Elton Pinto

eltonp3103@gmail.com | eltonpinto.me

Education

Ph.D in Computer Science, Georgia Institute of Technology, Atlanta, GA — August 2024 - present
M.S. in Computer Science, Georgia Institute of Technology, Atlanta, GA — August 2022 - May 2023
B.S. in Computer Science, Georgia Institute of Technology, Atlanta, GA — August 2018 - May 2022

Research Experience

Graduate Research Assistant *advised by Alexey Tumanov* — August 2024 - present

- **Co-author of Mirage**, an emulation-driven performance modeling system that transparently predicts the runtime of DLT workloads with less than 5% error without access to a GPU cluster. Developed a configuration search system that leverages Mirage to find Megatron-LM configs that reduce training costs by up to 56% . Submitted to EuroSys'25.
- **Developing AutoSched**, a novel compiler-driven DAG scheduling framework that dynamically instantiates a job scheduler based on the workload and system state, enabling cluster managers to dynamically navigate the scheduler runtime vs quality trade-off space.

Graduate Research Assistant *advised by Vivek Sarkar and Daan Leijen (MSR)* — August 2022 - May 2023

Replaced OCaml's garbage collector with the Perceus reference counting scheme, achieving upto 28% speed-up and 40% memory savings on representative programs. Presented at ICFP'23 ML Workshop.

Independent project, TINKER Lab — August 2022 - December 2022

Prototyped Neko, a quantum map-filter-reduce programming language that leverages superposition and interference for large-scale data processing. Presented at POPL'23 SRC. *Awarded NSF GRFP fellowship for this project.*

Work Experience

Software Engineer, Observe, OPAL compiler team — August 2023 - August 2024

- Developed a stratified sampling system to optimize RED chart queries, reducing the worst-case runtime from 2min to 30s for tables containing up to 25 billion records, modifying the compiler to accurately extrapolate sampled results
- Led the design, implementation, and public launch of reference tables, introducing non-temporal source table support to Observe's temporal stream processing system and bridging a key feature gap with competitors like Splunk and Datadog
- Enhanced OPAL with new primitives for data pivoting, bug fixes in SQL code generation of temporal primitives, and a Tree-sitter parser powering auto-complete, enabling faster and more accurate data analysis for customers

Software Engineering Intern, Meta, Privacy Language Experience (PLeX) team — May 2022 - August 2022

- Developed a distributed callgraph artifact generation system that feeds into a Hack typed-AST static analyzer to detect privacy-centric data leaks through global variables
- Built a pipeline that ingests over 100M records of Hack callgraph data into Glean, enabling large-scale program analysis
- Optimized Glean queries using derived predicates, resulting in a 280x speedup, and incrementally ported the system from Python to Rust, leveraging data-level parallelism for a 4.5x speedup.

Software Engineering Intern, Meta, PyTorch Dev Infra team — May 2021 - August 2021

- Setup infrastructure to build, test, and deploy a fork of clang-tidy in PyTorch CI using Docker and GitHub Actions
- Added support for the `max-tokens` pragma in clang-tidy guarding against commits that increase compilation times
- Authored a clang-tidy check that detects infinite loops caused by integer/floating-point overflow

Select Publications

[1] Elton Pinto and Daan Leijen. "Exploring Perceus for OCaml". In: *ML Family Workshop*. ICFP'23.

[2] Elton Pinto. "Neko: A quantum map-filter-reduce programming language". In: *Student Research Competition (SRC)*. POPL '23.

[3] Elton Pinto, Jeffrey Young, Thomas Conte, Austin Adams, and Eugene Dumitrescu. "An Implementation of the Quantum Verification of Matrix Products Algorithm". In: *Quantum Resource Estimation (QRE) Workshop*. ISCA '22.