

# Yuning Chen

✉ cyn10221017@gmail.com • 🌐 ycucm.github.io • in LinkedIn • 📄 Google Scholar

## RESEARCH INTEREST

---

LLM post-training, generative recommendation, agentic systems, semantic personalization, and sequential user modeling.

## EDUCATION

---

Ph.D. in Electrical Engineering and Computer Science, **University of California, Merced** Aug. 2021 - May. 2025  
Advisor: Prof. [Wan Du](#). Research Topic: Machine Learning; Generative AI; Networked Systems.

B.E. in Software Engineering, **Tsinghua University** Aug. 2017 - May. 2021  
Advisor: Prof. [Mingsheng Long](#). Honor: Outstanding Bachelor's Graduate (10%).

## PROFESSIONAL EXPERIENCE

---

**Research Scientist, Machine Learning @ Meta, Bellevue, WA** May. 2025 - Now  
Team: Core Ads Growth. *Ads Ranking, SFT, RL post-training, Agentic Post-training, Generative Recommendation.*

### Project: Semantic Personalization for Lead Ads Quality

- Designed and deployed an LLM-based pipeline to analyze Q&A in user-submitted lead forms and infer the commerce intent level. Leveraged RL post-training (GRPO) to enhance cross-answer analysis and attribution. Designed the reward to be a combination of rubrics and LLM-as-a-judge to mitigate reward hacking. Implemented hierarchical KV cache to boost throughput by 545.9%.
- Utilized the output as a proxy label to enhance the ads reranking teacher model with multi-task learning. Improved normalized entropy by 1.96%. Obtained stat-sig win in production A/B tests with sanitized ads value improved by 11.6%.

### Project: Self-evolving Oncall Automation Agent for Lead Ads Serving

- Built an internal long-horizon tool-calling agent to automate oncall alert diagnosis for the org, enabling root-cause analysis and auto-fix for >40 types of tasks of >10 minimum steps. Improved alert response speed by 60% and saved 520 engineering hours/year.
- Implement trajectory-level SFT and step-level Agentic post-training (HGPO w/ step attribution) for customization. Rolled out >42k trajectories during the Agentic RL on 8xA100 machines to eliminate reward hacking and achieve converged performance.
- Enabled the system to self-evolving by enriching the record docs by labeling each alert via reading actions and discussions in work chats. For new categories of alerts, designed and triggered a Trace2Skill module to learn new analysis pattern as additional skill files. Maintained long-term performance and improving long-term accuracy by 18.1% versus w/o self-evolving.

### Project: Generative Recommender as User Simulator for Ads Ranking Training Data Synthesis

- Developed a generative recommender framework as a high-fidelity user behavior simulator, with SFT + RL post-training with GSPO to improve long-sequence generation quality and new-user synthesis via composing and perturbing preference profiles to improve cold-start, long-tail, and out-of-distribution coverage.
- Demonstrated that the synthesized traces provide high-fidelity user profiles and transferable training signal for downstream sequential recommenders, improving Pass@32 by 8.4%-13.1%, while also supporting offline simulator-based RL policy learning.

### Project: Refining Sequential User Modeling for Lead Ads Quality

- Developed attention-based sequential modeling components to capture multi-step user intent across user interaction events in the lead generation funnel, improving normalized entropy by 1.30% with stat-sig online wins.
- Implemented Bayesian optimization based pipelines for automated feature selection across embedding and sparse features in large-scale ranking models, improving training efficiency and stability across >100 models.

**Software Engineer Intern @ Bytedance, San Jose, CA** May. 2024 - Aug. 2024  
Advisor: Dr. [Yongping Tang](#). Machine Learning for Systems, Network Engineering, Software-Defined Networks.

- Designed and developed a proactive ingress behavior probing tool for TikTok DNS and CDN users on IP anycast. After the successful deployment on 10+ edges globally, an ingress report for more than 240k DNS resolver IPs can be generated within 10 minutes.

**Research Intern @ Alibaba Cloud U.S., Sunnyvale, CA** May. 2023 - Aug. 2023  
Advisor: Dr. [Pan Hu](#), Dr. [Yunfei Ma](#). Code Generation, Large Language Model Evaluation, Cloud Computing, IP Anycast.

- Proposed a benchmark with 1011 problems for cloud-native configuration generation and a scalable evaluation platform for code assessment, and analyzed the performance of 13 LLMs, offering strategies for quality and efficiency improvement.
- Built an IP anycast optimization system that formulates the AS path prepending optimization as a weighted MAX-SAT problem and proposed an efficient scanning method to obtain the potential connections for global IP anycast clients.

## PUBLICATIONS

---

(\* denotes equal contribution)

- Kang Yang, [Yuning Chen](#), Wan Du. **Generalizable Radio-Frequency Radiance Fields for Spatial Spectrum Synthesis**. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) 2026. [Code](#)
- Minyuan Zhou\*, [Yuning Chen](#)\*, Yifei Xu, Jiaqi Zheng, Guihai Chen, Wanchun Dou, Pan Hu, Yongping Tang, Wendong Yin, Jie Lin, Qingyan Yu, Yuanhao Su, Songwu Lu, Wan Du. **AnyPro: Preference-Preserving Anycast Optimization based on Strategic AS-Path**

**Prepending.** USENIX Symposium on Networked Systems Design and Implementation (NSDI) 2026. [Code](#)

- Kang Yang\*, Yuanlin Yang\*, Yuning Chen\*, Sikai Yang, Wan Du. **SoilX: Calibration-Free Comprehensive Soil Sensing Through Contrastive Cross-Component Learning.** ACM/IEEE International Conference on Embedded Artificial Intelligence and Sensing Systems (SenSys) 2026. [Best Paper Award](#). [Code](#)
- Yuning Chen, Kang Yang, Zhiyu An, Brady Holder, Luke Paloutzian, Khaled M. Bali, Wan Du. **MARLP: Time-series Forecasting Control for Agricultural Managed Aquifer Recharge.** ACM KDD 2024. [Code](#)
- Yifei Xu\*, Yuning Chen\*, Xumiao Zhang\*, Xianshang Lin, Pan Hu, Yunfei Ma, Songwu Lu, Wan Du, Z. Morley Mao, Dennis Cai, and Ennan Zhai. **CloudEval-YAML: A Practical Benchmark for Cloud Configuration Generation.** MLSys 2024. [Code](#)
- Yifei Xu, Xumiao Zhang, Yuning Chen, Pan Hu, Xuan Zeng, Zhilong Zheng, Xianshang Lin, Yanmei Liu, Songwu Lu, Z. Morley Mao, Wan Du, Dennis Cai, Ennan Zhai, and Yunfei Ma, **Roaming Free in the VR World with MP2.** USENIX ATC 2025.
- Kang Yang, Yuning Chen, and Wan Du. **OrchLoc: In-Orchard Localization via a Single LoRa Gateway and Generative Diffusion Model-based Fingerprinting.** ACM MobiSys 2024 (Extended version published in ACM TMC). [Code](#)
- Kang Yang\*, Yuning Chen\*, Xuanren Chen, and Wan Du. **Link Quality Modeling for LoRa Networks in Orchards.** ACM/IEEE IPSN 2023 (Extended version published in ACM TOSN). [Code](#)
- Yuanlin Yang\*, Yuning Chen\*, Kang Yang\*, Sikai Yang, Wan Du. **Demo: Comprehensive Wireless Soil Component Sensing Through Contrastive Learning.** ACM SenSys 2025. [Best Demo Award](#). [Code](#)
- Yifei Xu\*, Yuning Chen\*, Xumiao Zhang\*, Xianshang Lin, Pan Hu, Yunfei Ma, Songwu Lu, Wan Du, Z. Morley Mao, Ennan Zhai, and Dennis Cai. **CloudEval-YAML: A Realistic and Scalable Benchmark for Cloud Configuration Generation.** Machine Learning for Systems Workshop, NeurIPS 2023. [Code](#)
- Zhizhang Hu\*, Shangjie Du\*, Yuning Chen, Xuan Zhang, Wan Du, Asa Bradman, Shijia Pan. **Poster: Enhancing Fault Resilience of Air Quality Monitoring in San Joaquin Valley: A Data Equity Analysis.** ACM SenSys 2023.
- (1st author, under submission) **Trace2Skill: Continual Skill Learning for LLM Agents from Interaction Traces.**
- (1st author, under submission) **Efficient RL Post-Training for Commercial Intent Extraction from Lead Ads Q&A.**
- (1st author, under submission) **GenRTC: Enabling Generative Visual Captions in Live Video Streaming.**
- (1st author, under submission) **GrangerNet: Multi-variate Time Series Modeling with Structural Granger Causality.**

## TECHNICAL SKILLS

---

- **Programming:** Python, C++, Go, JavaScript, SQL, MATLAB.
- **ML & AI:** LLM post-training, supervised fine-tuning, reinforcement learning, reward modeling, agentic systems, large embedding models, PyTorch, vLLM, Hugging Face, TensorFlow.
- **Systems & Infrastructure:** Linux, Kubernetes, Git, WebRTC, OpenCV, FFMPEG, MySQL.
- **RL System Architecture:** VeRL, SLIME, OpenEnv.

## ACADEMIC SERVICES

---

- **Conference Area Chair:** KDD'25-26 (Applied AI).
- **Conference Program Committee:** MLSys'26 (Agentic Systems).
- **Conference Reviewer:** NeurIPS'24-25, KDD'25-26, ICML'25, ICLR'25-27, AISTATS'25-26, MM'25, IMWUT'26, CIKM'24 TRAI.
- **Conference Artifact Evaluation Committee:** MobiCom'24-25, MobiSys'24-25.
- **Journal Reviewer:** IEEE TPAMI, IEEE TMC, IEEE TNSE, ACM TOSN, IEEE IoTJ, ACM TIOT.