

Qinglin Zhu

Links: [HomePage](#); [Google Scholar](#)

Mobile: +44-7552-752225

Email: zhuqlfx@gmail.com

PROFILE

PhD candidate in NLP at King's College London. My research focuses on improving the generation, reasoning, and code intelligence of large language models, with recent work on *repository-level code editing*, *latent-space reasoning*, and *retrieval-augmented generation*. I have hands-on experience in large-scale data construction and LLM training on hundreds of GPUs, with industry experience at MSRA, Baidu, Shopee, and Tencent Music.

EDUCATION

- **King's College London** London, UK
Ph.D. in Computer Science (NLP). Supervisors: Prof. Lin Gui, Prof. Yulan He. Oct. 2023 - Present
- **Harbin Institute of Technology** Shenzhen, China
M.Sc. - Computer Science and Technology; GPA: 3.667/4.0; Rank: 1/69 Sept. 2020 - Mar. 2023
- **Harbin Institute of Technology** Shenzhen, China
B.Eng. - Computer Science and Technology; GPA: 3.773/4.0; Rank: 14/70 Sept. 2016 - Jun. 2020

SELECTED PUBLICATIONS

* denotes equal contribution

LLM Reasoning, Decoding & Code Intelligence

- **Qinglin Zhu**, Tianyu Chen, Shuai Lu, Lei Ji, Runcong Zhao, et al. *Pull Requests as a Training Signal for Repo-Level Code Editing.* **ICML'2026.**
- **Qinglin Zhu**, Runcong Zhao, Hanqi Yan, et al. *Soft Reasoning: Navigating Solution Spaces in Large Language Models through Controlled Embedding Exploration.* **ICML'2025 (Spotlight).**
- Hanqi Yan*, **Qinglin Zhu***, Xinyu Wang, et al. *Mirror: A Multiple-perspective Self-Reflection Method for Knowledge-rich Reasoning.* **ACL'2024.**
- **Qinglin Zhu**, Yizhen Yao, Runcong Zhao, et al. *Latent Refinement Decoding: Enhancing Diffusion-Based Language Models by Refining Belief States.* Submitted to **NeurIPS'2026.**

Narrative & Retrieval

- Runcong Zhao*, **Qinglin Zhu***, Hainiu Xu, et al. *SymbolicThought: Integrating Language Models and Symbolic Reasoning for Consistent and Interpretable Human Relationship Understanding.* **ACL'2026 Demo.**
- Runcong Zhao*, **Qinglin Zhu***, Hainiu Xu, et al. *Large Language Models Fall Short: Understanding Complex Relationships in Detective Narratives.* **ACL'2024 Findings.**
- Zhanghao Hu, **Qinglin Zhu**, Siya Qi, et al. *Spectrum Projection Score: Aligning Retrieved Summaries with Reader Models in RAG.* **AAAI'2026 (Oral).**
- Zhanghao Hu*, **Qinglin Zhu***, Runcong Zhao, et al. *Beyond RAG for Agent Memory: Retrieval by Decoupling and Aggregation.* Submitted to **NeurIPS'2026.**

Previous work on *Stance Detection & Sentiment Analysis*:

- **[SIGIR'2022] Qinglin Zhu**, et al.; **[ACL'2022] Bin Liang***, **Qinglin Zhu***, et al.; **[SemEval'2021] Qinglin Zhu**, et al.; **[CCL'2020] Qinglin Zhu**, et al.

EXPERIENCE

- **Microsoft Research Asia (MSRA)** Beijing, China
Research Intern May. 2025 - Present
 - **Tokenizer for Code LLM Pre-training** (Mar. 2026 - Present):
 - * Proposed **SyntaxBPE**, a syntax-aware tokenizer for code LLMs that couples deployable boundary-rule selection with syntax-aware BPE vocabulary induction.
 - * Designed a randomized calibration tokenizer and language model to select syntax/lexical boundary relaxations via paired respect-vs-cross scoring under byte-normalized predictive code length.
 - * Validated SyntaxBPE through **1B/3B** base pre-training on hundreds of GPUs for up to **800B tokens** and SFT, achieving consistent code-generation gains over GPT-style BPE baselines.
 - **Repo-Level Code Editing** (May 2025 - Feb. 2026; **ICML'26**):
 - * Proposed **Clean-PR**, a scalable mid-training paradigm that turns noisy GitHub pull requests into verified Search/Replace edit blocks through filtering, reconstruction, and round-trip validation; built the largest publicly available corpus of **2M PRs** across 12 programming languages.
 - * Conducted LLM mid-training on 100+ GPUs and Agentless-aligned SFT with error-driven data augmentation, achieving **+13.6%** on SWE-bench Lite and **+12.3%** on SWE-bench Verified over instruction-tuned baselines.
 - * Demonstrated that repository-level editing capability can be internalized into model weights under a simplified agentless protocol, with a 32B model outperforming 72B baselines and transferring to Multi-SWE-bench Flash (7 languages) and OpenHands.

- **Baidu** Shanghai, China
Advertising Algorithm Engineer (Full-time) *Apr. 2023 – Oct. 2023*
 - Developed and optimized the Phoenix Nest advertising system’s central dispatch unit (“Mixer”) for online ad allocation, focusing on low-latency ranking/dispatch logic, experiment iteration, and user-preference-aware e-commerce title refinement.
- **Shopee** Shanghai, China
NLP Algorithm Engineer (Intern) *Jul. 2022 – Sept. 2022*
 - Built sentiment knowledge-enhanced pre-training and active learning pipelines for e-commerce review understanding, improving aspect sentiment triplet extraction by +8.6% F1 and reducing annotation pressure through targeted sample selection.
- **Tencent Music Entertainment Group** Shenzhen, China
NLP Algorithm Engineer (Intern) *Aug. 2021 – Apr. 2022*
 - Led aspect-based sentiment analysis for online music content, designing hierarchical classification and active learning workflows to improve training data quality and model iteration efficiency.

HONORS AND AWARDS

- Outstanding Master’s Graduate & Outstanding Dissertation for Master’s Degree, *HIT* (**3%**) Mar. 2023
- Sailvan Times Scholarship (**2%**), *Harbin Institute of Technology* Mar. 2023
- **National Scholarship** (for Graduate Student, **2%**) Oct. 2021
- NLPCC-2021 Shared Task: Argument Pair Extraction (**Top 1**) Mar. 2021
- SemEval-2021 Task 5: Toxic Spans Detection (**Top 1, Team Leader**) Feb. 2021
- Outstanding Bachelor’s Graduate & Outstanding Dissertation for Bachelor’s Degree, *HIT* (**3%**) Jun. 2020
- **National Scholarship** (for Undergraduate Student, **2%**) Oct. 2019

SKILLS

- LLM Training & Alignment: Pre-training, SFT, RLHF/GRPO; PEFT with LoRA
- Distributed LLM Training: DeepSpeed, FSDP, Megatron
- Inference & Serving: vLLM, Sglang
- Programming: Python, C/C++, PyTorch

ACADEMIC SERVICE

Reviewer for: ICML, NeurIPS, ICLR, ACL, AACL, KDD, EMNLP, NAACL, NLPCC

TEACHING EXPERIENCE

- Teaching Assistant for Data Structure King’s College London, 2025 Spring
- Teaching Assistant for Data Science King’s College London, 2025 Spring
- Teaching Assistant for Natural Language Processing Harbin Institute of Technology, 2020 Fall