

Mingyue Liao

Senior Machine Learning Engineer · LLM Post-training (SFT / DPO / GRPO)

Japan · lmyueruc@gmail.com · github.com/liaomingyue · mingyue-liao.pages.dev

5+ years of machine learning engineering at Baidu Search and Alibaba's LLM Application Division; now an LLM technical expert in Japan. Focused on LLM post-training: I train GRPO-based hallucination-detection judges and use their verdicts to improve generators through DPO and self-reward. End-to-end delivery experience in the medical domain, from continued pretraining to retrieval and generation.

Experience

LLM Technical Expert — SERVICEMEME

2025.10 – Present

Japan

- **News compliance screening assistant (deployed; used by the client to batch-process its full database).** Automatically identifies, from news reports, whether a loan applicant has criminal records or is a sanctioned party. The client's Qwen3 + prompt solution degraded as the prompt grew with each edge case, and the team had no fine-tuning experience. Built an error taxonomy (inconsistent labels, uncovered data, mis-linked relations, missing relations, entity-boundary errors) and measured its distribution, then improved in three steps: unified labeling guidelines (+1.2%), synthetic data for coverage (+1.5%), and DPO on preference pairs cross-constructed over the full dataset to fix error-prone cases (+3.5%). Established an iteration loop: expand eval → diversify data → build pairs → SFT → DPO → evaluate.
- **Text-to-Python tool for a US healthcare organization (Azure).** Converts natural language into data-analysis Python. Designed a four-stage workflow: a domain-tuned model drafts code → execute and self-correct (up to 3 rounds; success rate 82% → 93%) → LLM-as-a-judge review → select the final code from trial logs and reviews. DPO on correct/incorrect code pairs plus more multi-table-join and rare-field data raised overall accuracy by 10%; final code accuracy 90%+, analysis hours reduced by 20%+.
- **AI authentication for luxury handbags (in production).** Takes multiple photos of a bag and outputs an authenticity verdict and 8 structured attributes. YOLO11s detects 32 authentication points (mAP50 0.91); authenticity uses a three-stage design of ConvNeXtV2 / DINOv2 local scoring + XGBoost whole-bag verdict, end-to-end AUC 0.898 (closed set); official colorways are matched by ArcFace metric-learning retrieval, so new colorways need no retraining. Deployed as AWS microservices for first-pass screening alongside human review.

Algorithm Expert — Alibaba Group, LLM Application Division

2023.10 – 2025.09

Hangzhou, China

Product: RAG-based medical literature review generation (two-way recall → rerank → outline → body). Owned the generation stage.

- **GRPO-based hallucination-detection judge (solo owner).** Manually attributed errors into three classes: missing premise, unsound reasoning, misattribution; split long answers into referenced claims judged one by one. All negatives synthetic: Qwen 72B rewrote correct text into a target error type, filtered by the detector over three rounds to obtain hard cases. Trained a Qwen 14B judge with GRPO (correctness + format rewards). On 1,000+ human-labeled claims, F1 improved 7% relative to the open-source RefChecker and 10%+ relative to an SFT baseline.
- **RAG generation quality.** Measured by human-rated expertise score (all relative): self-reward alignment +10% with claim-level hallucination rate down ~50%; DPO on judge-built preference data +5%; bilingual terminology dictionary +5%; RAFT noise-robust training +4%.
- **Medical continued pretraining.** Mined medical terminology and synthesized QA and review-outline data with search augmentation to continue-pretrain an open-source base model; powers three production capabilities: Chinese summaries of English papers, document translation and document QA.

Senior Algorithm Engineer — Beijing Yiyong Technology

2022.07 – 2023.09

Beijing, China

- **LLM virtual annotator (offline validation).** Clinical QA rules can't be enumerated and carry no semantics; validated replacing a multi-stage pipeline (NER → relation extraction → entity linking → rules) with end-to-end LLM judgment. Based on LLaMA 13B, using QLoRA / LOMO for low-memory fine-tuning and record segmentation + RoPE extension to overcome the 2,048-token context limit. Offline accuracy: treatment response 67% → 85%, surgical margin 53% → 87%, radiotherapy 76% → 91%.
- **Medical continued pretraining.** Entity-level masking + SBO objective (after SpanBERT); a 2-layer ALBERT-tiny + span model replaced the production 4-layer ALBERT+LSTM+CRF: 3× faster inference, 30% smaller, F1 +1%.
- **Medical terminology standardization.** Led standardization across 5 domains (procedures, diagnoses, lab tests, exams, drugs) serving 1,000+ hospitals; procedure / diagnosis F1 ≥ 90%, lab / exam accuracy ≥ 95%.

Algorithm Engineer — Baidu, Search

2020.07 – 2021.12

Beijing, China

- **General NER for search queries.** Multi-label NER with 70+ entity types in a 3-level hierarchy (ERNIE pretraining + multi-task learning; high-frequency queries cached offline, long-tail queries served by an online 3-layer ERNIE+CRF); accuracy and recall +25% over the Baidu UNIT baseline on novels / film / songs.
- **Parser migration and Du Xiaoxiao.** Migrated 200+ parsing models to the main search architecture; distilled a 3-layer ERNIE sentiment classifier for the virtual assistant Du Xiaoxiao, 90%+ online accuracy, covering 40% of recommendation traffic.

Algorithm Engineer (Intern) — Ke.com (Beike), Search & Language Intelligence

2019.10 – 2020.07

Beijing, China

- Shipped controllable text generation for dialogue (SeqGAN + GPT) and intent classification models; conversation service-quality scoring model reached 95.21% online accuracy.

Education

M.Eng., Software Engineering Renmin University of China

2017.09 – 2020.06

B.Eng., Internet of Things Engineering Hefei University of Technology

2013.09 – 2017.07

Skills

LLM post-training: SFT, DPO, GRPO, PPO, self-reward, RAFT, continued pretraining, knowledge distillation, MergeKit

RAG & retrieval: hallucination detection (judge models), reranking, embedding fine-tuning, chunking strategy

Frameworks & tools: PyTorch, TensorFlow, PaddlePaddle, Transformers, vLLM, Python, C++, Java

Cloud & infrastructure: GCP (Cloud Run, Vertex AI, IAM, Terraform), Azure AI Foundry, Salesforce API

Models: Qwen family (incl. local RTX 5090 deployment), LLaMA, ERNIE, BERT / ALBERT family, GPT-4 (as teacher / judge)

Awards

- National Post-Graduate Mathematical Contest in Modeling, **Third Prize** (2018)
- Sensors Cup Keyword Prediction Competition, **preliminary round 5 / 641** (2018)