Ken/Yuangang Li

yuanganl@uci.edu | (+1) 213-421-9602 | yuangang.me | Irvine, CA

SUMMARY

Yuangang Li is a Ph.D. student at UCI with research and engineering capabilities in AI, specializing in **Agent Evaluation and Optimization**, **Trustworthy LLMs, LLM Reasoning,** and **AI for Software Engineering**. He has conducted research at Stanford, USC, UVA, and Texas Instruments @ Kilby Labs, and previously worked at SenseTime developing production-grade **cloud-native infrastructure**. He brings a combination of deep research expertise and hands-on production experience. He also serves on the program committee for top-tier conferences, including KDD 2025 (Outstanding Reviewer) and 2026 (Senior Area Chair), ICLR 2026, ICML 2025, AAAI 2026, ACL 2026, CAIS 2026, MLSys 2026, IEEE MLSP 2025, ICASSP 2025, WWW 2026, and NeurIPS 2026

PUBLICATIONS

- [1] ***Harbor: A framework for evaluating and optimizing agents and models in container environments***(GitHub). Authors**: Yuangang Li** with Harbor Framework Team.
- [2] ***Harbor Adapters and Harbor-Mix: Infrastructure and a Curated Meta-Dataset for Large-Scale Agentic Evaluation*** *(GitHub, Submit neurips 2026). Authors: Yuangang Li with Harbor Adapters Team.*
- [3] ***Agents' Last Exam*** *(arxiv, Submit neurips 2026). Authors: Yuangang Li with Team.*
- [4] ***Beyond Output Correctness: Benchmarking and Evaluating Large Language Model Reasoning in Coding Tasks*** *(arxiv, , Submit EMNLP 2026).* Authors**: Yuangang Li,** et.al.
- [5] ***Mitigating Hallucinations in LLMs via Causal Reasoning*** *(AAAI 2026).* Authors: **Yuangang Li**, et.al.
- [6] ***S2D-Align: Shallow-to-Deep Auxiliary Learning for Anatomically-Grounded Radiology Report Generation*** *(AAAI 2026).* Authors: Jiechao Gao, Chang Liu, **Yuangang Li**
- [7] ***LLM-Guided Semantic Bootstrapping for Interpretable Text Classification with Tsetlin Machines*** *(ACL 2026).* Authors: Jiechao Gao, Rohan Kumar Yadav, **Yuangang Li**, Jie Wang
- [8] ***NLP-ADBench: NLP Anomaly Detection Benchmark*** *(EMNLP 2025).* Authors: **Yuangang Li**, et.al.
- [9] ***AD-LLM: Benchmarking Large Language Models for Anomaly Detection*** *(ACL 2025).* Authors: Tiankai Yang, Yi Nian, Shawn Li, Ruiyao Xu, **Yuangang Li**, et.al.
- [10] ***A Large-Scale Simulation on Large Language Models for Decision-Making in Political Science (arxiv).*** Authors: Chenxiao Yu, Jinyi Ye, **Yuangang Li**, Zheng Li, et.al.
- [11] ***H-FedSN: Personalized Sparse Networks for Efficient and Accurate Hierarchical Federated Learning for IoT Application*** *(arxiv).* Authors: **Yuangang Li***, et.al.
- [12] ***FedMetaMed: Federated Meta-Learning for Personalized Medication in Distributed Healthcare Systems*** *(BIBM 2024).* Authors: Jiechao Gao, **Yuangang Li**
- [13] ***FedLDR: Federated Local Data-infused Graph Creation with Node-centric Model Refinement*** *(ICDM 2024).* Authors: Jiechao Gao, **Yuangang Li**, et.al.
- [14] ***Artificial Intelligence-Aided Digital Twin Design: A Systematic Review*** *(Preprints).* Authors: **Yuangang Li**, et.al.
- [15] ***FedBCGD: Communication-Efficient Accelerated Block Coordinate Gradient Descent for Federated Learning*** *(ACM MM 2024).* Authors: Junkang Liu, Fanhua Shang, Yuanyuan Liu, Hongying Liu, **Yuangang Li**, et.al.

RESEARCH EXPERIENCES

University of California, Irvine | Research Assistant *09.2025-Present*

Research in Coding Agent and Scalable AI/Agent Evaluation

- Core contributor to Harbor, large-scale AI agents evaluation and optimizing framework, 3,400+ stars on GitHub [1]
- Core member in developing the Harbor Adapters infrastructure and the Harbor-Mix curated meta-dataset for large-scale agentic evaluation, and co-designing the "Agents' Last Exam" benchmark to systematically assess advanced AI agents[2,3]
- Evaluation of LLMs coding reasoning quality, focusing on coherence and causal consistency in step-by-step reasoning[4]

University of Southern California | Research Assistant *07.2023-08.2025*

Research in Trustworthy AI

- Proposed CDCR-SFT, a novel **fine-tuning method** that internalizes causal DAG-based reasoning, boosting **causal reasoning to 95.3% (from 72.9%)** on the CLADDER benchmark, and **reducing hallucinations by 11%** on HaluEval[5]
- Proposed NLP-ADBench, the **first comprehensive NLP anomaly detection benchmark**, establishing a standardized foundation by showing the strong effectiveness of LLM-based features across 19 methods and 8 datasets[8]
- Developed AD-LLM to evaluate LLMs for NLP-AD, showing strong generalization in zero-shot detection, data augmentation, and model selection, and positioning LLMs as a robust backbone for future NLP-AD systems[9]
- Designed a multi-step reasoning framework for LLMs to simulate voter decisions, cutting election prediction error by 77% and exposing bias and overfitting issues in models like GPT-4o and LLaMA 3.1[10]

Stanford University | Research Assistant *05.2025-08.2025*

Research in Trustworthy AI and Explainable AI

- **Fine tune (LoRA, DPO, etc.) open-source LLMs** on the first self-constructed dataset with causal structures, improving causal reasoning and mitigating hallucinations, offering the first empirical evidence linking causality to hallucination[5]
- Proposed S2D-ALIGN, a progressive fine-tuning framework that learns anatomically grounded alignment from coarse to fine using reference reports and anatomical key phrases, achieving SOTA results on MIMIC-CXR and IU X-RAY[6]

- Developed an LLM-guided semantic bootstrapping method that distills LLM knowledge into interpretable Tsetlin Machines, achieving BERT-level accuracy without embeddings or inference, while preserving clause-level transparency[7]

University of Virginia | Research Collaboration *12.2023-12.2024*

Research in Effective Machine Learning Systems

- Designed an automated method to generate optimally compressed LLMs, focusing on enhancing model accuracy and deployment efficiency through innovative quantization and calibration techniques
- Proposed H-FedSN, a personalized sparse network method for hierarchical federated learning in IoT, reducing communication cost by up to 238× while preserving accuracy via structured masking and Bayesian aggregation[11]
- Developed the FedMetaMed, integrating federated learning and meta-learning to enhance personalized medication strategies across distributed healthcare systems, improving model adaptability and privacy preservation[12]
- Single-handedly developed FedLDR, a brand new federated learning algorithm that employs GCN to enhance spatio-temporal data analysis through local data integration and node-centric optimization[13]
- FedMetaMed[12] and FedLDR[13] were accepted by BIBM and ICDM respectively, and I gave talks on these research

Chinese Academy of Sciences | Research Assistant *05.2021-09.2021*

Research in Software Engineering

- Contributed to the R&D of a web-based automated testing tool using Record and Playback technology, significantly enhancing test case management by enabling streamlined recording, editing, execution, analysis, and result generation. This implementation boosted end-to-end testing efficiency by 300% and saved over 15 hours per week

WORK EXPERIENCES

Texas Instruments @ Kilby Labs | Machine Learning Research Intern *06.2026-Present*

Coding agents and agentic AI systems for complex design task

- Developed a multi-agent AI system for automated chip design on custom silicon, delivering a **500x** reduction in design cycles (from months to hours) and surpassing human expert designs
- Designed a multi-agent Reinforcement Learning method to post-train Qwen3.6-27B on chip design tasks, matching the capability of frontier models (Claude Opus) while reducing external LLM API token cost to zero

University of Southern California | Research Engineer *03.2023-07.2023*

Developing a system for ML distributed execution

- Contributed to "Ablator", an open-source Deep Learning framework used by 40+ USC researchers for horizontal scaling of ablation experiments and hyperparameter tuning, encompassing 70 pull requests
- Implemented distributed experiment execution with **Ray**, managed open-source projects, set up **CI pipelines** via GitHub Actions, oversaw release management and version control, and authored **pytest** unit tests with **97%** coverage

SenseTime | Infrastructure Engineer *01.2022-01.2023*

Developing a SaaS Platform for AI and production-grade Application

- Built scalable, production-grade AI infrastructure with Kubernetes-native tooling, cloud-native messaging systems, and GPU-accelerated ML pipelines
- Developed "RocketMQ as a Service", akin to "RabbitMQ as a Service" in AWS Marketplace, offering fully managed SaaS-based RocketMQ clusters, increasing **300%** creation speed and saving **10+ hours/week** in manual operations
- Utilized Operator SDK to build a **Kubernetes**-based RocketMQ Operator and CRD automating lifecycle management
- Integrated **Prometheus** & **Grafana** for real-time monitoring, **cutting 35%** downtime and **boosting 40%** failure detection
- Optimized a machine learning training pipeline using **GPUs on Kubernete**s for enhanced computational efficiency
- Researched container runtimes (sysbox, crun, youki) to enhance **AI workload** isolation and resource management in K8S

Xiaoniu Translations (Beijing) Technology Co., Ltd. | FullStack Developer *01.2021-04.2021*

- Developed an AI document translation system with Java/Spring/Maven, independently created a PDF/XML parsing module attracting 30,000 MAUs, used Nginx for reverse proxy, and managed version control with Git

PROJECT EXPERIENCES

Large-Scale Agent Evaluation and Optimization Framework {Code} | *Agent, Reinforcement Learning* *11.2025-Present*

- Integrated of Multi-SWE-Bench into Harbor, a multilingual benchmark featuring over **1600+** real-world software issues
- Supported RL training pipelines and SFT trace generation through with SkyRL and GEPA for agent optimization

SLinux OS (Similar to Linux 0.11) {Code} | *C, File System, System Call Interface* *06.2023-09.2023*

- Developed a functional Unix-like Operating System, providing basic os functions and ensuring performance after tests
- Gained hands-on experience with kernel development, file systems, process scheduling, and other core OS concepts
- Developed modules for memory management, process scheduling, file system management, and process communication
- Implemented multi-threading and concurrency control to manage and execute multiple processes simultaneously

Job Posting Analytics System {Demo} | *ETL, Data Mining, Database, Visualization, SpringBoot,Docker* *05.2019-06.2020*

- Developed a text mining system for collecting and analyzing recruitment data, which won the "National Level Innovative Excellence Project" award and was adopted by the university
- Automated a **ETL** pipeline to scrape, deduplicate (Bloom Filters), and store **10M+** recruitment records

EDUCATION

University of California, Irvine | Ph.D. in Software Engineering *09.2025 - 06.2029*

University of Southern California | M.S. in Computer Science *(GPA: 3.81/4.0)* *01.2022 - 05.2024*