Huan Wang

📍 Palo Alto, CA   ✉ joyousprince@gmail.com   🔗 huan-december.github.io   in huanwangyale

🎓 Google Scholar

Research Interests

Machine Learning, Natural Language Processing, Computer Vision, Large Action Models, AI Agents, Reasoning, and Planning.

Education

Ph.D.	**Yale University**, Computer Science	2008–2013

- Advisor: Prof. Daniel Spielman ↗, Prof. John Wright ↗
- Research on clustering, regression, and dictionary learning

M.Phil.	**The Chinese University of Hong Kong**, Information Engineering	2005–2007

- Advisor: Prof. Xiaoou Tang ↗, Prof. Shuicheng Yan ↗, Prof. Jianzhuang Liu ↗
- Research on face recognition, manifold learning, subspace learning, semi-supervised learning

B.Eng.	**Zhejiang University**, Information Engineering, Chu Kochen Honors College	2000–2004

Professional Experience

Senior Director, Salesforce Research — Palo Alto, CA — Feb 2025 – Present

- Large Action Models (xLAM), AI Agents, Reasoning, Planning, SWE Agents, Agentforce

Director, Salesforce Research — Palo Alto, CA — Nov 2021 – Feb 2025

- Led teams in AI for Operational Research (AIOps), AI for Software, Conversational AI, Time-Series Anomaly Detection, Uncertainty Estimation, and Data Hardness Evaluation

Senior Manager, Salesforce Research — Palo Alto, CA — Nov 2019 – Nov 2021

- Deep learning theory and applications, reinforcement learning, multi-task learning, language modeling, multilingual NER, knowledge graph

Senior Research Scientist, Salesforce Research — Palo Alto, CA — 2018 – Nov 2019

- Deep learning research in large-scale language modeling and vision-language integration

Senior Applied Researcher, Microsoft AI Research — Sunnyvale, CA — 2015 – 2018

- Developed deep learning systems for recommendation, ranking, and intelligent Q&A
- Improved web and local search relevance using neural embeddings

Research Scientist, Yahoo! Labs — New York, NY — 2013 – 2015

- Designed large-scale ML algorithms for search ads prediction and account security prediction
- Leveraged Hadoop, Spark, and Storm for data processing

Adjunct Professor, New York University

New York, NY
2012 – 2013
- Taught the "Machine Learning" class.

Adjunct Professor, Baruch College

New York, NY
2012 – 2013
- Taught the "Algorithm Design" class.

Research Intern, Microsoft Research

Redmond, WA
2011 – 2011
- Projects on anomaly detection and time series modeling

Research Intern, Microsoft Research Asia

Beijing, China
2010 – 2010
- Projects on anomaly detection, image modeling, and dictionary learning

Honors and Awards

- Best Paper Award, Conference on Learning Theory (COLT), 2012
- Award of Excellence – Stars of Tomorrow, Microsoft Research, Asia
- Bachelor's Degree with Honors, Zhejiang University

Representative Publications

Full publication list: Google Scholar ↗

Large Language Models (LLMs)

- APIGen-MT: Agentic Pipeline for Multi-Turn Data Generation via Simulated Agent-Human Interplay. ↗ NeurIPS Datasets & Benchmarks Track, 2025. (*co-corresponding author*), Data ↗, Model ↗
- APIGen: Automated Pipeline for Generating Verifiable and Diverse Function-Calling Datasets. ↗ NeurIPS, 2024. [Data] ↗, [Model] ↗
- xLAM: A Family of Large Action Models to Empower AI Agent Systems. ↗ NAACL, 2024. (*co-corresponding author*), [Code] ↗
- Retroformer: Retrospective Large Language Agents with Policy Gradient Optimization ↗, ICLR, 2024. (*co-corresponding author*)

AI Agents and Multi-Agent Systems

- AgentLite: A Lightweight Library for Building and Advancing Task-Oriented LLM Agent System. ↗ Arxiv, 2024. [Code] ↗
- CRMArena: Understanding the Capacity of LLM Agents to Perform Professional CRM Tasks in Realistic Environments. ↗ NAACL, 2025. [Code] ↗
- MCPEval: Automatic MCP-based Deep Evaluation for AI Agent Models. ↗ Arxiv, 2025. [Code] ↗
- REX: Rapid Exploration and eXploitation for AI Agents. ↗ Arxiv, 2023.

LLM Reasoning

- Language Models are Hidden Reasoners: Unlocking Latent Reasoning Capabilities via Self-Rewarding. ↗ Arxiv, 2024.
- PRACT: Optimizing Principled Reasoning and Acting of LLM Agent. ↗ SIG CoNLL, 2024.
- LATTE: Learning to Think with Vision Specialists. ↗ Arxiv, 2024.

Reinforcement Learning

- Policy Finetuning: Bridging Sample-Efficient Offline and Online Reinforcement Learning. ↗ NeurIPS, 2021.
- WarpDrive: Extremely Fast End-to-End Deep Multi-Agent Reinforcement Learning on a GPU. ↗ arXiv, 2021.
- On the Generalization Gap in Reparameterizable Reinforcement Learning. ↗ ICML, 2019.

Uncertainty Estimation & Reliability

- Improved Online Conformal Prediction via Strongly Adaptive Online Learning. ↗ ICML, 2023.
- Understanding the Under-Coverage Bias in Uncertainty Estimation. ↗ NeurIPS, 2021.

Earlier Work in Machine Learning and Sparse Representation

- Exact Recovery of Sparsely-Used Dictionaries. ↗ COLT, 2012. ***Best Paper Award.***
- Trace ratio vs. ratio trace for dimensionality reduction. ↗ CVPR, 2007.
- Adaptive Dropout with Rademacher Complexity Regularization. ↗ ICLR, 2018.

Skills

Technical: Machine Learning, Deep Learning, Reinforcement Learning, NLP, Computer Vision, Algorithms, AI Agents, Data Mining

Programming: Python, PyTorch, TensorFlow, Spark, Hadoop

Languages: English, Chinese