

Yunxiang Wei

+86-151-0203-0398 | yunxwei@zju.edu.cn
Hangzhou, Zhejiang, China

OBJECTIVE

My primary research interest lies in **code LLMs and agents**. I am particularly interested in deploying LLMs in real-world tasks such as coding and scientific research. In the context of coding, my work focuses on ensuring the **reliability** of generated code, including detecting data risks and estimating/verifying correctness. In scientific discovery, I am interested in equipping AI systems with **interdisciplinary knowledge** to support idea generation, evaluation, and verification. My long-term research goal is to build an automated systems with continual knowledge learning capability as well as reliability.

EXPERIENCE

- **Research Intern (Prof. Yang Liu)** Mar 2025 - Present
Code LLMs Nanyang Technological University, Remote
 - **Pretraining Data Detection (PDD) for code LLMs**: developed **CodeSnitch**, a multilingual function-level benchmark (**9K samples, 3 programming languages**) for detecting training data leakage in code LLMs; designed mutation strategies based on **Type-1–Type-4 code clone** taxonomy and performed large-scale evaluation of **7 PDD methods across 8 code LLMs**; proposed a **Reverse Code Mutation (RCM)** module, which substantially enhances the robustness of existing PDD approaches under code mutations.
 - **Uncertainty estimation in code generation**: proposed **Ensemble Semantic Entropy (ESE)**, an uncertainty metric aggregating semantic consistency across multi-LLM ensembles to detect false positive samples, which **improves prediction accuracy at 10% false positive rate by 53.4%** in selective generative tasks; designed **CAS, a cascading test-time scaling framework** utilizing ESE as a decision signal to adaptively allocate inference budgets, achieving **64.9% FLOPs reduction** compared to single-model scaling while maintaining performance.
- **Research Intern (ZJUKG, Prof. Ningyu Zhang)** Oct 2025 - Present
Knowledge Graph, LLM Agents Zhejiang University
 - Developed **InnoEval**, a knowledge-grounded **multi-agent evaluation framework** for scientific ideas; implemented a heterogeneous **deep knowledge retrieval engine** combining fast search and slow reading to achieve a balance between effectiveness and efficiency, with a multi-round reranking and query refinement mechanism to ensure both depth and breadth of retrieved sources.
 - Constructed **evaluation datasets** derived from peer-reviewed research submissions and conducted extensive benchmarking on point-wise, pair-wise, and group-wise evaluation tasks, where InnoEval demonstrates **strong alignment with human expert judgments**.

EDUCATION

- **Zhejiang University** August 2023 - June 2027
Undergraduate, Bachelor of Engineering Zhejiang, China
 - Major: Computer Science and Technology
 - Rank: 8/143
 - GPA: 3.97/4.0

PUBLICATIONS

C=CONFERENCE, J=JOURNAL, S=IN SUBMISSION

- [S.1] Tianlin Li, **Yunxiang Wei** et al. (2025). Are AI Coders Snitches? An Empirical Study of Training Data Detection on Code Large Language Models. [[arXiv:2507.17389](https://arxiv.org/abs/2507.17389)] [[Code](#)]
- [S.2] **Yunxiang Wei**, Tianlin li et al. (2026). Predicting Program Correctness With Ensemble Semantic Entropy.
- [S.3] Shuofei Qiao, **Yunxiang Wei** et al. (2026). InnoEval: On Research Idea Evaluation as a Knowledge-Grounded, Multi-Perspective Reasoning Problem. [[arXiv:2602.14367](https://arxiv.org/abs/2602.14367)][[Code](#)][[Demo](#)]

HONORS AND AWARDS

- **Tianzhou Chen Scholarship** September 2025
Zhejiang Province
 - Award for outstanding performance. (3/143)
- **Zhejiang Province Scholarship** September 2024
Zhejiang Province
 - Award for outstanding performance.
- **National Olympiad in Informatics in Provinces (NOIP) 2019 First Prize** November 2019
China Computer Federation (CCF)
 - Award for advanced skills in coding and solving algorithmic problems skills

ADDITIONAL INFORMATION

Languages: Chinese (Native), English (TOEFL: 106; CET-6: 600)