Yuxuan Jiang

ML Reliability Through Observability

(+1) 217-954-3744 | jyuxuan@umich.edu | Essoz | essoz.github.io

Research Interests

I build principled **observability layers for ML and distributed systems** to proactively detect, diagnose, and prevent silent failures. My work combines systems techniques (runtime monitoring, invariant inference, fault tolerance) with AI-based reasoning, producing tools like **TrainCheck** (*OSDI'25*) that have caught bugs in PyTorch, HuggingFace, and DeepSpeed, and frameworks like **SilosBridge** and **XPert** (*ICSE'24*) deployed in Microsoft production clusters. My goal is to make large-scale AI systems more reliable, efficient, and trustworthy.

Education

University of Michigan
Ph.D. in Computer Science and Engineering
08/2023 – 04/2028 (expected), Ann Arbor, MI

- Advisor: Prof. Ryan Huang
- Research focus: Reliability of Cloud-Scale Distributed Systems

University of Illinois Urbana-Champaign
B.S. in Computer Engineering
08/2019 – 05/2023, Urbana, IL

Zhejiang University
B.Eng. in Computer Engineering
08/2019 – 05/2023, Haining, Zhejiang, China

Publications

- **Training with Confidence: Catching Silent Errors in Deep Learning Training with Automated Proactive Checks**
  Yuxuan Jiang, Ziming Zhou, Boyu Xu, Runhui Xu, Beijie Liu, Ryan Huang
  OSDI 2025
- **One-Size-Fits-None: Understanding and Enhancing Slow-Fault Tolerance in Modern Distributed Systems**
  Ruiming Lu, Yunchi Lu, **Yuxuan Jiang**, Ryan Huang
  NSDI 2025
- **Xpert: Empowering Incident Management with Query Recommendations via Large Language Models**
  Yuxuan Jiang, Chaoyun Zhang, Shilin He, Zhihao Yang, Minghua Ma, Si Qin, Yu Kang, Yingnong Dang, Saravan Rajmohan, Qingwei Lin, Dongmei Zhang
  ICSE 2024
- **Acto: Automatic End-to-End Testing for Operation Correctness of Cloud System Management**
  Jiawei Tyler Gu, Xudong Sun, Wentao Zhang, **Yuxuan Jiang**, Chen Wang, Mandana Vaziri, Owolabi Legunsen, Tianyin Xu
  SOSP 2023

Research & Industry Experience

Microsoft Research, Systems Innovation Group
Research Intern
05/2025 – 08/2025, Redmond, WA

- Led development of **SilosBridge**, a multi-agent framework for cloud incident triage.
  - Evaluated on **133 high-severity production incidents**; reducing outage response time by an average of **66%**.
  - Deployed in **4+ Microsoft production teams**, credited with reducing on-call toil.
  - Advanced multi-agent orchestration under noisy and incomplete signals, pushing real-world practicality of multi-agent systems.
  - In submission to *FSE'26*.

University of Michigan, Ordered Systems Lab
Graduate Research Assistant
08/2023 – Present, Ann Arbor, MI

- Creator of **TrainCheck** (*OSDI'25*), a proactive runtime checking framework for ML training reliability.
  - Studied **88 real-world silent training failures**; reproduced **20** high-severity ones, TrainCheck detected **18 within a single iteration**.
  - Found **6 new bugs** in PyTorch, HuggingFace Transformers, and DeepSpeed.
  - Python-based instrumentor with **<2% overhead**; invariant inference + precondition synthesis for precise, reusable checks.
  - Awarded the **Michigan ADVANCE Translational Research Grant** and accepted at the **PyTorch Conference 2025**.
  - Open-sourced at github.com/OrderLab/TrainCheck (49 stars, growing community interest).

Microsoft Research Asia
Research Intern
05/2021 – 08/2021, Beijing, China

- Developed **XPert** (*ICSE'24*), an LLM-based framework for incident diagnosis in Azure.
  - Analyzed **346,508 incident tickets** and **712,222 KQL queries**; found **50% of incidents** managed with only one query → strong case for automation.
  - Designed query synthesis/validation pipeline; introduced **Xcore**, a semantics-based metric, beating CodeBLEU in accuracy.
  - Achieved **state-of-the-art query recommendation**; deployed internally to support on-call engineers in production.

University of Illinois Urbana-Champaign, XLAB
Undergraduate Research Assistant
09/2018 – 07/2020, Urbana, IL

- Co-designed **Acto** (*SOSP'23*), the first end-to-end testing framework for Kubernetes controllers.
  - Evaluated on **11 major operators**, finding **56 new serious bugs** (42 confirmed, 30 fixed), plus **6 additional bugs** in Kubernetes/Go runtime.
  - Achieved a **false alarm rate of only 0.19%** in nightly campaigns on an 8-machine cluster.
  - Open-sourced tool (**127 GitHub stars**, featured at **KubeCon**, covered in **USENIX ;login**) and adopted by the operator developer community.

Teaching Experience

University of Michigan, CSE Department
Teaching Assistant, CSE 582: Advanced Operating Systems
Fall 2025, Ann Arbor, MI
University of Illinois Urbana-Champaign
Teaching Assistant, CS 425: Distributed Systems
Spring 2023, Haining, China

Awards & Honors

- **Michigan ADVANCE Translational Research Grant** (2025)
- **PyTorch Conference Poster Acceptance** (2025)
- **Tomorrow Star Award**, Microsoft Research Asia (2023)
- **Zhejiang University Scholarship, China** (2020, 2021)
- **Learn Student Ambassador**, Microsoft (2019–2021)