Yutong Dai

roth.rbind.io · rothdyt@gmail.com · GScholar

Education

Lehigh University, PA, USA — Sept.2019 – Dec.2023
Ph.D. in Industrial and Systems Engineering

University of Illinois at Urbana-Champaign, IL, USA — Sept.2017 – May.2019
M.S. in Statistics

Sichuan University, Chengdu, China — Sept.2013 – Jun.2017
B.S. in Mathematics with honors (concentration in Statistics)

Working Experience

Salesforce — Jan.2024 – Now

Applied Scientist → Senior Applied Scientist — Palo Alto, CA

- **Computer-Use Agents**. Built computer-use agents to achieve strong performances in academic and enterprise benchmarks by improving **grounding, planning, and action space**; built a Salesforce computer-use benchmark (SCUBA) to comprehensively evaluate agents in enterprise CRM tasks.
- **On-device Blip-3**. Built on-device versions of the multi-modal large language model Blip-3 via quantization; supported field service use-cases.
- **Text-to-Flow Generation**. Built proprietary models for generating complex Salesforce Flow using natural languages. The product is general availability (GA) on 02/2025 and nearly 8,000 customers enrolled in 2 months.

Adobe — May.2023 – Aug.2023

Machine Learning Engineer Intern — San Jose, CA

- Designed a multi-objective optimization algorithm to production data to improve both the ranking lists quality and diversity; with hit@k metric increased from 8% to 20% and the diversity score increased by up to 43%.

Salesforce — May.2022 – Aug.2022

Research Intern — Palo Alto, CA

- Proposed a novel method to tackle *data heterogeneity* with the class imbalance in *personalized Federated Learning* by combining the uniformity and semantics of class prototypes; published in AAAI2023.

Selected Projects & Publications

Computer/Browser-Use Agents

My works focus on two aspects: 1) improving the agents' performance in terms of the task success rates, efficiency, and robustness; 2) building evaluation tasks and a reinforcement learning environment for agents.

SCUBA: Salesforce Computer Use Benchmark [project page] [media coverage] 2025

- Built on a realistic and live Salesforce sandbox environment, this benchmark is designed to evaluate computer/browser-use agents' performances on the enterprise CRM tasks via three dimensions: accuracy, latency, and costs.

WALT: Web Agents that Learn Tools [tech report] [media coverage] 2025

- By reverse-engineering website functionality into reusable tools, this work advocates shifting the sequences of fragile low-level GUI actions to reliable tool invocation. WALT achieves the SOTA performance with fewer steps in VisualWebArena benchmark, and the average success rate is 52.9%. This established a robust and efficient paradigm for browser automation.

CoAct-1: Computer-using Agents with Coding as Actions [tech report] [media coverage] 2025

- To address the efficiency and reliability issues for complex and long-horizon tasks, we introduce a more robust and flexible agentic paradigm: combining GUI-based control with direct programmatic execution. The approach is the first to achieve more than 60% success rates on the OSworld benchmark and remains SOTA until 09/2025.

GTA-1: GUI Test-time Scaling Agent [tech report] [media coverage] 2025

- We improve the computer-use agents' planning via test-time scaling to select the most appropriate action proposal; meanwhile, the agents' grounding ability is boosted via the reinforcement learning (RL), the simple rewarding design. GTA-1 achieve more than 60% success rates on the OSworld benchmark and remains SOTA until 09/2025.