Fengyuan Zhu

Research Assistant Professor, Siebel School of Computing and Data Science
The Grainger College of Engineering, University of Illinois Urbana-Champaign
zhufy@illinois.edu | fengyuan-zhu.com

Research Interests

I build and measure AI agents that carry out parts of scientific research, running on self-hosted open-weight models so that every result can be repeated.

- **Research automation.** Agents that reproduce published results from the paper alone and map large research literatures.
- **Cost and capability of open-weight models.** Serving cost of agent workloads across GPU generations, token overhead, and contamination-controlled benchmarks of what open-weight models can do.
- **Tools.** Software that lets agents operate real systems and lowers the cost of deploying models.

Appointments

University of Illinois Urbana-Champaign — Urbana, IL
Research Assistant Professor, Siebel School of Computing and Data Science — Aug 2026 – present

Education

University of Toronto — Toronto, Canada
Ph.D., Computer Science (Human-Computer Interaction). Advisor: Tovi Grossman — 2018 – 2025
New York University — New York, NY
M.P.S., Interactive Telecommunications Program, Tisch School of the Arts — 2015 – 2017
Peking University — Beijing, China
B.S., Physics — 2011 – 2015

Professional Experience

Google — San Francisco, CA / Toronto, ON
Student Researcher, Google Labs — May 2023 – Mar 2024
NVIDIA Research — Durham, NC
Research Intern — May – Aug 2018
Holojam Inc. — New York, NY
Research Engineer — Sep 2017 – May 2018

Current Research Activity

- **AI agents for research automation.** Building agent harnesses on self-hosted open-weight models that attempt to reproduce published research from the papers alone.
- **Literature mapping at field scale.** Expanding and labeling a corpus of about 125,000 abstracts on laboratory automation with open-weight models, to measure how far automation has actually spread in a field.
- **LLM-driven measurement at population scale.** Building LLM- and agent-driven pipelines that make population-scale empirical studies feasible: linking large collections of academic records into career data and labeling them with locally served open-weight models, with match-quality checks and labeling reproducibility tested across model architectures.
- **Benchmarks for open-weight models.** Building a contamination-controlled benchmark from frontier mathematics formalized in Lean after the training cutoff of the models under test, to compare open-weight and closed models on identical problems.
- **Evaluating open-weight models in practice.** Training-data contamination tests across 10 open-weight base models using competition write-ups dated around each model's cutoff; a comparison of seven open-source document OCR pipelines on 100 journal articles.

- **Price of capability over time.** Reconstructed the daily history of a public LLM leaderboard (January 2024 to present, 473 base models, open and closed weights) to track what a fixed level of capability costs, with reproducible acceptance checks against the live source.
- **Workload shape of LLM agents.** Analysis of real agent coding sessions shows serving cost is dominated by prefill (prefill:decode token ratio of roughly 300–400:1) and repeated fixed prefixes, a profile unlike chat traffic.
- **LLM serving on legacy accelerators.** Characterized inference of large open-weight models on eight Tesla V100-32GB GPUs (sm_70, PCIe, no NVLink), including a 744B-parameter mixture-of-experts model at 1-bit quantization, long-context (262,144-token) serving of a 27B model, and video diffusion models on FP16-only hardware. Results public at fengyuan-zhu.com/v100.

Publications

Full papers

[1] H. Wang, **F. Zhu**, B. Huang, Z. Wang, L. Sidenmark. Prop-Chromeleon: Adaptive Haptic Props in Mixed Reality through Generative Artificial Intelligence. *ACM DIS 2026*. **Honorable Mention Award.**

[2] **F. Zhu**, X. Qian, D. Kalmar, M. Tayarani, E. J. Gonzalez, M. Gonzalez-Franco, D. Kim, R. Du. Beyond the Phone: Exploring Phone-XR Integration through Multi-View Transitions for Real-World Applications. *IEEE VR 2025*, pp. 770–780.

[3] **F. Zhu**, M. Sousa, L. Sidenmark, T. Grossman. PhoneInVR: An Evaluation of Spatial Anchoring and Interaction Techniques for Smartphone Usage in Virtual Reality. *ACM CHI 2024*.

[4] **F. Zhu**, L. Sidenmark, M. Sousa, T. Grossman. PinchLens: Applying Spatial Magnification and Adaptive Control Display Gain for Precise Selection in Virtual Reality. *IEEE ISMAR 2023*, pp. 1221–1230.

[5] **F. Zhu**, Z. Lyu, M. Sousa, T. Grossman. Touching the Droid: Understanding and Improving Touch Precision with Mobile Devices in Virtual Reality. *IEEE ISMAR 2022*, pp. 807–816.

[6] **F. Zhu**, T. Grossman. BISHARE: Exploring Bidirectional Interactions Between Smartphones and Head-Mounted Augmented Reality. *ACM CHI 2020*, pp. 1–14.

[7] P. Li, T. Qin, B. Hu, **F. Zhu**, S. Shen. Monocular Visual-Inertial State Estimation for Mobile Augmented Reality. *IEEE ISMAR 2017*, pp. 11–21.

[8] X. Ma, K. Fang, **F. Zhu**. From Breakage to Icebreaker: Inspiration for Designing Technological Support for Human-Human Interaction. *ACM DIS 2016*, pp. 403–414.

Extended abstracts

[9] Z. He, **F. Zhu**, K. Perlin. PhyShare: Sharing Physical Interaction in Virtual Reality. *UIST 2017 Adjunct*, pp. 17–19.

[10] K. Perlin, Z. He, **F. Zhu**. Chalktalk VR/AR. *CreMedia* 2017/2 (2018), pp. 30–31.

[11] **F. Zhu**, K. Fang, X. Ma. Exploring the Effects of Strategy and Arousal of Cueing in Computer-Human Persuasion. *CHI EA 2017*, pp. 2276–2283.

[12] **F. Zhu**, W. Sun, C. Zhang, R. Ricks. BoomChaCha: A Rhythm-Based, Physical Role-Playing Game that Facilitates Cooperation among Players. *CHI EA 2016*, pp. 184–187.

Data notes (public, techwave.cs.illinois.edu, 2026)

- **F. Zhu**. Most Self-Driving Labs Don't Call Themselves That.
- **F. Zhu**. When Do Open Weights Catch Closed Models?
- **F. Zhu**, J. Liu. Anatomy of the XR Cycle.
- **F. Zhu**. What Running a Model Actually Costs (three field notes on AI adoption cost).

Patents

- K. Perlin, **F. Zhu**. Head Mounted Display and Method. US Patent 11,635,627.
- K. Perlin, **F. Zhu**. System for a Spectator to View a Virtual Reality of a Performer and Method. US Patent 11,836,285.
- R. Du, D. Kim, **F. Zhu**, D. Kalmar. Determining Input Modality of Head-Mounted Device Based on Input Modality of Handheld Device. US Patent App. 18/884,920.

Awards

2023, 2024 Doctoral Completion Award, University of Toronto
2016, 2017 Departmental Scholarship for Outstanding Academic Work, Tisch School of the Arts, NYU
2012, 2013 Departmental Scholarship, School of Physics, Peking University

Teaching and Mentoring

Instructor, Agentic AI course (planned), Siebel School of Computing and Data Science, UIUC, Spring 2027.

Teaching Assistant, University of Toronto (2019–2023): Introduction to Human-Computer Interaction (3 terms), The Design of Interactive Computational Media, Topics in Interactive Computing, Topics in Intellectual Property, Topics in Blockchain Technology.

Mentored student researchers: Harry Wang (2024–25), Grace Xu (2023–24), Zhuoyue Lyu (2019–21), University of Toronto; Wangshu Sun (2016–17), New York University.

Service

Reviewer: ACM CHI (2020, 2022–2025), ACM UIST (2021–2024), IEEE ISMAR (2020–2025), IEEE VR (2025), ACM SIGGRAPH (2025), ACM SIGGRAPH Asia (2024), ACM VRST (2021), Pacific Graphics (2024).