

Alexander Kozik

215-264-2104 | alex.kozik3141@gmail.com | linkedin.com/in/alex-kozik | alexkozik.com | github.com/LambdaAK

Education

- Carnegie Mellon University** — *M.S. in Computer Science* Aug 2026 – Dec 2027
- Concentration in Machine Learning and Systems
- Cornell University** — *B.A. in Computer Science & Mathematics* Aug 2022 – Dec 2025
- **GPA: 4.11/4.3 (Top 1%)** | Concentration in Machine Learning and Software Engineering
 - **Awards:** Marx Memorial Prize (2 recipients/yr), Phi Beta Kappa, Summa Cum Laude, Outstanding TA
 - **Teaching Assistant:** Machine Learning, Analysis of Algorithms, Data Structures & Functional Programming

Experience

- Amazon** — *Software Development Engineer Intern, AI for Customer Service* June 2026 – Aug 2026
- Architected an **ML-gated damage-claim verification feature** to block fraud and reduce manual review.
 - Engineered a secure image-upload pipeline (**S3** + virus scanning) passing scanned images to a **computer-vision service** that synchronously verifies reported damage and gates approval.
 - Built the **Java/Kotlin** backend orchestrating the multi-step upload-to-decision sequence across services.
- Salesforce** — *Software Engineer Intern, Metadata-Driven Services* June 2025 – Aug 2025
- Built **virtual formula fields** in Salesforce Data Cloud for no-code ML features at enterprise scale.
 - Implemented **query-time computation** over data objects, eliminating staleness and storage overhead.
 - Engineered the **Java/Apache Spark** layer for formula evaluation over cross-object joins and aggregations.
- Cornell Course Management System** — *Software Engineering Lead* Aug 2023 – Jan 2026
- Led a team of **20 engineers** building Cornell's course management system, used by **10,000+ students**.
 - Architected the **React/TypeScript** student-facing frontend, reducing page load times by **90%**.
 - Implemented APIs with **serializable data structures** for secure frontend/backend communication.
- Generative AI @ Cornell** — *Machine Learning Engineer* Feb 2025 – Dec 2025
- Developed **AI agent** for QuickFi, a commercial lending technology company, automating insurance certificate validation and compliance verification processes, replacing a manual review step.
 - Built frontend and **vector database** for ClassGPT, enabling professors to create AI tutors for classes.
- Carnegie Mellon University** — *Software Engineer & Research Intern* May 2024 – Aug 2024
- Worked on the **interpreter** for SASyLF, a language for writing proofs, implementing core features in Java.
 - Implemented a **polymorphic type system** for SASyLF, similar to Java generics, enabling type-safe code reuse and significantly reducing duplication in proofs by allowing abstraction over multiple data types.

Selected Projects

- C-Torch** | *Distributed Machine Learning Library* | C++, CUDA, Multithreading, TCP Networking
- Implemented **15+ ML models** in C++ (neural networks, RL), achieving 99%+ accuracy and 97%+ win rates.
 - Built distributed training across **8 nodes** using TCP AllReduce and dataset sharding, achieving **5× speedup**.
 - Achieved **26× speedup** via **CUDA kernels**, compiler optimizations, and parallelization of tensor operations.
- Atlas** | *End-to-End LLM Infrastructure* | Python, PyTorch, FastAPI, Redis, PostgreSQL, Docker, Kubernetes, React
- Built a **50M-param GPT-style transformer** from scratch for conversational reasoning and state tracking.
 - Engineered a distributed inference system with Redis + batching, with **110 req/s** at **140 ms P95 latency**.
- Forge** | *Functional Programming Language* | OCaml, LLVM
- Implemented an **LLVM-based compiler** that lowers to native code, yielding a 2.5× speedup over Python.
 - Created a tree-walking interpreter and constraint-based **type checker** for 20+ language features.

Technical Skills

Languages: Python, C, C++, Rust, TypeScript, JavaScript, Java, OCaml, SQL
Libraries: PyTorch, CUDA, Scikit-learn, NumPy, Pandas, React, Next.js, Node.js, Express.js, Flask
Tools: Linux, Docker, Kubernetes, AWS, Redis, S3, PostgreSQL, gRPC, Apache Spark, TCP Sockets, LLVM, Git