

433 Jarvis Street  
Toronto, ON M4Y 2G9  
Canada (Permanent Resident)

# SAUBHIK MUKHERJEE

AI COMPILER ENGINEER

USA: +1 (470) 313-0534  
CAN: +1 (226) 336-3865  
[saubhik@gatech.edu](mailto:saubhik@gatech.edu)

## EMPLOYMENT

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

- Built [MLIR](#) & [LLVM](#)-based compiler layers & tools to transform, optimize, debug, execute state-of-the-art AI models on proprietary (non-Von Neumann, dataflow-based) ML accelerator architectures; standardized compiler infrastructure to support for rapid new model bringup (10 per month) with industry-leading model performance
- Engineered and refined an embedded DSL abstracting memory & compute units, ensuring robustness and adaptability to evolving accelerator architectures and emerging ML models; designed optimizations to improve resource count by 5%
- Analyzed & improved compile-time & run-time performance across multiple AI hardware architectures & ML frameworks, such as [TensorFlow](#) & [PyTorch](#); achieved a 20% reduction in compilation time for large AI models, resulting in faster model iteration & development cycles
- Tools:* [C++](#), [Clang](#), [MLIR](#), [LLDB](#), [cmake](#), [Bazel](#), [ninja](#), [gperf](#), [tmux](#), [neovim](#), [ctags](#), [clangd](#), [Synopsys VCS](#)

**Software Engineer** **AlpacaJapan, Co. Ltd.** **Jan 1, 2019 - Jan 8, 2021**  
**Machine Learning (ML) systems** **Tokyo, Japan**

- Designed, developed, maintained, & tested production stock price prediction system, leading to 20% improvement in model deployment efficiency; developed workflows for fast production recovery reducing downtime by 30%
- Managed installation & maintenance of new OHLCV data sources & developed the data platform used for ML model R&D, resulting in a 25% reduction in data acquisition & processing time
- Collected & documented client requirements for future releases & made extensible & robust software design decisions for developing server & client web apps, contributing to a 10% annual revenue growth
- Tools:* [Python](#), [React](#), [JavaScript](#), [Flask](#), [PostgreSQL](#), [SQLAlchemy](#), [Alembic](#), [Kubernetes](#), [Docker](#), [PyTorch](#), [P&as](#), [NumPy](#), [SciPy](#), [Luigi](#), [CircleCI](#), [Argo CD](#), [Auth0](#), [Datadog](#)

**Data Scientist** **Anheuser-Busch InBev** **Jun 19, 2017 - Dec 28, 2018**  
**ML in pricing research** **Bangalore, India**

- Developed ML (multinomial logit, neural nets) models to estimate ABInBev's market share & revenue in different pricing scenarios of beer SKUs across multiple countries, using [R](#) & [Python](#); developed game-theoretic techniques & numerical optimization algorithms (Nelder-Mead, BFGS, Simulated Annealing) based on pricing analysis results to maximize business KPIs; created UI dashboards for consuming model reports, using [Shiny](#), [RStudio](#)

## INTERNSHIPS & RESEARCH

**Graduate Research Assistant** **Georgia Institute of Technology** **Aug 23, 2021 - May 7, 2022**  
**Networked Systems** **Atlanta, GA, USA**

- Developed [QuicNIC](#), a software NIC to accelerate the [QUIC](#) stack, to move segmentation, pacing, & encryption from the application to a dedicated core implementing software NIC functionality; achieved TCP-like single-connection throughput.
- Ported Meta's production QUIC (network protocol) implementation, [mvfst](#) (& dependencies [folly](#), [fizz](#)), to rely on the efficient kernel-bypass network stack (custom threading & socket libraries) APIs provided by MIT's [Shenango](#) ([caladan](#)) & achieve low tail latency & increase CPU efficiency; involved CPU profiling using [flame graphs](#); [research doc](#)

**Open Source Contributor** **Google Summer of Code** **Jun 7, 2021 - Aug 23, 2021**  
**The Linux Foundation**

- Analyze & fix race condition bugs in the Linux Kernel 5.4 device drivers based on software verification static analyzer checker, [Klever](#). Accepted [patches](#) to kernel mainline. *Skills:* [Linux kernel development](#), C.

## EDUCATION

- Master of Science in Computer Science** **Jan 2020 - May 2022**
  - [Georgia Institute of Technology, USA](#)
  - Major:* Computer Science, *Concentration:* Computing Systems. *GPA:* 4.00; [transcript link](#).
- Master of Science in Quantitative Economics** **Jul 2015 - Jun 2017**
  - [Indian Statistical Institute, India](#)
  - Passed in *First Division with Distinction*; [transcript link](#)
- Bachelor of Science in Mathematics & Computer Science** **Aug 2012 - Jul 2015**
  - [Chennai Mathematical Institute, India](#)
  - GPA:* 06.93 (out of 10); [transcript link](#).