Guanxu Chen (陈冠旭)

18360196678 | WeChat quantumcgx lm.cgx@sjtu.edu.cn | quantumcgx@gmail.com https://biuboomc.github.io

Education

PhD. Student: Shanghai Jiao Tong University **Major:** Information and Communication Engineering 2025 – Now

B.S. Degree: Shanghai Jiao Tong University **English Level:** CET6 2021 – 2025

Dual Bachelor of Degrees **Major 1:** Information Security **Major 2:** Mathematics and Applied Mathematics

Research Interests

Trustworthy AI, Reinforcement Learning, Looped Transformers, Interpretability

Research Topic & Publications

Large Reasoning Model & Reinforcement Learning

- ICLR 2026 (Accept) Conditional Advantage Estimation for Reinforcement Learning in Large Reasoning Models
  Guanxu Chen, Yafu Li, Yuxian Jiang, Chen Qian, Qihan Ren, Jingyi Yang, Yu Cheng, Dongrui Liu, Jing Shao
- Arxiv Rethinking Entropy Regularization in Large Reasoning Models
  Yuxian Jiang, Yafu Li, **Guanxu Chen**, Dongrui Liu, Yu Cheng, Jing Shao

Trustworthy AI

- ICML 2026 (Accept) Beyond External Monitors: Enhancing Transparency of Large Language Models for Easier Monitoring
  Guanxu Chen, Dongrui Liu, Tao Luo, Lijie Hu, Jing Shao
- Blog/Report SafeWork-R1: Coevolving Safety and Intelligence under the AI-45° Law
  One of the Core Contributors.
- Blog/Report Frontier AI Risk Management Framework in Practice: A Risk Analysis Technical Report v1. 5
  One of the Core Contributors.

Looped Transformers & Latent Reasoning

- Blog/Report Loop as a Bridge: Can Looped Transformers Truly Link Representation Space and Natural Language Outputs?
  Guanxu Chen, Dongrui Liu, Jing Shao

Masked Diffusion Large Language Model & Reinforcement Learning

- Arxiv Taming Masked Diffusion Language Models via Consistency Trajectory Reinforcement Learning with Fewer Decoding Step
  Jingyi Yang, **Guanxu Chen**, Xuhao Hu, Jing Shao

Project & Internship Experience

Shanghai AI Lab Internship (Lead by Jing Shao) 2024.6 – now

Professional Skills

- Python/C++ (Python≫C++), PyTorch
- Familiar with transformers, generative models and reinforcement learning
- Driven by curiosity and a desire for innovation