Yan Zhan

2001-01 | Male | The 2027 session
2401210760@stu.pku.edu.cn | https://miku2g.github.io

EDUCATION

Beijing University 985 211 C9 Sep 2024 - Jun 2027

Software Engineering Master Full-time Beijing

- Major Course: ML Natural Language Processing Large Language Model Development and Application Artificial Intelligence Practice Algorithm Analysis and Design
- GPA: **3.71/4.0**

Beijing Forestry University 211 Sep 2020 - Jul 2024

Professional in digital media Bachelor Beijing

- Grade Ranking: 2020-2021 Academic Year **2/37** | 2021-2022 Academic Year **3/37** | 2022-2023 Academic Year **1/37** | Total Ranking **2/37**
- School Honors: Outstanding Student Scholarship, Sports, Outstanding Scholarship, Academic Excellence Scholarship, Three Good Students, Outstanding Student Organization Officer

PERSONAL PROFILE

- Love artificial intelligence, master basic ML algorithms and large language model related theories and algorithms
- Proficient in using Python, PyTorch and other programming languages and deep learning frameworks for data analysis processing and model training
- Knowledge of LLM related technologies such as RLHF, PPO, GRPO, DPO, VERL, Scaling Law, RAG, Agent, vLLM, Chain-of-Thought, MoE, Agent, etc
- Familiar with natural language processing models, large language models and multi-modal models, such as BERT, GPT, LLaMA, Qwen, DeepSeek, etc
- Familiar with CNN, ResNet, RNN, LSTM, GRU, Seq2seq,Transformer and other classic deep learning algorithms
- Learning ability, good communication skills, team spirit and high execution

PROFESSIONAL EXPERIENCE

Shenzhen Tencent Computer System Co., Ltd Jul 2026 - Present

WorkBuddy Code Agent algorithm research intern Youtu Lab Beijing

- Project Background: **WorkBuddy**'s in-house model, designed to compete with **Cursor Composer,** lagged well behind frontier models on web-agent tasks. The core issue was that the **legacy evaluation framework** (the current open-source version of WorkBuddyBench) could no longer reliably assess model performance or data quality. Its main limitations were: **inaccurate evaluation**, as the evaluator relied on scripted pipelines and stitched static screenshots without simulating human interaction with webpages; **poor task taxonomy**, as benchmark tasks lacked fine-grained categorization, resulting in limited diversity and discriminative power and making it difficult to identify specific weaknesses; and **weak data mining**, with large volumes of production feedback data left underutilized and no closed-loop **data → evaluation → training** flywheel.
- Solution: Propose and lead **evaluation-driven optimization** technology mainline, refactoring four-tier infrastructure: **Agent All Judge**, reconstructing the past miscellaneous referees into a task-level Agent review of one topic and one session. The Agent can operate the page, collect source code/screenshots/interactive evidence on demand, and rule the whole topic in one session; **mission x Ability** matrix, building 131000 Web task three-level classification trees from 4 million lines of data and a five-axis system (ability/knowledge/change surface/operation/difficulty) from the real execution trajectory, put **total score by task** upgrade **competency Matrix** diagnosis; **automatic problem-making engine**, build a three-stage problem-making pipeline (plan & rarr; build environment & rarr; build rubric), sample according to the "scene × ability" cell, daily production is about **500** problem; **data Flywheel**, rubric is used to mark the weak items of the backflow trajectory mining model, and SFT data is fine-screened according to topic × ability ratio from about 100000 backflow data. After training, the backflow is verified by the same evaluation system.
- Project results: the evaluation system and the expert label live consistent rate. **80.8%→90.7%**(Manslaughter Rate Drops 75%); Coverage rate of problem-making ability **55%→70%**40% of nearly repeated questions & rarr;20%; The passing rate of rubric in the post-SFT domain **+ 18.3pp**(53.9%→72.1%) **weak Tasks** equally divided **+ 78%**(0.36 & rarr;0.64); the system has been **landing** it is a standard evaluation line for multiple versions of the self-developed model, directly supporting access evaluation and data production decisions.

Shenzhen Tencent Computer System Co., Ltd Oct 2025 - Jun 2026

LLM Algorithm Research Intern (Research + Product) PCG Social Platform Architecture Center QQ Algorithm Group Shenzhen

- Project Background: QQ Overseas Chat System Needs to Realize Similar to WeChat" **write while translating**"The fast translation function is required in **eurasian languages** accurate translation, **reasoning speed** fast and able to adapt to shorter **colloquial** sentence translation scene. Existing mixed meta-translation models are in **asian Languages** there are short boards that require independent production of datasets and training of small translation models.
- Solution: with high quality **film and television dialogue** with **voice dialogue** data-based, supplemented by literary data, the construction of 13 languages across Europe and Asia translation data. **Article 1 million**, involving data screening and cleaning, missing language distillation and completion, NER entity repair and other standardization. At the same time, the translation ability and reasoning delay of each 4B level small model are evaluated, and finally the training base is selected as the gemma-4-E4B for SFT.
- Project Outcome: Within the Model Domain **performance improvement** about **10 percentage points** extraterritorial FLORES-200 with WMT25(GEMBA/XCOMET indicator) **universal Translation Assessment** both with Hy-MT2-30B-A3B **flat**.

Perfect World (Beijing) Software Technology Development Co., Ltd Jun 2023 - Dec 2023

NLP algorithm intern Advertising language recognition based on LLM Beijing

- Project Background: In order to prevent competitive advertisements from appearing in the player WeChat communication group, resulting in customer loss, it is necessary to **identify competitive, malicious and game-related advertisements** in time.
- Solution: **chi-square pre-annotation** and **manual verification** are used to obtain the annotation data **knowledge** base, **LoRA** technology is used to train Qwen1.5-14B, **RAG** is combined to improve recall effect, and **vLLM** deployment is used to improve efficiency.
- Project Results: F1-score **increased to 85%** and successfully **landed** on the hour-level offline **batch advertisement recognition task**.

RESEARCH

Length-Adaptive Decoding for Masked Diffusion Machine Translation. Jan 2026 - May 2026

One work. EMNLP 2026 Master Conference. https://arxiv.org/pdf/2608.22274

- Project background: The mask diffusion language model is required in machine translation. **Advance** determine **target Length**, length **too short** would lead to information **omission**, **too long** easily generated **repeat**, hallucinations or redundancy. The existing decoding research mainly focuses on the token decoding order, and **target Length Selection** this key bottleneck is insufficiently explored. This study aims **no training** under the condition of increasing the length of diffusion machine translation. **Adaptive ability** with **integrity of translation**.
- Experimental scheme: proposed **no need for training** of **entropy-Valley(EV)** length **adaptive decoding** method, **candidate Length Construct** with **full mask prediction entropy** scoring, for diffuse machine translation **dynamic selection** the target length. in LLaDA-8B. **English-Chinese, Chinese-English, English-German translation** on the task, compared with the decoding results of fixed-scale length decoding and the reference translation length as the theoretical upper limit, it is verified that the method can be used in **does not rely on the length of the reference translation** under the circumstances **approaching Ideal Length Selection Effect** and combined with automatic indicators, manual evaluation and cross-model experiments to analyze its stability.
- Research significance: proved the purpose. **Standard Length Selection** yes **fixed Canvas** mask **diffusion Machine Translation** in the core **performance bottleneck**, and **no need** add a training target, length predictor, or external model, and you can **significant improvement** translation adequacy and coverage. This method is a diffusion language model in machine translation. **Length-sensitive generation tasks** the decoding optimization in provides **lightweight, migratable** the solution.

SLCA-GRPO:Resolving Cross-Segment Credit Misattribution in Tool-Calling RL Sep 2025 - May 2026

One work NeurIPS 2026 in the cast

- Project Background: The work originated from the need for efficient tool use in **QQ's built-in agent**. I focused on the credit assignment problem in **long-horizon agent tasks**. A **Tool-Calling** agent produces both structured **tool-call** segments and **natural-language** summary segments. Standard GRPO broadcasts a trajectory-level advantage uniformly to all tokens, allowing summary quality to contaminate the advantage assigned to tool-call tokens **across segments**, which can lead to **incorrect** credit **assignment** and training **instability**.
- Solution: Propose **SLCA-GRPO** framework, rewards for tool segments and summary segments **independent normalization** and **route separately**, route the execution reward to the tool call token, route the preference reward to the summary token, and implement scalable API interaction training and construction using Golden ToolCalling based on the Schema constrained LLM simulator. **Layered Dense Rewards**.
- Project results: improved intra-domain evaluation on 7B model compared to standard GRPO, ToolPO and RLTR **2.53pp**, BFCL boost **1.36pp**, & tau;²-Bench Lift **9.15pp**, effective **suppress redundancy** tool invocation; paper submission attempts to use training methods **Agent Class** data and migrate the segmented scoring ideas **business Eval System**, build tool call and summarize quality **independent scoring system**.

TrustRoboReward: Preference-Ordered Isotonic Score Editing for Multi-Paradigm Robot Reward Models. Jul 2025 - May 2026

A total of one. NeurIPS 2026 in the cast. https://arxiv.org/pdf/2608.08491

- Project background: **RoboReward** mainly 1-5 points **track Progress** scoring, missing **pairwise** preference supervision and video scene understanding dimensions; the direct introduction of multi-paradigm supervision in turn produces **pointwise score** with **pairwise preference** between **conflict**, Impact Reward Model **training stability**. This study builds a more general and stable task-oriented. **multi-paradigm** multimodal reward models.
- Experimental scheme: building Score-A, Score-B, Pair-A and Pair-B based on RoboReward video samples **four types of supervision data** covered separately. **Track Progress** score, **Video-QA answer quality** score, **track Video Preferences** comparison and **VQA Answer Preference** comparison. For **reverse score-pair** conflict, introducing **POISE consistency correction square** method, the pairwise label is modeled **partial order constraint**, and through **PAVA algorithm** correction **pointwise** fractions; for small pairwise rings and **transitive conflict** proceed **repair** and the cleaned data is used for SFT and GRPO training.
- Significance of the study: Extending the evaluation dimension of the reward model through multi-paradigm supervision and correcting it through explicit consistency. **Decrease** in the training corpus **conflict** signal to improve the stability and credibility of the model evaluation. In the experiment, the training corpus score-pair conflict can be reduced **0%** the comprehensive index of the same parameter model is reached. **78%**, higher than RoboReward-4B **10%**, when testing **score-pair agreement rate** lift **72%**.

HONORS & AWARDS

The 8th China International Internet+ College Students Innovation and Entrepreneurship Competition Beijing Division rematch High First Prize 2022.08

The third prize was 2. in the national finals of the national college students' digital media science and technology works and creativity competition 2021.11

Second prize in the annual contest of 2021 national modeling contest 2021.12