John Pluchino

New York, NY · john.pluchino7@gmail.com · github.com/johnpl-io

Education

The Cooper Union — New York, NY
B.E. Electrical Engineering, Computer Engineering Track; Minor in Computer Science, GPA: 3.91/4.00 — 2020 - 2024

Experience

Persistent Systems, LLC — New York, NY
Software Engineer II — Mar 2026 - Present
Software Engineer — Sep 2024 - Mar 2026
Software Engineer Intern — Jun 2024 - Sep 2024

- Lead development of a FastAPI service allocating MAC addresses across multiple product lines; implemented idempotent, race-safe allocation using row locking and unique-constraint conflict recovery, keyed on device type + serial number.
- Cut query latency 82% (5.1s to 0.9s at 400k rows) by using execution plans to trace a bottleneck to a redundant forced sort in a SQL Server view; designed a 12k–400k-row scaling benchmark to validate the fix.
- Built the team's first auth/authz layer (scoped API principals with per-device-type permissions); own the containerized (Docker Compose) deployment.
- Prototyped an observability pipeline for the MAC service: structured JSON logs shipped via Vector to Loki, surfaced in Grafana dashboards.
- Develop and extend a Java desktop application that prepares Android-based MANET radios for production, automating firmware installation, configuration via radio APIs, and network validation (TCP, video streaming), and persisting unit data to a database; added support for a new device product line used by production teams and external customers.
- Refactored legacy multi-threaded code to unblock new feature support, replacing polling with condition-variable-based blocking synchronization; built a companion Kotlin Gradle plugin that scans the codebase for user-facing error codes and generates HTML documentation embedded in the Java application, letting production teams and external customers self-diagnose issues.

Newrez — New York, NY
Software Engineer Intern — Jun 2023 - Jul 2023

- Delivered bug fixes and new features in C++ and SQL on a legacy proprietary mortgage application.
- Built a proof-of-concept user story formatter with Transformer-based grammar checking and Levenshtein-distance scoring; demoed to company executives.

Carnegie Mellon University — Pittsburgh, PA
REU Software Engineering Research Intern — May 2022 - Aug 2022

- Conducted dynamic and static analysis of WASM benchmarks across four production WebAssembly engines (V8, SpiderMonkey, JavaScriptCore, and Wasmtime), advised by Dr. Heather Miller and Dr. Ben Titzer.
- Contributed to WebAssembly benchmark suite research, performing statistical analysis (PCA) in Python to identify redundant and underrepresented workload characteristics and inform suite diversification.

Projects

C99 Compiler *C, GNU Bison, Flex* · github.com/johnpl-io/C99-Compiler

- Designed and implemented a compiler in C for a large subset of C99, with a Flex/Bison front end and semantic analysis over the resulting AST.
- Built a backend lowering the AST to a three-address-code (quad) IR organized into a basic-block control-flow graph, emitting x86-32 assembly; compiles and runs test programs (recursion, pointer arithmetic, N-dimensional arrays).

Generative Art Recommendation System (Cooper Union Capstone) *Python, TensorFlow, PyTorch, SDXL Lightning*
Honorable Mention, HackAI (Dell & NVIDIA Challenge, 4000+ participants) · github.com/johnpl-io/gars_nvidia

- Designed a content-based recommendation system driven by user engagement with images generated in real time, using an adaptive algorithm with a decaying exploration rate over a Milvus vector similarity search pipeline; achieved a 95% positive rating rate vs. 55% random baseline on explicit user ratings in IRB-approved user studies.
- Trained emotion and focus classification models and fused their outputs into a unified engagement score computed from a live webcam feed.
- Engineered real-time image generation with SDXL Lightning featuring progressive decoding of latents and CPU offloading for consumer GPU deployment.

Smart Park *PostGIS, PostgreSQL, Python, React.js*

- Designed and built a full-stack app that recommends free NYC street parking from logged user parking events, using PostGIS geospatial queries by time and proximity to the user's location.

Skills

Languages:	Python, Java, C++, C, Kotlin, SQL
Technologies:	FastAPI, SQLAlchemy, PostgreSQL, SQL Server, PostGIS, Docker, Git, Linux, Grafana, Loki, Spring Boot, Gradle, Jenkins, JUnit, React.js, TensorFlow, PyTorch, Milvus