

MELISSA ZHIYANG PAN

+1(412) 708-6280 ◇ Berkeley, CA

melissapan@berkeley.edu

[Website](#) ◇ [LinkedIn](#)

RESEARCH SUMMARY

I design and build efficient and reliable machine learning systems through joint optimization across the systems stack. Specifically, I study how to build responsible and efficient frameworks for emerging agent systems, with sustainability as a first-order objective. At the same time, I study how to reliably harness modern language models in agent systems to accelerate scientific discovery in systems research and beyond.

EDUCATION

University of California, Berkeley, Berkeley, CA

Expected Graduation: Jun 2029

Ph.D. in Computer Science, GPA: 4.0/4.0

Advisor: Prof. Matei Zaharia

Carnegie Mellon University, Pittsburgh, PA

Graduation: Jun 2024

Master of Science in Electrical and Computer Engineering – Applied Study with Research Track, GPA: 4.0/4.0

Advisor: Prof. Akshitha Sriraman

University of Toronto, Toronto, Canada

Graduation: Jun 2020

Bachelor of Applied Science in Electrical and Computer Engineering, High Distinction

Engineering Business Minor, AI Engineering Certificate, Core GPA 3.83/4.0

SELECTED PUBLICATIONS

Refereed Conference Papers

[Measuring Agents in Production](#)

Melissa Z. Pan*, Negar Arabzadeh*, Riccardo Cogo, Yuxuan Zhu, Alexander Xiong, Lakshya A. Agrawal, Huanzhi Mao, Emma Shen, Sid Pallerla, Liana Patel, Shu Liu, Tianneng Shi, Xiaoyuan Liu, Jared Quincy Davis, Emmanuele Lacavalla, Alessandro Basile, Shuyi Yang, Paul Castro, Daniel Kang, Koushik Sen, Dawn Song, Joseph E. Gonzalez, Ion Stoica, Matei Zaharia, Marquita Ellis. *Proceedings of the 43rd International Conference on Machine Learning (ICML 2026)*, Seoul, South Korea, **Oral & Spotlight**. Earlier version presented at the *ICLR 2026 Workshop on Agents in the Wild: Safety, Security, and Beyond*. (* equal contribution)

[EvoX: Meta-Evolution for Automated Discovery](#)

Shu Liu, Shubham Agarwal, Monishwaran Maheswaran, Mert Cemri, Zhifei Li, Qiuyang Mang, Ashwin Naren, Ethan Boneh, Audrey Cheng, **Melissa Z. Pan**, Alexander Du, Kurt Keutzer, Alvin Cheung, Alexandros G. Dimakis, Koushik Sen, Matei Zaharia, Ion Stoica. *COLM 2026*.

[Why Do Multi-Agent LLM Systems Fail?](#)

Melissa Pan*, Mert Cemri*, Shuyi Yang*, Lakshya A. Agrawal, Bhavya Chopra, Rishabh Tiwari, Kurt Keutzer, Aditya Parameswaran, Dan Klein, Kannan Ramchandran, Matei Zaharia, Joseph E. Gonzalez, Ion Stoica. *NeurIPS 2025 Datasets and Benchmarks Track*, **Spotlight**. Earlier versions presented at the *ICLR 2025 Workshop on Building Trust in LLMs and LLM Applications*, the *MLSys 2025 Young Professional Symposium*, and the *NeurIPS 2025 Workshop on Evaluating the Evolving LLM Lifecycle*. (* equal contribution)

[Semantic Operators and Their Optimization: Towards AI-Based Data Analytics with Accuracy Guarantees](#)

Liana Patel, Siddharth Jha, **Melissa Z. Pan**, Harshit Gupta, Parth Asawa, Carlos Guestrin, Matei Zaharia. *VLDB 2025*.

Face Recognition and Rehabilitation: A Wearable Assistive and Training System for Prosopagnosia
Steve Mann, **Zhiyang Pan**, Yi Tao, Anqi Gao, Xingchen Tao, Danson Evan Garcia, Dawei Shi, Georges Kannan.
IEEE International Conference on Systems, Man and Cybernetics 2020, Toronto, Canada. [[Presentation](#)]

Peer-Reviewed Workshop Papers, Demos, and Posters

Fantastic Adaptive Taxonomies and How to Use Them

Mert Cemri, Andrei Cojocaru, **Melissa Z. Pan**, Shu Liu, Shubham Agarwal, Alexander Krentsel, Jay Tang, Kannan Ramchandran, Joseph E. Gonzalez, Matei Zaharia, Alex Dimakis, Ion Stoica. *ICML 2026 Workshop on Failure Modes of Agentic AI*, *Best Paper Award*.

SkyDiscover: A Flexible, Adaptive Framework for AI-Driven Scientific and Algorithmic Discovery

Shu Liu, Mert Cemri, Shubham Agarwal, Alexander Krentsel, Ashwin Naren, Qiuyang Mang, Zhifei Li, Akshat Gupta, Monishwaran Maheswaran, Audrey Cheng, **Melissa Z. Pan**, Ethan Boneh, Kannan Ramchandran, Koushik Sen, Matei Zaharia, Alexandros G. Dimakis, Ion Stoica. *ACM CAIS 2026, Demo Track*, *Spotlight*.

Towards Automatically Optimizing Retrieval Augmented AI Systems

Melissa Z. Pan, Negar Arabzadeh, Mathew Jacob, Fiodar Kazhamiaka, Esha Choukse, Matei Zaharia. *Machine Learning for Systems 2025*; also presented at the *OSDI 2025 Poster Session*.

ReServe: Giving Old Servers A Second Chance at Hyperscale

Jaylen Wang, **Melissa Pan**, Udit Gupta, Akshitha Sriraman. Presented at the *SOSP 2024 Poster Session*, the *SOSP 2024 Young Professional Symposium*, *SYSDW 2024*, *HotCarbon 2025*, and the *OSDI 2025 Poster Session*.

Under Submission

Natural Language Query to Configuration for Retrieval Agents

Melissa Z. Pan, Negar Arabzadeh, Mathew Jacob, Fiodar Kazhamiaka, Esha Choukse, Matei Zaharia. *arXiv preprint*, 2026.

Barbarians at the Gate: How AI is Upending Systems Research

Melissa Pan*, Audrey Cheng*, Shu Liu*, Zhifei Li, Bowen Wang, Alex Krentsel, Tian Xia, Mert Cemri, Jongseok Park, Shuo Yang, Jeff Chen, Lakshya Agrawal, Aditya Desai, Jiarong Xing, Koushik Sen, Matei Zaharia, Ion Stoica. *arXiv preprint*, 2025.

Let the Barbarians In: How AI Can Accelerate Systems Performance Research

Audrey Cheng*, Shu Liu*, **Melissa Pan**, Zhifei Li, Shubham Agarwal, Mert Cemri, Bowen Wang, Alexander Krentsel, Tian Xia, Jongseok Park, Shuo Yang, Jeff Chen, Lakshya Agrawal, Ashwin Naren, Shulu Li, Ruiying Ma, Aditya Desai, Jiarong Xing, Koushik Sen, Matei Zaharia, Ion Stoica. *arXiv preprint*, 2025.

PATENT PUBLICATIONS

“Method and Apparatus of Rank Based Setup and Execution of Federated Learning Workflow”, **Melissa Pan**, Derek Roy, Kairui Dong, Nazim Uddin Bhuiyanr, Rafiur Rashid, Shaikh Quader, Petr Novotny, [IPCOM000270087D](#), Jun. 02, 2022, IBM LLC.

“Method and System of Redaction of Sensitive Enterprise-Wide Data”, **Melissa Pan**, Derek Roy, Kairui Dong, Nazim Uddin Bhuiyanr, Shaikh Quader, Petr Novotny, [IPCOM000270088D](#), Jun. 02, 2022, IBM LLC.

AWARDS AND HONORS

Amazon AI Fellowship	2025 – 2027
Laude Institute AI Residency	2025 – 2026
Laude Slingshots // TWO	2026
UC Berkeley The Katherine & James Lau Electrical Engineering and Computer Science Fellowship	2024 – 2025
Carnegie Mellon University Members of IEEE Eta Kappa Nu	2023
IBM Jumpstart Judges’s Choice Best Project Award	2021
IBM DB2 Manager’s Recognition Award	2020
University of Toronto ECE Distinction Graduation Capstone Project Award	2020
University of Toronto Faculty of Engineering Dean’s Honored List	2015 – 2020
University of Toronto President’s Entrance Scholarship Award	2015

TALKS

Why Do Multi-Agent LLM Systems Fail?

LangChain, hosted by: Harrison Chase	Aug 2026
University of Oxford and the Alan Turing Institute, hosted by: Naman Goel	Apr 2026
Edge Delta AI Weekly	Feb 2026
UC Berkeley Epic Data Systems Lab	Oct 2025
Universal AGI	Sept 2025
UC Berkeley Agent AI Summit	Aug 2025
Accenture Course	Aug 2025
MIT Data Systems Group	Jul 2025
IBM Watson Lab	Jul 2025
Microsoft Research Asia, hosted by Tianyin Xu	Jun 2025
UC Berkeley BAIR-RDI Agent Workshop	May 2025
University of Copenhagen Natural Language Processing Group	Apr 2025
UC Berkeley SkyDeck Demo Day	Apr 2025
Gravity Foundation	Apr 2025
UC Berkeley Programming Systems Seminar	Mar 2025
UC Berkeley Systems Reading Group	Mar 2025

Measuring Agents in Production

University of Washington, SyFI Lab	Apr 2026
Microsoft Azure Systems Research Group, hosted by Fiodar Kazhamiaka	Apr 2026
AGI House Agent Build Day Speaker, invited by Rocky Yu	Apr 2026
Amazon Gen AI Monthly, hosted by Ghazal Jaber	Apr 2026
UC Berkeley Kannan Ramchandran’s Group Meeting	Mar 2026
UC Berkeley Sewon Min’s Group Meeting	Feb 2026
NVIDIA Tech Talk, hosted by Vinod Grover	Feb 2026
Amazon Foundation Model Team	Feb 2026
NICE: NLP Academic Exchange Program	Jan 2026
UC Berkeley Programming Systems Seminar	Dec 2025
SOSP 1st Workshop on Systems for Agentic AI	Oct 2025

Natural Language Query to Configuration for Retrieval Agents

Stanford University, Prof. Azalia Mirhoseini’s Group	Jun 2026
Arena Intelligence [link]	Jun 2026

Barbarians at the Gate: How AI is Upending Systems Research

Meta, Systems Optimization Organization	Feb 2026
---	----------

OTHER RESEARCH EXPERIENCE

Machine Learning Research Intern, Arena Intelligence

Jul 2026 – Present

Advised By: **Prof. Ion Stoica**

San Francisco, CA

- Working on coding agent efficiency measurement to push the cost–performance Pareto frontier

Graduate Researcher, Carnegie Mellon University

Aug 2023 – Present

Advised By: **Prof. Akshitha Sriraman** and **Prof. Udit Gupta**

Pittsburgh, PA

- Studied sustainable computing in hyperscale datacenters through carbon-aware scheduling of microservices on heterogeneous hardware for hardware lifetime extension
- Designed profiling experiments to characterize the relationship between microservice performance and server generations for online scheduling decisions
- Implemented RL-based and heuristic schedulers using Kubernetes and Docker Swarm, achieving 23% carbon footprint reduction while meeting latency SLOs across three popular microservice benchmarks

Software Engineer, IBM (system design & research group – side project)

Jan 2021 – Jun 2022

Mentored By: **Dr. Petr Novotny** and **Shaikh Quader**

Markham, Canada

- Worked on database-system and ML-system integration for enterprise federated learning and privacy-sensitive ML workflows
- Developed and integrated a scripting interface to streamline data transfer from the `ibm_db` framework into IBM Cloud Pak for Data’s federated learning client API
- Automated local environment setup for enterprise federated learning workflows and authored two patent publications on rank-based federated learning and sensitive-data redaction

Undergraduate Researcher, University of Toronto, Mann Lab

Aug 2019 – Oct 2020

Advised By: **Prof. Steve Mann**

Toronto, Canada

- Built a wearable assistive and training system for prosopagnosia to support both real-time interaction and long-term rehabilitation
- Designed and implemented a full system architecture with real-time face recognition, asynchronous on-device training, and a self-training mode in an Android application
- Programmed a custom eye-tracking pipeline that mapped pupil coordinates to outward camera regions for AR overlays, achieving 96% recognition accuracy within a 0.5-meter range

TEACHING & MENTORSHIP

UCB : Expanding Your Horizons conference, Creative Coding: Workshop Creator & Instructor	Spring 2025
CMU : 18-847, Data Center Computing: Graduate Teaching Assistant	Spring 2024
CMU : 18-847, Modern Computer Systems: Graduate Teaching Assistant	Fall 2023
IBM : Intern Mentor: Catherine Mo Zhou (UofT, undergraduate student)	2021 – 2022
IBM : STEM For Girls, Introduction to Coding: Course Designer & Instructor	Summer 2022
IBM : STEM For Girls, Design Thinking: Course Instructor	Summer 2021
IBM : STEM For Girls, Logical Thinking: Course Designer & Instructor	Summer 2019
BitTiger : Teaching Assistant Volunteer	2017 – 2018
Toronto District School Board : Homework Peer Mentor	2014 – 2015
Toronto District School Board : K-12 After School Education Volunteer	2013 – 2015

PROFESSIONAL SERVICE

NeurIPS : Reviewer	2026
ACM CAIS : Reviewer	2026
USENIX : OSDI Artifact Evaluation Committee (OSDI '25, OSDI '26)	2025 – 2026
UCB : NEXT Meetup Organizer	Mar 2026 – Present
UCB : Organizer for Programming Systems Seminar	May 2025 – Aug 2025

STUDENT SERVICE & LEADERSHIP

UCB: EECS Graduate Student Assembly, Co-President for CS	AY 2026 – 2027
UCB: CS Faculty Hiring Student Committee Chair	AY 2025 – 2026
UCB: EECS Graduate Student Assembly, Student Representative	2025 – Present
UCB: CS Graduate Entrepreneur, Seminar Chair	Nov 2024 – Present
UCB: CS Faculty Hiring Student Committee	AY 2024 – 2025
UCB: Graduate Women of Engineering, Career & Professional Development Co-Chair	Nov 2024 – Dec 2025
UCB: Equal Access for Application Assistance	Nov 2024
Google: Community Resources for Science – Science Kit Volunteer	Jul 2024
CMU: ECE Diversity and Inclusion Committee	Oct 2023 – May 2024
CMU: ECE Graduate Organization, General Member and Volunteer	Jan 2023 – May 2024

INDUSTRY EXPERIENCE

Software Engineering Intern (Research-Oriented) May 2024 – Jul 2024
Google, EdgeTPU *Mountain View, CA*

- Explored production opportunities for GPU-TPU co-compilation in Pixel neural networks
- Identified TFLite partitioning strategies that achieved up to 10% end-to-end inference speedup
- Designed and developed an integrated Python toolchain for TFLite model splitting and on-device evaluation across TPU and GPU

Software Engineering Intern (Research-Oriented, [github link](#)) May 2023 – Aug 2023
Google, CoreML *Sunnyvale, CA*

- Designed and implemented NCCL GPU collective clustering optimizations end-to-end in C++ within TensorFlow DTensor
- Reduced training step time by up to 5% and device idle time by 78% for BERT and T5 workloads
- Conducted 180 training experiments across five transformer settings to profile bottlenecks and evaluate optimization performance
- Authored design documentation from literature review, prior designs, and technical discussions with production leads

Software Engineer May 2018 – Sept 2019, May 2020 – Dec 2022
IBM, Data Analytics and AI *Markham, Canada*

- Built a full-scale distributed monitoring system from scratch and modernized test infrastructure for 488 test suites, reducing manual effort by 75%
- Redesigned Db2 VARCHAR data structures for analytics workloads, achieving 66% memory reduction and 10% query speedup
- Maintained C++ code in the Db2 core engine across multiple releases and represented the Backup, Restore, and Recovery domain in a cross-team initiative spanning seven teams

Backend Developer (& Technical Product Manager) Jun 2020 – Sept 2020
BeenThere – Harvard Innovation Lab *Remote*

- Developed backend services using Spring, Spring MVC, and MySQL for integration with existing frontends
- Designed relational schemas to support machine learning workflows in psychology research applications
- Led domain migration from WordPress to AWS and evaluated cloud infrastructure options

PRESS

Nebius Science: “ICML 2026: Papers That Matter” Jul 2026
The Daily Californian: “UC Berkeley researchers use AI to tackle climate problems” Mar 2026

REFERENCES

Matei Zaharia

Associate Professor, University of California, Berkeley

matei@berkeley.edu

Ion Stoica

Professor, University of California, Berkeley

istoica@berkeley.edu

Joseph E. Gonzalez

Professor, University of California, Berkeley

jegonzal@cs.berkeley.edu

Akshitha Sriraman

Assistant Professor, Carnegie Mellon University

akshitha@cmu.edu

Esha Choukse

Principal Researcher, Microsoft Azure Systems Research

Esha.Choukse@microsoft.com

Fiodar Kazhamiaka

Senior Researcher, Microsoft Azure Systems Research

fkazhamiaka@microsoft.com