Tengchao Yang

Email: krisy0409@gmail.com Personal Website: https://tangciuyueng.github.io/

EDUCATION

Tongji University Shanghai, China
Bachelor of Engineering in Software Engineering Sep. 2021 – Jun. 2026 (Expected)
GPA: 4.92/5.0 **Ranking:** 2/197
Relevant Courses: Algorithm Design and Analysis, Distributed Systems, Principles of Compilers, Software Design Patterns, Information Security, Statistics Analysis and Modeling, Large-Scale Database Application Development
Research Interests: Trustworthy AI, AI Safety, Multimodal LLM, AI for Education

PUBLICATIONS

[1] W. Wang, D. Huang, J. Li, **T. Yang**, Z. Zheng, D. Zhang, *et al.*, "CMPhysBench: A Benchmark for Evaluating Large Language Models in Condensed Matter Physics," *under review (submitted to ICLR)*, Sep. 2025. Available: https://doi.org/10.48550/arXiv.2508.18124

[2] **T. Yang**, S. Guo, M. Jia, J. Su, Y. Liu, Z. Zhang, and M. Jiang, "MMTutorBench: The First Multimodal Benchmark for AI Math Tutoring," *under review (submitted to ACL)*, Oct. 2025. Available: https://doi.org/10.48550/arXiv.2510.23477

[3] Q. Wei, **T. Yang** (co-first author), Y. Wang, X. Li, L. Li, Z. Yin, *et al.*, "A-MemGuard: A Proactive Defense Framework for LLM-Based Agent Memory," *under review (submitted to ICLR)*, Sep. 2025. Available: https://doi.org/10.48550/arXiv.2510.02373

RESEARCH EXPERIENCE

MMTutorBench: Multimodal Benchmark for AI Math Tutoring Jun. 2025 - Oct. 2025
Research Assistant, iSURE Program Supervisor: Prof. Meng Jiang

- Co-developed MMTutorBench, the first benchmark for assessing multimodal large language models (MLLMs) on math tutoring tasks involving reasoning, diagnosis, and step-by-step guidance.
- Constructed a dataset of 685 pedagogically grounded problems covering algebra to calculus, encompassing three task types—Insight Discovery, Operation Formulation, and Execution.
- Designed a six-dimensional rubric-based evaluation framework and implemented an LLM-as-a-Judge pipeline for scalable, human-aligned assessment.
- Conducted model evaluation, providing quantitative analysis of reasoning gaps between proprietary and open-source MLLMs and identifying key limitations in multimodal understanding.

A-MemGuard: A Proactive Defense Framework for LLM-Based Agent Memory Mar. 2025 - Sep. 2025
Research Assistant Supervisors: Dr. Xinfeng Li & Prof. XiaoFeng Wang

- Co-developed A-MemGuard, a proactive defense system addressing memory poisoning in LLM-based autonomous agents.
- Designed the consensus-based validation module to detect anomalous reasoning by comparing multi-memory inference paths without modifying the agent architecture.
- Contributed to the dual-memory mechanism, enabling adaptive self-correction by distilling detected anomalies into structured "lesson" memories.
- Conducted large-scale experiments across reasoning and agent benchmarks, demonstrating a 95% reduction in attack success rate with minimal performance degradation.

CMPhysBench: Condensed Matter Physics Benchmark for LLMs May 2025 - Sep. 2025
Research Assistant Supervisor: Dr. Shufei Zhang

- Led dataset construction: extracted and curated over 520 complex, open-ended problems from graduate-level condensed matter physics textbooks to form a novel benchmark dataset.
- Designed the data curation pipeline, performed rigorous data cleaning and standardization on intricate physics problems, ensuring the accuracy of complex mathematical equations and theoretical concepts for evaluating Large Language Models (LLMs).
- Contributed to the development of CMPhysBench, a pioneering benchmark designed to test the advanced reasoning and mathematical capabilities of LLMs in a frontier area of physics.

LLM Genealogy and Provenance Research Jan. 2025 – Present
Research Assistant Supervisor: Prof. Kaifeng Huang

- Benchmarked state-of-the-art provenance methods, identifying bottlenecks in high-cost dynamic and fragile static approaches.
- Proposed a novel static-analysis framework using timestamp features to overcome the limitations of unstable metrics and parameter cosine similarity.
- Developed a machine learning solution on a large-scale dataset to automatically learn the optimal distance between models, engineering a low-cost, data-driven solution for traceability and compliance.

INTERNSHIP

Research Assistant**, Shanghai Artificial Intelligence Laboratory** Supervisor: Prof. Shufei Zhang Sep. 2025 - Present

- Design and implement the SciDR Bench evaluation framework for large language models in scientific reasoning, covering multimodal QA, experimental design, structured data extraction, and code generation tasks.
- Develop a quantitative scoring mechanism that combines rule-based metrics with LLM-as-judge evaluations to assess model outputs across heterogeneous task categories.
- Model Deep Research multi-step tasks, defining atomic operations and multi-hop planning processes tailored to physical and chemical research scenarios.

AWARDS AND HONORS

National Scholarship (Top 0.2%)	Oct. 2024
University-level Undergraduate Scholarship (Top 5%)	Oct. 2023 & Oct. 2024
Merit Student (Top 10%)	Oct. 2022 & Oct. 2024

PROFESSIONAL SKILLS

Language: English (Proficient, CET-6: 548, IELTS: 7), Chinese (Native)
Programming: Python, Java, C/C++, SQL, JavaScript
AI & Machine Learning: PyTorch, TensorFlow, Huggingface, Scikit-learn, LangChain, Pandas, Numpy, Matplotlib
Developing Tools: Git, Linux, Docker, Latex, Weights & Biases, Makefile