

## YANXIN LU

+86 131-4994-1411 | yanxin4@illinois.edu

718 Haizhou East Road, Haining, Zhejiang, 314400, China

### EDUCATION

---

#### Zhejiang University (ZJU)

Haining, China

B. Eng. in Computer Engineering | GPA: 3.97/4.00 | Rank: 5/65

Sep. 2021 - Expected May 2025

#### University of Illinois Urbana-Champaign (UIUC)

Urbana, IL

B. Sc. in Computer Engineering | GPA: 3.97/4.00

Sep. 2021 - Expected May 2025

- **Core Courses:** Basic Discrete Mathematics, Intro to Electronics, Analog Signal Processing, Data Structures, Database Systems, Computer Systems Engineering, Artificial Intelligence, Game Development, Applied Parallel Programming.
- **Awards & Honors:** Zhejiang Provincial Government Scholarship; Zhejiang University – Second Prize Scholarship; ZJU-UIUC Institute - Dean’s List; ZJU-UIUC – Third Prize Scholarship; The 9th China International College Students’ “Internet+” Innovation and Entrepreneurship Competition - Bronze Award.

### PUBLICATION

---

Minyang, Tian, ..., **Yanxin, Lu**, ..., et al. “SciCode: A Research Coding Benchmark Curated by Scientists.” *NeurIPS 2024 Datasets and Benchmarks Track*, Accepted.

### RESEARCH EXPERIENCE

---

#### Enhancing LLM’s Reasoning Capabilities with Multi-Tool Integration

Urbana, IL

Core Researcher; Supervisor: Prof. Tong Zhang

Jul. 2024 - Present

- Designed and developed multi-turn dialogue datasets for multi-tool reasoning tasks, enabling step-by-step problem solving by calling specific APIs at each reasoning step.
- Conducted an extensive literature review to identify and select appropriate benchmarks and methodologies.
- Applied Monte Carlo Tree Search (MCTS) algorithms to explore solution paths and identify valid tool usage chains for solving mathematical and logical reasoning problems.
- Fine-tuned open-source models using the newly created datasets to enhance LLM’s reasoning performance.

#### SmoothNN: Towards Better Generalization by Learnable Smoothing on Neural Networks

Urbana, IL

Team Member; Supervisor: Prof. Tong Zhang

Nov. 2024 - Present

- Proposed a technique to enforce model smoothness by perturbing input covariates and penalizing prediction variance induced by these perturbations to improve generalization capabilities.
- Experimented with various neural network architectures, including ResNet and Vision Transformer, applying “smoothing” procedure to evaluate its effectiveness.
- Tuned hyperparameters of the “smoothing” procedure to optimize performance and compared them with non-perturbation structures.

#### LA-Eval: Evaluating the Long-context Capabilities of LLMs in AI-assisted Research Scenarios

Urbana, IL

Team Member; Supervisor: Prof. Tong Zhang

Oct. 2024 - Nov. 2024

- Developed LA-Eval, a benchmark for evaluating the long-context reasoning capabilities of LLMs in AI-assisted research scenarios involving multi-turn interactions.
- Conducted in-depth analysis and proposed enhancements, such as incorporating additional statistical metrics and using claude-3.5-sonnet as an external evaluator to assess article quality more robustly compared to GPT-4-based evaluations.
- Contributed to writing the research paper and developing the blog website (<https://yanxinlu.github.io/LA-Eval/>)

## Scicode: A Research Coding Benchmark Curated by Scientists

Urbana, IL

Team Member & Core Programmer; Supervisor: Prof. Hao Peng

Mar. 2024 - May 2024

*Open-source code link:* <https://github.com/scicode-bench/SciCode>, *Paper Accepted by NeurIPS 2024*

- Developed a challenging scientist-curated coding benchmark, SciCode, to evaluate the ability of language models to generate code to solve real-world scientific research problems.
- Designed high-level API interface and led cross-team members to finish coding implementation using Python; Responsible for over 80% of the development and implementation tasks.
- Processed scientific problem descriptions based on domain knowledge and designed test cases to build evaluation datasets.
- Established a robust evaluation pipeline to assess accuracy and efficiency of code generated by LLMs.
- Designed and optimized LLM prompts for various tasks to evaluate the versatility and programming capabilities of models.
- Evaluated proprietary models against established benchmarks, including *GPT*, *Claude*, and *Gemini*.

## Children's Education Software Based on GPT and YOLOv5

Hangzhou, China

Team Leader; Supervisor: Dr. Liuqing Chen

Jun. 2023 - Jul. 2023

- Developed an educational software on Windows that allowed children to identify and learn about objects outside the car by simply tapping the screen without distracting their parents while driving.
- Led a team to determine project strategy, conduct daily management and lead the development of interactive features.
- Integrated YOLOv5 for real-time object detection and recognition from live video stream, and leveraged GPT to enable interactive learning scenarios including concept explanation, knowledge QA and story telling.
- Designed a user-friendly interface to enhance the interactivity and accessibility of the educational software using Qt.

## INTERNSHIP

---

### Shanghai Lilith Technology Corporation

Shanghai, China

Large Language Model (LLM) Algorithm Engineer

Jul. 2024 - Oct. 2024

- Developed an automated quality assurance agent based on screenshots for interactive game UI, capable of adaptively arranging test steps and automatically generating test scripts based on test tasks, objectives, and interface changes.
- Utilized multi-modal large language models (LLMs) to detect anomalies in game scenes, such as text exceeding boundary and garbled text.
- Enhanced machine translation with LLM by integrating translation memory to augment the terminology database.

## COURSE PROJECT

---

### GameUniverse: A Game Searching Website | Game Development

Urbana, IL

Developer

Sept. 2023-Dec. 2023

- Led a four-person team to achieve Steam game searching using MySQL database by designing and implementing core database functionalities.
- Developed a website to show search results by implementing a user system, advanced search filters and game descriptions.

### Design of A Unix-Like Operating System | Computer Systems Engineering

Urbana, IL

Developer

Oct. 2023 - Dec. 2023

- Developed various components like RTC, terminals and system calls in the kernel from bare-bones.
- Added features such as signal handling, text editor and auto-completion using Tab to build a functional OS.

## LANGUAGES & SKILLS

---

- **Programming Languages:** Python, C, C++, SQL, x86 Assembly
- **Tools:** Pytorch, Qt, MySQL, Git, QEMU, Quartus
- **Language:** Native in Chinese, Fluent in English