# Nguyen Nguyen

Boston, MA | 315-708-3375 | phucnguyen1999.nn@gmail.com | linkedin.com/in/nguyenston |
https://github.com/nguyenston | https://nguyenston.github.io

Software engineer focused on backend systems and data-intensive applications, with research training in optimization and machine learning and experience building end-to-end systems through self-directed and academic projects.

## PROJECTS

**Mokuro Library** 11/2025 - Present

- Built a self-hosted, multi-user Japanese OCR reader and editor for NAS/home servers, with authentication, binary file streaming, and reproducible Docker Compose deployment.
- Designed and implemented a database-backed versioned data model for OCR documents using immutable patches and copy-on-write branching, supporting undo/redo, snapshot caching, and deterministic state reconstruction via patch replay.
- Implemented collaborative editing infrastructure for shared OCR content, including an admin-maintained baseline, per-user private branches, undo/redo, reset, and a rebase-based merge workflow with explicit conflict resolution.
- Built a browser-based OCR editing engine for text and bounding boxes, implementing zoom-invariant geometry transforms, keyboard-driven text mutation, and Japanese vertical and horizontal layout semantics.
- Designed a hybrid shared-library system combining public admin-owned content with private user uploads, per-user reading state, and a submission/review workflow for promoting contributions.

**Quantum Virtual Machine (QViM)** 03/2021 - 05/2021

- Created a Domain-Specific Language in Julia using meta-programming to define quantum circuits with readable syntax.
- Implemented a circuit simulator that optimizes gate operations via linear algebra basis changes to reduce computational overhead.

## WORK EXPERIENCE

**Paschalidis NOC Lab** **05/2022 - Present**
**Graduate research assistant | Full-time** **Boston, MA**

- Formulated a novel spectral algorithm for discovering latent policies from demonstrations, guaranteeing global convergence in a single data pass and overcoming EM method pitfalls.
- Designed and implemented a flexible framework for discovering latent processes, with applications in bioinformatics (mutation analysis) and remote sensing (hyperspectral unmixing).
- Built and deployed an ML pipeline in 6 months, trained on a 100,000-patient dataset and is now running weekly inference for BMC's clinical trial for hypertension prescription.
- Built a data pipeline for a complex 10-year dataset of 30,000 appointments for 6,000 patients to power ML prediction of missed CT screenings, enabling targeted patient support.

## SKILLS

**Programming languages:** Python, TypeScript, Bash, Rust, Julia, Lua, SQL, C++

**Technical skills:** large-scale data processing, exploratory data analysis, data integration, statistics, probability, optimization, machine learning, dynamic programming, predictive modeling, data structures

**Framework and libraries:** PyTorch, Hugging Face, pandas, Weights & Biases, cvxpy, pydantic, scikit-learn, Jupyter, SvelteKit, Prisma, Fastify

**Tooling:** Cargo, Conda, Linux/Unix, Nix, Git, Latex, Docker

**Languages:** English (fluent), Vietnamese (native), Japanese (conversational)

## EDUCATION

**PhD. in Systems Engineering** | Boston University Boston, MA | 08/2021 - 06/2026

**B.S. in Computer Science** | Syracuse University Syracuse, NY | 08/2017 - 05/2021

**B.A. in Physics** | Syracuse University Syracuse, NY | 08/2017 - 05/2021