MELISSA ZHIYANG PAN

+1(412) 708-6280 ◇ Berkeley, CA, United States
melissapan@cmu.edu ◇ linkedin.com/in/zhiyang-pan-melissa ◇ www.melissa-pan.github.io

RESEARCH INTERESTS

I aspire to build next-generation systems that embody responsibility in the core of computing stacks. To this end, I am passionate about systems and infrastructure tailored for machine learning and data-intensive tasks at large scale, that prioritize sustainability alongside efficiency, performance, and reliability.

EDUCATION

University of California, Berkeley, Berkeley, CA — Expected Graduation: Jun 2029
Ph.D. in Computer Science

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

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

PUBLICATIONS & INDUSTRIAL PATENT PUBLICATIONS

"Method and Apparatus of Rank Based Setup and Execution of Federated Learning Workflow", **Zhiyang 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", **Zhiyang Pan**, Derek Roy, Kairui Dong, Nazim Uddin Bhuiyanr, Shaikh Quader, Petr Novotny, IPCOM000270088D, Jun. 02, 2022, IBM LLC.

"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, Conference Paper]

RESEARCH EXPERIENCE

Graduate Researcher, Carnegie Mellon University — Aug 2023 – Present
Advised By: **Prof. Akshitha Sriraman** and **Prof. Udit Gupta** — *Pittsburgh, PA*

- *Research topic*: introducing sustainability as a first-order objective in hyper-scale data center systems via distributed microservice scheduling on heterogenous hardware to promote hardware lifetime extension
- Designed and conducted profiling experiments to identify correlation between microservice performance and hardware generations to construct scaling policy in combination with carbon intensity for online scheduling
- Developed the heterogeneous auto-scaling online scheduler using Docker Swarm to execute workload generated from real-world application traces, ensuring latency service level objectives while reducing carbon footprint

Graduate Researcher, Carnegie Mellon University, TEEL Lab — Sept 2023 – Dec 2023
Advised By: **Dr. Jaromir Savelka** — *Pittsburgh, PA*

- *Research topics*: investigating the application of GPT models for creating structured, university-level course materials in specialized, research-focused disciplines, particularly in areas with limited online resources.
- Developed Python infrastructure to pipeline the generation of learning objectives and corresponding multiple choice questions with prompt templates designated to each type of questions
- Designed rubrics to evaluate the quality of the generated multiple choice questions based on Bloom's taxonomy and curated input learning objectives iteratively based on the evaluation results

Software Engineer, IBM (system design & research group – side project) Jan 2021 – Jun 2022
Mentored By: **Dr. Petr Novotny** and **Shaikh Quader** *Markham, Canada*

- *Research topics*: innovating the combination of database systems with machine learning and machine learning systems using relational databases for enterprise-level data and applications.
- Developed and integrated a scripting interface to streamline the transfer of Db2 data from ibm_db framework into IBM Cloud Pak for Data's federated learning client API, and automated the setup process for local environments.
- Authored two full system design proposals, delivered presentations to the IP stakeholder review board and successfully submitted them for publication

Undergraduate Researcher, University of Toronto, Mann Lab Aug 2019 – Oct 2020
Advised By: **Prof. Steve Mann** *Toronto, Canada*

- *Research topic*: addressing social challenges faced by individuals with prosopagnosia through the development of assistive devices for immediate support and long-term rehabilitative training
- Designed and implemented a full system architecture with two modes being real-time face recognition implemented with asynchronous on-device training on BlazeNet and FaceNet in Python, alongside a self-training mode designed to emulate clinical treatment, developed within an Android mobile application using Java and Android Studio
- Programmed a customized eye tracker for pupil detection that translates pupil image from inward raspberry pi infrared camera to coordinates on image region of the outward facing camera, achieving a recognition accuracy of 96% within a 0.5-meter range for real-time application

SELECTED PROFESSIONAL EXPERIENCE

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

- Explored the methodology, benefits, and opportunities in Pixel production models to do GPU-TPU co-compilation
- Identified specific TFLite splits for Pixel production models to achieve up to 10% end-to-end latency speed-up
- Designed and developed an integrated tool chain in Python to enable TFLite model splitting and on-device evaluation for both 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 using C++ in Tensorflow DTensor API enabling up to 5% reduction in step time and 78% reduction in device idle time for BERT and T5
- Formulated and conducted 180 training experiments on 5 types of transformers and or models in Python to profile and evaluate optimization performance; Crafted comprehensive performance study report from analysis of the experiment result
- Authored design documentation independently from extensive research on state-of-the-art literatures in distributed computing, previous design documentation and proactive engagement in design discussions with technical leads

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

- Spearheaded end-to-end development of two new features to handle tablespace failure during backup and recovery in Db2 v11.5.7, and to optimize storage of recovery objects in Db2 v12.1; presented in montely stakeholder meeting
- Maintained C++ code for Db2 core engine in Backup, Restore and Recovery (BAR) components across six Db2 releases, core features include: availability, data replication in distributed database systems, log space management
- Represented BAR domain end-to-end in a project involving 7 teams from design to implementations; facilitated PM to drive progress by providing technical consultations on code migration and infrastructure for other teams
- Refactored and migrated 488 full domain test sets to GIT, and developed standardized Perl APIs for modernization
- Built a full-scale monitoring system for distributed infrastructure from scratch, reducing manual effort by 75%

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

- Developed web backend using Spring, Spring MVC, and MySQL for a seamless integration with existing frontends

- Designed relational database in MySQL with optimized schemas to streamline ML training in psychology research
- Led migration of the domain from WordPress to Amazon AWS and conducted analysis on various cloud providers

Software Engineer (Professional Experience "Co-op" Year) May 2018 – Sept 2019
IBM, Data Analytics and AI *Markham, Canada*

- Participated in redesign of VARCHAR data structure for Db2 blu acceleration leading up to 66% runtime memory reduction; Implemented major APIs in C++ integrated with 200+ referencing functions in the source code
- Developed and managed a new testing framework with primary functional verification tests in Perl for a new feature in Db2 led to a 10% total speed enhancement and covered SQL commands in 12 categories with 25 data types

TEACHING & MENTORSHIP

CMU: 18-847, Data Center Computing: 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

AWARDS AND HONORS

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 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

TECHNICAL SKILLS

Programming Languages	Python, C++/C, Bash, MLIR, Java, Perl, Scala, x86 Assembly, HTML, Javascript
AI & Big Data	SQL, NoSQL, Tensorflow, JAX, NumPy, Pandas, MapReduce, Spark, Jupyter
Databases & Servers	DB2, MySQL, PostgreSQL, MongoDB, Neo4j, Kubernetes, Docker, Google Cloud, Azure, AWS EC2, Drogon, Python Flask
Tools	Git, Linux perf, Vim, Xcode, Android Studio

ACTIVITIES AND OTHER EXPERIENCES

Google Community Resources for Science - Science Kit Volunteer — Jul 2024
CMU Carnegie AI Safety Initiative, Graduate Member, facilitated by: **Dr. Vojta Kovarik** — Sep 2023 – Jan 2024
CMU ECE Diversity and Inclusion Committee — Oct 2023 – May 2024
CMU ECE Grad Organization, General Member, Volunteer — Jan 2023 – May 2024
Hack the North: Participant [project link] — 2018
UofT Chinese Scholars and Students Association, VP of Media & IT — 2015 – 2016
Toronto Science Fair: Marine Ecosystem COWEB Agent-Based Modelling, Advised By: **Dr. Brad Bass** — 2015