface for real-time model prediction visualization.

Software Engineering, MUC — Mar 2020 - Jun 2020
Teaching Assitant

- Assisted students with coding exercises, providing technical support and in-depth explanations for projects.

INTERESTS

In my free time, I enjoy reading, road biking, hiking, cooking, and exploring international art films.