

# Nguyen Nguyen

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

Ph.D. candidate with 3 years of experience developing interpretable machine learning models for data-driven modeling. Experience includes designing novel, efficient algorithms and building automated ML pipelines for large-scale data applications.

## WORK EXPERIENCE

### Paschalidis NOC Lab

#### Graduate research assistant - Full-time

05/2022 - Present

Boston, MA

- Implemented and tested a novel filter for computer vision tasks; simulation shows improved vehicle speed estimation accuracy by ~26% under heavy noise (30% corrupted frames).
- 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, Bash, Julia, R, Rust, SQL, Lua, C, C++

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

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

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

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

## PROJECTS

### Mokuro Library

11/2025 - Present

Full-stack, Dockerized server application (Fastify/Prisma backend, SvelteKit frontend) that hosts a centralized, multi-user comic library with embedded OCR data. It exposes a comprehensive REST API for user-scoped resource management (auth, progress, secure file serving) and features an in-place web reader for editing OCR text and bounding boxes, with all changes written directly back to the source files on the server's disk.

### Custom LoRA for SDXL

01/2025 - 03/2025

Fine-tuned the SDXL model by training LoRA to generate images with non-standard anatomical compositions. Curated and preprocessed a custom dataset and experimented with various training parameters in the Hugging Face ecosystem to achieve desired model outputs.

### Matter phase simulation

01/2021 - 05/2021

Developed a particle simulation in Rust with a Bevy UI to explore complex physical systems, implementing an efficient grid method to visualize the breakdown of the ideal gas law, crystalline formation, and the process of annealing.

### Quantum Virtual Machine (QViM)

03/2021 - 05/2021

Designed a domain-specific language (DSL) in Julia using meta-programming to create an intuitive syntax for defining quantum logic circuits. Includes an efficient simulator, with gate operations optimized through the use of basis change.

## 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

