

OSHIN REX

omr6@cornell.edu | linkedin.com/in/oshin-rex | github.com/oshinrex

EDUCATION

Cornell University

Ithaca, NY

B.S. Computer Science and Operations Research (GPA: 3.79/4.0)

Aug. 2024 – May 2028

- **Relevant Coursework:** Analysis of Algorithms, Object-Oriented Programming, Functional Programming, Linear Algebra, Probability, Optimization, Databases, Machine Learning

EXPERIENCE

Software Engineering Intern

May – July 2026

IBM Consulting

Chicago, IL

- Architected a vendor management portal from scratch serving 60,000+ consumers and 6,000+ vendors using JavaScript, HTML/CSS, Dataverse, and Liquid templates, enabling vendors to view their assigned purchase-of-service requests.
- Developed and integrated FetchXML queries across multiple Dataverse entities to retrieve, filter, and relate vendor, purchase-of-service, and associated request records, powering dynamic data retrieval throughout the application.
- Developed a Microsoft Copilot-based intake workflow that analyzed unstructured service requests, extracted key requirements, and recommended the vendors best suited to each request.
- Won 1st place overall out of 30+ teams in the IBM Consulting Intern Challenge by engineering a full-stack fan intelligence platform and technical roadmap for the Seattle SuperSonics, integrating Python-based vector embeddings, propensity scoring, and real-time streaming chat APIs to analyze fan behavior and support personalized engagement strategies.

Software Developer

Sep. 2025 – Present

Hack4Impact Cornell

Ithaca, NY

- Built and deployed a React Native application for Alaska Children's Trust, digitizing 75 safety cards across 15 decks with offline persistence, child profiles, deck filtering, and progress tracking.
- Built and deployed a full-stack inventory platform for the Hudson Valley Textile Project using React, Express.js, and Firebase, replacing spreadsheet-based tracking with centralized inventory, sales, and location data.
- Implemented role-based access and public inventory views, separating internal warehouse operations from consumer-facing stock data while supporting inventory search, filtering, and updates.

Software Developer Intern

Dec. 2025 – Feb. 2026

GreenSand AI

Remote

- Engineered an automated Django pipeline to process 5,000+ PDFs, parsing documents into structured tokens and indexing word coordinates for interactive highlighting to replace manual OCR review.
- Reduced end-to-end document processing time by 30% and improved text extraction quality by integrating MinerU into the parsing architecture to normalize raw OCR outputs.

PROJECTS

High-Performance Limit Order Book | C++20, STL, CMake

2026

- Processed 1M+ synthetic orders with deterministic FIFO execution and 429 ns median matching latency by implementing a price-time priority engine supporting market orders, limit orders, partial fills, and cancellations.
- Reduced order cancellation latency to 1.38 μ s, achieving 99.4% improvement over a naive linear-scan baseline by indexing orders with hash maps and organizing price levels with ordered data structures.
- Achieved 1.14M orders/sec throughput at sub-3 μ s p99 latency (matching: 1.9 μ s, cancellation: 3.0 μ s) across 1M+ randomized orders by developing a C++ benchmarking harness to measure matching, insertion, and cancellation performance under sustained, high-throughput order flow.
- Reduced memory usage by 12.5% and allocations by 26.9% by profiling execution paths and optimizing object lifetimes, memory access patterns, and cache locality across the matching engine.

Case Match (Agentic Memory) | Python, LangGraph, CockroachDB, AWS Bedrock

2026

- Evaluated semantic retrieval across 48+ consulting engagements using precision and recall at 5 results, benchmarking embedding-based search against service-line similarity as a relevance proxy across representative consulting queries.
- Built a self-correcting LangGraph retrieval workflow using LLM-based relevance evaluation across business problems, solution approaches, and outcomes, expanding from 5 to 8 candidates across industries when initial results are insufficient.
- Achieved 175ms median vector-search latency and 330ms end-to-end latency across 48 benchmark queries using 1,024-dimensional AWS Titan embeddings stored and retrieved through CockroachDB's vector search infrastructure.
- Overrode top vector-similarity match in 52% of cases across 48-query blind evaluations, using AWS Bedrock and Pydantic to evaluate business problem, solution, outcome, and cross-industry transferability when ranking consulting precedents.

TECHNICAL SKILLS

Languages: Python, C++, Java, SQL, JavaScript, HTML/CSS, OCaml

Frameworks/Libraries: React, React Native, Express.js, Django, FastAPI, LangGraph, OpenCV

Cloud/AI: AWS Bedrock, Microsoft Copilot, Power Automate, RAG

Databases/Tools: CockroachDB, PostgreSQL, Firebase, Microsoft Dataverse, Git, CMake