# Wonjin Eum

(831) 884-2492 | we46@cornell.edu | LinkedIn | GitHub | Website | Monterey, CA

## Education

**Cornell University, College of Engineering** Ithaca, NY

*Bachelor of Science in Computer Science* *Expected May 2028*

**Relevant Coursework**: Object-Oriented Programming and Data Structures (Java), Machine Learning, Analysis of Algorithms, Functional Programming (OCaml), Discrete Math, Backend Development, Networks

**Campus Involvement**: Cornell AppDev, Cornell Data Science, Women in Computing at Cornell

## Experiences

**LinkedIn** Jan 2026 – Present

*Software Engineering Intern* *Remote*

- Architected a multi-stage inference platform combining a custom-trained CNN with an LLM-based scoring service, processing 8,000+ product images monthly into structured material, care, and sustainability attributes.
- Designed a retrieval-grounded generation system constraining LLM outputs to a structured sustainability knowledge base (material impact, supply chain signals, brand ratings), reducing hallucinated ratings 15% in production.
- Optimized the CNN inference layer via request batching and loss-function tuning, cutting end-to-end scoring latency 50% and lifting eco-alternative click-through 40%.
- Built a versioned schema-contract system (Pydantic, TypeScript) spanning the model-serving, API, and browser extension layers, eliminating a class of cross-service integration defects and reducing schema-related failures 46%.

**Curate** Jun 2026 – Present

*Software Engineering Intern* *Los Angeles, CA*

- Architected an agentic retrieval-augmented system (FastAPI, Pinecone, Voyage AI embeddings) integrating tool-calling, ticket escalation, and live order lookups across 50+ restaurant clients.
- Designed validated tool-calling handlers for order lookup and escalation logic, reducing manual triage time 42% and deflecting 18 recurring FAQ tickets/month.
- Built a hybrid retrieval system combining PostgreSQL full-text search with dense Voyage embeddings over a 28-article knowledge base, tuning chunk size and top-$k$ to raise grounded-answer precision from 68% to 84%.
- Engineered a configuration-parsing pipeline extracting Toast POS data to auto-generate loyalty program mappings, cutting client onboarding time from 2.5 hours to 12 minutes.

## Projects

**Eatery** | *Node.js, TypeScript, Prisma, PostgreSQL, Docker* GitHub

- Designed a ratings subsystem with a Prisma/PostgreSQL schema and recency-weighted aggregation algorithm, exposing 4 versioned REST endpoints serving 10,000+ monthly active users.
- Containerized the backend infrastructure with Docker, standardizing developer environments across the team and cutting local setup time from 2 hours to 30 minutes.
- Migrated a legacy backend service onto a containerized Node.js/TypeScript stack, redesigning the deployment pipeline to cut release time 60% and eliminate 3 categories of environment-specific failures.
- Led a 13-person cross-functional team across iOS, Android, backend, and design, coordinating technical planning, code reviews, Git workflows, and feature delivery for an application with 40,000+ downloads.

**Learning Based Cache Eviction** | *Rust, Python, PyTorch* GitHub

- Designed a Rust-based caching system with a learned eviction policy using PyTorch MLP predicting future access patterns from real-world (Wikipedia CDN cache server) and synthetic trace data (Looping, Bursty, Zipf).
- Built an online retraining pipeline updating the eviction model every 5,000 misses, achieving a 27% hit-rate improvement over LRU on Zipf-distributed workloads without full model retraining.
- Quantified latency-vs-hit-rate tradeoffs against traditional heuristics (LRU, FIFO), isolating a 0.18 ms per-decision inference overhead against the 27% hit-rate gain.

## Technical Skills

**Languages**: Python, C/C++, SQL (PostgreSQL, MySQL), Java, JavaScript/TypeScript, OCaml, Rust
**Frameworks & Libraries**: React & Next.js, Express & Node.js, NumPy, Pandas, PyTorch
**Developer Tools**: Git, Docker, MongoDB, Vercel, AWS, Postman