1735 Woodland Ave, Apt 62
East Palo Alto, CA 94303
United States

Saubhik Mukherjee

Mobile: (470) 313-0534
saubhik.mukherjee@gmail.com

Employment

Senior Software Engineer — **SambaNova Systems, Palo Alto** — **May 23, 2022 - Present**
Machine Learning (ML) compiler

- Build MLIR and LLVM-based compiler layers and compiler tools to transform, optimize, debug, and execute ML models on proprietary ML accelerator architectures.
- Build scalable and high-quality production compiler infrastructure using well-established and emerging techniques and push the boundaries of compiler design.
- Develop, maintain, and debug compiler optimization algorithms on ML graphs and add compiler support for new hardware architectures.
- Analyze and improve compile-time and run-time performance across multiple AI hardware architectures and ML frameworks, such as TensorFlow and PyTorch, to support new state-of-the-art training and inference.
- Collaborate with ML researchers and engineers to guide compiler development for future ML trends.
- *Tools*: C++, Clang, MLIR, LLDB, cmake, ninja, gperf, tmux, neovim, ctags, clangd, Synopsys VCS.

Software Engineer — **AlpacaJapan, Tokyo** — **Jan 1, 2019 - Jan 8, 2021**
Machine Learning (ML) systems

- Design, develop, maintain, and test live production software systems for delivering stock price predictions.
- Collaborate with data science & engineering team to integrate different software systems and deploy and upgrade ML models in live production financial forecasting software.
- Handle installation & maintenance of new data sources and develop the data platform used for ML model R&D.
- Collect and document client requirements for future releases and make extensible and robust software design decisions for developing server & client web applications; responsible for 10% annual revenue growth.
- Manage software releases with an agile mindset and develop workflows for fast production recovery in case of failures.
- Driving innovation by evaluating new technologies, original financial data sources, and recent research papers that add value to Alpaca's products.
- *Tools*: Python, React, JavaScript, Flask, PostgreSQL, SQLAlchemy, Alembic, Kubernetes, Docker, PyTorch, Pandas, NumPy, SciPy, Luigi.

Data Scientist — **Anheuser-Busch InBev, Bangalore** — **Jun 19, 2017 - Dec 28, 2018**
Machine Learning (ML) pricing research

- Develop machine learning models to estimate ABInBev's market share and revenue in different pricing scenarios of beer SKUs across multiple countries, using both R statistical programming language and Python.
- Conduct extensive experiments to determine the significant variables in ML models and create automated scripts to replicate the process, using Keras, TensorFlow.
- Create pricing conjoint survey questionnaires and handle data management and pre-processing using customized scripts and workflows; used dplyr, tidyr.
- Interact and collaborate with business heads in different countries to include different pricing scenarios in conjoint based on the business requirements and present the pricing analysis results for business actions; used ggplot2.
- Develop various optimization algorithms based on pricing analysis results to maximize the business objective, such as market share or revenue; used nloptr.
- Create UI dashboards that display conjoint analysis results for business to gain actionable insights, using Shiny, RStudio.

Internships & Research

Graduate Research Assistant — **Georgia Tech, Atlanta** — **Aug 2021 - May 2022**

- Ported Facebook's QUIC implementation, mvfst, to rely on the efficient kernel-bypass network stack (threading & socket) APIs provided by MIT's Shenango and achieve low tail latency & increase CPU efficiency.
- *QuicNIC*: Offloaded GSO & crypto (encryption, decryption) to a dedicated CPU core to obtain record QUIC throughputs (x5). *Skills*: QUIC, caladan, mvfst, folly, fizz, profiling prod C++ codebase, CPU FlameGraphs.

Linux Contributor, Intern — **Google Summer of Code** — **May '21 - Aug '21**

- Analyze and fix race condition bugs in the Linux kernel 5.4 device drivers based on software verification static analysis tool, Klever. Part of *Linux Standards Base* & *Linux Driver Verification*.
- Accepted patches to kernel mainline. *Skills*: Linux kernel development, C.

Education

Atlanta, GA — **Georgia Institute of Technology** — **Jan '20 - May '22**

- **Master of Science in Computer Science** with *Systems Specialization*, May 2022. **GPA: 4.0**
- *Courses*: Operating Systems; Computer Architecture; Compilers; Networks; Distributed Systems; Databases; HPC; Algorithms.

Kolkata, IN **Indian Statistical Institute** **Jul '15 - Jun '17**

- **Master of Science in Quantitative Economics**; full scholarship & monthly stipends
- *Relevant Courses*: Optimization; Game Theory; ACM-ICPC regional qualifier.

Chennai, IN **Chennai Mathematical Institute** **Aug '12 - Apr '15**

- **Bachelor of Science in Mathematics and Computer Science**, Innovation in Science Pursuit Scholar
- *Courses*: Algorithms; Programming Languages; Discrete Math; Theory of Computation; Logic.

Projects

- *Xen's Credit Scheduler* Credit scheduler C implementation in a user-level threads library.
- *TinyFile* Shared memory-based file compression service with client library in C.
- *Distributed KVS* Distributed key-value store using XML-RPC with partitioning, replication & consistency.

Languages and Technologies

- **About 10k lines**: C++, Python, C, Java; **About 500 lines**: JS, HTML, CSS
- Linux, git, tmux, vim, ctags, LSPs (clangd), docker, k8s, AWS, Auth0, CircleCI, DataDog, ArgoCD