He Du (杜贺)

elyndendu@gmail.com | github.com/kinza99 | Google Scholar | +86 18324568312

Education

Fudan University Shanghai, China
Ph.D. Candidate in Artificial Intelligence Sept. 2024 – Present
School of Computer Science and Technology

- **Research Interests:** Large Language Models (LLM), Reinforcement Learning (RL), Code Intelligence.

Harbin Institute of Technology (HIT) Harbin, China
Bachelor of Engineering in Computer Science and Technology Sept. 2020 – June 2024

Internship Experience

Shanghai AI Laboratory (InternLM Team) Sept. 2024 – Apr. 2026

Qwen, Alibaba Apr. 2026 – Present

Publications & Manuscripts

Confidence as a Reward: Transforming LLMs into Reward Models *ICASSP 2026*

He Du, Bowen Li, Chengxing Xie, Chang Gao, Kai Chen, Dacheng Tao

- Systematically investigated Confidence-as-a-Reward (CRew).
- Proposed CRew-DPO, a training strategy that constructs preference data from confidence scores combined with correctness signals

EvoSyn: Generalizable Evolutionary Data Synthesis for Verifiable Learning *ACL 2026 main*

He Du, Bowen Li, Aijun Yang, Siyang He, Qipeng Guo, Dacheng Tao

- Introduced an evolutionary, task-agnostic, strategy-guided, executably-checkable data synthesis framework.
- Demonstrated the effectiveness of the proposed approach under both RLVR and model distillation training paradigms.

Swe-fixer: Training Open-source LLMs for Effective and Efficient GitHub Issue Resolution *ACL 2025 findings*

Chengxing Xie, Bowen Li, Chang Gao, **He Du**, Wai Lam, Difan Zou, Kai Chen

DevBench: A Comprehensive Benchmark for Software Development *COLING 2025*

Bowen Li, Wenhan Wu, Ziwei Tang, Lin Shi, John Yang, Jinyang Li, Shunyu Yao, Chen Qian, Binyuan Hui, Qicheng Zhang, Zhiyin Yu, **He Du**, Ping Yang, Dahua Lin, Chao Peng, Kai Chen

Large Language Models Meet Symbolic Provers for Logical Reasoning Evaluation *ICLR 2025*

Chengxing Qi, Ren Ma, Bowen Li, **He Du**, Binyuan Hui, Jinwang Wu, Yuanjun Laili, Conghui He

Open Source & Projects Experience

Qwen3.8-Max Foundation Model Delivery Qwen Internship
Hybrid RL

- Core contributor of the coding hybrid RL stage for the Qwen3.8-Max foundation model delivery.
- Designed and conducted mixed RL training recipe over SWE-bench and Terminal-bench task families.
- Introduced filtering mechanisms for non-sequential multi-turn rollout tasks to ensure training stability.

- Supports training with various agent scaffolds, including swe-agent, terminus-2, and claudecode.
- Optimized the RL pipeline to defend against reward hacking in web-retrieval question-answering tasks.
- Monitor the Qwen3.8-max training process and resulting teacher was directly incorporated into the final delivered stage.

Kernel LLM InternLM Internship

- Target on optimizing high-complexity and common GPU kernels (e.g., MLP) to accelerate inference process.
- Curated a large-scale Torch module dataset representing real-world scenarios by scraping, standardizing, and filtering extensive GitHub repositories.
- Responsible for the complete RL lifecycle: environment setup, integration, training, and evaluation.
- Developed EvoRL by coupling RL training with the OpenEvolve pipeline, achieving substantial performance gains in automated kernel optimization.
- Implemented on-policy distillation algorithms using VeRL to facilitate the research of efficient distillation paradigms.
- In charge of the reinforcement learning component for the GPU kernel generation task, attaining SOTA results on KernelBench(triton) by building upon Qwen3-235B.

AI scientist InternLM Internship

- Focus on developing an AI Scientist framework to autonomously manage the end-to-end specific model training lifecycle.
- Developed a Code Agent system equipped with a comprehensive suite of tools for data processing, task submission, and execution monitoring, granting the model autonomous tool-use capabilities. Leveraging this framework, we synthesized high-quality datasets to perform SFT, further enhancing model performance.
- By leveraging our Code Agent, the resulting domain-specific models achieve performance on par with those meticulously tuned by human experts.

VeRL (multi-turn RL training support) **PR Link**

- **Key Contribution:** Designed and implemented the **Multi-turn RL training support**, enabling the model to handle complex, multi-turn reasoning tasks and interactive environments.

Long-Context LLM Post-training InternLM Internship

- Constructed a taxonomy-driven data synthesis pipeline to generate long-context training data across multiple domains.
- Delivered **internlm2-7b-chat-1m**, among the first open-source long-context LLMs supporting 1M tokens, outperforming open-source counterparts on long-context tasks.

Technical Skills

- **Languages:** Python, C++, SQL, Bash, LaTeX.
- **Frameworks & Tools:** PyTorch, HuggingFace, Transformers, Ray, DeepSpeed, FSDP, etc.
- **Research Areas:** RLHF (PPO, DPO, GRPO), LLM Fine-tuning, Code Generation, Automated Reasoning.

Honors & Awards

- Academic Scholarship, Fudan University 2024
- Outstanding Graduate, Harbin Institute of Technology 2024