

DongYeop Lee

Software Engineer · Full-Stack & Backend Systems · Java, Python, Node.js, React, SQL, REST APIs

dongyeop0810@gmail.com · (619) 870-6702 · linkedin.com/in/dongyeopl · github.com/DY0810

EDUCATION

University of Southern California

Bachelor of Science, Physics/Computer Science

Los Angeles, CA

Expected December 2028

- GPA 4.0/4.0 · Dean's List
- USACO Platinum Division (2024)

PROFESSIONAL EXPERIENCE

Hemut (YC P25)

Fresno, CA

Software Engineer, Backend & Infrastructure (AI Engineer)

April 2026 - Present

- Restored data integrity across a 9-org PostgreSQL estate by moving every database tool onto row-level-security-enforced, org-scoped sessions, closing a cross-tenant data leak, then optimized the query plans the rollout regressed to cut the worst endpoint from 10.2s to 487ms (21x).
- Owned the ECS Fargate release path end to end (SHA-pinned container images, per-deploy task definitions, pre-rollout Alembic migrations) behind GitHub Actions CI/CD, taking services from code and tests through to production deploys.
- Triaged a production latency issue and removed 570ms of time-to-first-token (1.3s to 730ms) through request parallelization, HTTP connection pooling, endpointing-threshold tuning, and preemptive generation.
- Raised end-to-end task accuracy from 33% to 99% on an internal AI copilot by rebuilding its agent layer: a 192-tool MCP server with a ReAct loop and embedding-based tool retrieval, eliminating fabricated tool output.

IDX Exchange

Los Angeles, CA

NLP Engineer & Team Lead

February 2026 - May 2026

- Engineered a multistage NLP pipeline over 10,000 listings (8 classes, 103 regex patterns, hybrid FAISS + BM25 retriever), backed by automated tests at every stage.
- Raised classification F1 40%, enabling automated structured search and removing manual listing review for analysts.

PROJECTS

EcoPrompt: AI Energy Efficiency · Next.js, React, TypeScript, AWS Bedrock, DynamoDB

- Built a full-stack AI chat application on AWS Bedrock (Claude Haiku/Sonnet, Titan embeddings) with semantic-deduplication caching, prompt-complexity-based model right-sizing, and a live energy dashboard.
- Cut per-query LLM energy up to 98.6% on cache hits and 90% on right-sized queries, quantified against EPA eGRID emissions factors.

Veracity: AI Fact-Checking Chrome Extension · Chrome Web Store

- Shipped a zero-backend, BYOK Manifest V3 extension (React 19, TypeScript) that extracts and verifies per-article claims against live sources, with exponential-backoff retries and a 4-level JSON repair pipeline for malformed model output.

TECHNICAL SKILLS

- **Languages:** Java, Python, JavaScript (Node.js), TypeScript, SQL, C++, Bash
- **Backend & Data:** REST APIs, PostgreSQL (RLS, query planning), Node.js/Express, FastAPI, Redis, DynamoDB, SQLAlchemy, Alembic, Pydantic, pandas
- **Frontend:** React, Next.js, Tailwind, Zustand, Zod
- **Infrastructure & DevOps:** AWS (ECS/Fargate, ECR, RDS, S3, Secrets Manager), AWS CDK, Terraform, Docker, Kubernetes, GitHub Actions, CI/CD, Linux, Git
- **Testing:** pytest, Jest, Vitest, automated integration testing
- **AI & LLM:** LLM agents (ReAct, tool use, MCP), RAG and hybrid retrieval (FAISS, BM25, embeddings), evals and error analysis