

YIFU QIU

yifu.qiu@ed.ac.uk | yfqi.netlify.app | Google Scholar | +44 07873801958

PROFESSIONAL SUMMARY

Final-year PhD Researcher at University of Edinburgh/Cambridge (via ELLIS) specializing in **Foundation Model (LLMs & VLMs) Post-Training, hallucination detection and mitigation**, and understanding their **internalized world model**, with first-author publications in **ICLR, NeurIPS, ACL, EMNLP, and NAACL**. Proven industrial research experience/collaboration at **Meta FAIR, Apple AIML (with Apple AIML PhD Fellowship), Nvidia Research**, and **Baidu NLP**, contributing to the large-scale pre/post-training for an omnimodal models (7B+ params), long-context LLMs' post-training for retrieval and reasoning, and hallucination mitigation. Expert in scaling distributed training pipelines and designing novel alignment algorithms (SFT/RL).

EDUCATION

University of Edinburgh & University of Cambridge <i>PhD in Natural Language Processing (ELLIS Program)</i>	Oct 2022 – Jun 2026 (Expected) United Kingdom
<ul style="list-style-type: none">– Topic: Internalized World Models in Foundation Models (Advisors: Shay Cohen, Anna Korhonen, Edoardo Ponti).– Fully funded by Apple AIML PhD Fellowship and Informatics Global PhD Scholarship.– Selected for the ELLIS PhD program (< 5% acceptance rate).	
University of Edinburgh <i>MSc in Cognitive Science (NLP Specialism) – Distinction</i>	Sep 2020 – Aug 2021 United Kingdom
BNU-HKBU United International College <i>BSc in Computer Science – 1st Class Honours</i>	Sep 2016 – Aug 2020 China

RESEARCH EXPERIENCE

Research Scientist Intern <i>Meta FAIR (Fundamental AI Research)</i>	June 2025 – Sep 2025 Paris, France
Project: Omnimodal Foundation Models (Large Concept Model) Host: Dr. Holger Schwenk	
<ul style="list-style-type: none">– Developed vSONAR, a vision-language embedding space supporting 200+ text and 37 speech languages.– Introduced vLCM, a scalable post-training pipeline for the 7B Large Concept Model (LCM), utilizing 1M instruction tuning samples and 200K interleaved multimodal Chain-of-Thought data.– Achieved SOTA performance on multilingual captioning and QA, outperforming existing baselines in 61 out of 62 languages.– Paper: <i>Unified Vision–Language Modeling via Concept Space Alignment</i> (ICLR 2026).	
Research Scientist Intern <i>Apple AIML</i>	June 2024 – Sep 2024 Seattle, WA
Project: Long-Context Reasoning & Inference Optimization Hosts: Dr. Benjamin Han, Dr. Yizhe Zhang	
<ul style="list-style-type: none">– Proposed ICR², a benchmark for evaluating In-Context Retrieval and Reasoning in realistic scenarios.– Engineered Retrieval-Attention-Probing, an inference-time alignment technique using attention heads to de-noise long contexts during decoding.– Demonstrated +17% Exact Match gain on LOFT and +13% on ICR² using Mistral-7B, outperforming GPT-4-Turbo on specific tasks.– Paper: <i>Eliciting In-context Retrieval and Reasoning for Long-Context Language Models</i> (ACL 2025).	
Research Scientist Intern <i>Apple AIML</i>	June 2023 – Sep 2023 Seattle, WA
Project: Trustworthy Generation & Hallucination Mitigation Host: Dr. Benjamin Han	
<ul style="list-style-type: none">– Developed TWEAK, a decoding-only hypothesis verification method to reduce hallucinations in knowledge-to-text generation without model retraining.– Created FATE (Fact-Aligned Textual Entailment), a dataset for training task-specific verifiers.– Improved faithfulness metrics (FactKB) by 7.17 points in out-of-distribution settings with negligible latency cost.– Paper: <i>Think While You Write: Hypothesis Verification Promotes Faithful Knowledge-to-text Generation</i> (NAACL 2024).	
Research Scientist Intern <i>Baidu, Inc. (NLP Department)</i>	Nov 2021 – Aug 2022 Beijing, China
Project: Dense Retrieval at Scale Hosts: Dr. Hongyu Li, Dr. Jing Liu	
<ul style="list-style-type: none">– Constructed DuReader-Retrieval, the largest Chinese passage retrieval benchmark derived from real-world search engine logs.	

- Implemented diversity-enhanced dense retrieval training strategies using Natural Language Inference supervision.
- Paper: *DuReader-Retrieval: A Large-scale Chinese Benchmark...* (EMNLP 2022).

PUBLICATION LIST

Self-Improving World Modelling with Latent Actions (Collaboration with Nvidia Research)

Yifu Qiu, Zheng Zhao, Waylon Li, Yftah Ziser, Anna Korhonen, Shay Cohen, Edoardo Ponti

Under Review, ICML 2026

Unified Vision–Language Modeling via Concept Space Alignment

Yifu Qiu, Paul-Ambroise Duquenne, Holger Schwenk

ICLR 2026

Bootstrapping Action-Grounded Visual Dynamics in Unified Vision–Language Models (Collaboration with Nvidia Research)

Yifu Qiu, Yftah Ziser, Anna Korhonen, Shay Cohen, Edoardo Ponti

Under Review, ACL 2026

Eliciting In-context Retrieval and Reasoning for Long-Context Language Models

Yifu Qiu, Varun Embar, Shay Cohen, Yizhe Zhang, Navdeep Jaitly, Benjamin Han

ACL 2025

Spectral Editing of Activations for Large Language Model Alignment

Yifu Qiu, Zheng Zhao, Anna Korhonen, Edoardo Ponti, Shay Cohen

NeurIPS 2024

Are Large Language Models Temporally Grounded?

Yifu Qiu, Zheng Zhao, Anna Korhonen, Edoardo Ponti, Shay Cohen

NAACL 2024

Think While You Write: Hypothesis Verification Promotes Faithful Knowledge-to-text Generation

Yifu Qiu, Varun Embar, Shay Cohen, Benjamin Han

NAACL 2024

Detecting and Mitigating Hallucinations in Multilingual Summarisation

Yifu Qiu, Anna Korhonen, Edoardo Ponti, Shay Cohen

EMNLP 2023

Abstractive Summarization Guided by Latent Hierarchical Document Structure

Yifu Qiu, Shay Cohen

EMNLP 2022

DuReader-Retrieval: A Large-scale Chinese Benchmark for Dense Information Retrieval from Real-World Applications

Yifu Qiu, Hongyu Li, Yingqi Qu, Ying Chen, Qiaoqiao She, Jing Liu, Hua Wu, Haifeng Wang

EMNLP 2022

COLLABORATION PUBLICATION LIST

Lost in Space? Vision-Language Models Struggle with Relative Camera Pose Estimation (Collaboration with Nvidia Research)

Ken Deng, Yifu Qiu, Yoni Kasten, Shay Cohen, Yftah Ziser

Under Review, ACL 2026

Iterative Multilingual Spectral Attribute Erasure

Shun Shao, Zheng Zhao, Yifu Qiu, Yftah Ziser, Shay Cohen

EMNLP 2025

What Is That Talk About? A Video-to-Text Summarization Dataset for Scientific Presentations

Dongqi Liu, Chenxi Whitehouse, Xi Yu, Louis Mahon, Rohit Saxena, Zheng Zhao, Yifu Qiu, Mirella Lapata, Vera Demberg

ACL 2025

Multiple Instance Learning with Coarse-to-Fine Self-Distillation

Shuyang Wu, Yifu Qiu, Ines Nearchou, Sandrine Prost, Jonathan Fallowfield, Hakan Bilen, Timothy Kendall

Under Review by TMI

EEE-QA: Exploring Effective and Efficient Question-Answer Representations

Zhanghao Hu, Yijun Yang, Junjie Xu, Yifu Qiu, Pinzhen Chen

LREC-COLING 2024

TECHNICAL SKILLS

Core Research Area: LLM/VLM Post-Training (SFT, RLHF), Multimodality, Reasoning, AI Safety (truthfulness and fairness)

Deep Learning: PyTorch, HuggingFace Transformers, Diffusers

Scaling and Infrastructure: DeepSpeed, FSDP, Slurm, Distributed Data Parallel (DDP)

Languages: Python, C++, SQL, Bash

AWARDS & HONORS

Apple AIML PhD Fellowship (2023) – *Awarded to top PhD students in AI worldwide.*

ELLIS PhD Program Member (2022) – *Selected for excellence in European AI research.*

Informatics Global PhD Scholarship, University of Edinburgh (2022).

Lord Frederick Cavendish Award, University of Cambridge (2022).

Top 1% Scholarship, BNU-HKBU UIC (2019, 2020).