

Yichen Feng

Ph.D. Student, University of Washington
<https://yichenfeng.me> | yfeng42@uw.edu

EDUCATION

University of Washington

Ph.D. Student

Master of Science in Electrical and Computer Engineering

Seattle, USA

2026–Present

2024–2026

Advisor: Professor Radha Poovendran

Research Interest: RSI, LLM Agent, Synthetic Data, LLM Evaluation

I explore the boundaries of AI intelligence and how to extend them, by measuring what models cannot yet do and building what helps them improve. My research asks: How far can AI go, and how do we know it is truly getting better?

Industry Experience:

Amazon Web Services (Seattle), Agent-Driven SQL Generation and Web Visualization, 2025 Summer

University of Electronic Science and Technology of China &

Chengdu, China

University of Glasgow

B.Eng. in Electronics and Electrical Engineering

2020–2024

RESEARCH EXPERIENCE

OpenRSI Index (*Project initiator and co-lead*)

Building an open index of real-world research tasks that evaluates whether AI agents can drive recursive self-improvement and advance scientific discovery, at scales from a single node to thousands of GPUs.

Ongoing project. [Project](#) | [Code](#)

JobBench (*Adopted by Meta Muse Spark, Kimi K3, and Qwen3.8*)

Developed a benchmark of 130 agentic tasks across 35 occupations, evaluating workflows that human experts prioritize for delegation and grading outputs with task-specific rubrics.

NeurIPS 2026 ED Track. [Paper](#) | [Project](#)

Visual Aesthetic Benchmark: Can Frontier Models Judge Beauty?

Introduced an expert-grounded benchmark for visual aesthetics across art, photography, and illustration, revealing a substantial gap between 20 frontier vision-language models and expert judgment.

COLM 2026. [Paper](#) | [Project](#)

VisualSphinx: Large-Scale Synthetic Vision Logic Puzzles for RL

Created large-scale synthetic visual logic puzzles using a rule-to-image pipeline. Applied reinforcement learning to improve vision-language models' logical reasoning and transfer across reasoning tasks.

DMLR 2026; DataWorld Workshop at ICML 2025. [Paper](#) | [Project](#)

BadScientist (*Best Paper Award, Agents4Science 2025*)

Studied whether automated research agents can produce convincing but unsound papers that deceive LLM reviewers, and evaluated weaknesses in automated peer review.

ACL 2026. [Paper](#) | [Project](#)

WebUniverse: Scaling Verified Training Data for Agentic Web Development

Built a pipeline that turns open-source repositories and real web pages into 15K verifiable front-end development tasks with verified agent trajectories; fine-tuning Qwen3.5-35B-A3B on them raises WebGen-Bench accuracy from 34.5% to 63.9%.

Preprint, 2026. [Paper](#) | [Project](#)

RAC: A Relational Agent Contract for Constraint-Governed Engineering Tasks

Introduced a contract in which engineering agents choose relational constraints and specialized tools derive exact consequences, raising the PCB clean-pass rate across nine models to 90.9% from 74.0% for the best baseline.

Preprint, 2026. [Paper](#)

TAPDreamer: Transferable Adversarial Patches for World Action Models

Proposed a transferable adversarial patch, optimized with only a public visual encoder, that cuts world action model FastWAM's success from over 90% to 0% without querying the target policy.

Preprint, 2026. [Paper](#) | [Project](#)

Polyhedral Instability Governs Regret in Online Learning

Characterized how changes in active polyhedral regions govern regret in online learning, with matching theoretical bounds and experiments on combinatorial decision problems.

Preprint, 2026. [Paper](#)

The WidthWall: A Strict Expressivity Hierarchy for Hypergraph Neural Networks

Established a strict expressivity hierarchy for hypergraph neural networks, identifying structural limitations of bounded-width architectures and motivating density-aware models.

Preprint, 2026. [Paper](#)

HONORS & AWARDS

Best Paper Award, Agents4Science 2025 (*BadScientist*)

ACADEMIC SERVICE

Initiator and Tech Lead: [OpenRSI Foundation](#)

Student Advisory Board Member: ACTION NSF AI Institute

Conference Reviewer: ICML, NeurIPS, ACL, ICLR