Howard (Hao) Ye

Email: hyedailyuse@gmail.com — Phone: (865) 456-1194 — Boston, MA

howardye.vercel.app | GitHub | LinkedIn

SUMMARY

Computer Engineering graduate with production experience in cloud-native architectures, AI/ML integration, and DevOps automation. Specialized in building scalable microservices platforms with Docker, Kubernetes, and AWS deployment. Proven track record deploying full-stack applications with LangChain AI agents, Spring Boot services, and real-time monitoring systems. Strong foundation in system modeling, distributed computing, and infrastructure automation.

TECHNICAL SKILLS

Cloud & DevOps: AWS (EC2, RDS, S3, EKS), Docker, Docker Compose, Kubernetes, Terraform (IaC), CI/CD, Prometheus, Grafana

Backend & APIs: Node.js/TypeScript, Python (FastAPI), Java (Spring Boot), Go, RESTful APIs, GraphQL, WebSocket, gRPC

Frontend: Vue.js, React/Next.js, TypeScript, Real-time Dashboards, Chart.js, Responsive Design

Databases & Caching: PostgreSQL, Redis, ChromaDB (Vector DB), Query Optimization, Connection Pooling

AI/ML Integration: LangChain, OpenAI API, RAG Systems, PyTorch, Hugging Face, Reinforcement Learning (PPO)

System Modeling: OMNeT++, HPC Simulation, Performance Optimization, IBM CPLEX, Mathematical Optimization

Tools & Monitoring: Git, JWT Auth, Prometheus, Grafana, Health Monitoring, Structured Logging, API Testing

EDUCATION

MS Computer Engineering — University of Tennessee Knoxville — 2025
MS Industrial Engineering — University of Tennessee Knoxville — 2016

PROJECTS

SmartOps AI Agent Platform

GitHub: github.com/yehao622/hpc-simulation-platform

Tech: Docker, Node.js/TypeScript, Python FastAPI, Java Spring Boot, PostgreSQL, Redis, ChromaDB, LangChain, AWS, Prometheus, Grafana

- Architected enterprise-grade microservices platform with 8 containerized services: API Gateway (Node.js), AI Agent (Python/LangChain), Analytics (Java Spring Boot), React frontend, and supporting infrastructure
- Deployed production-ready system on AWS EC2 with security groups, health checks, auto-restart policies, and memory optimization for free-tier hosting
- Integrated LangChain AI agent with RAG (Retrieval Augmented Generation) system using ChromaDB vector database for intelligent HPC system analysis and natural language querying
- Built RESTful and GraphQL APIs with JWT authentication, Redis caching, PostgreSQL optimization, and WebSocket real-time communication
- Implemented comprehensive observability stack with Prometheus metrics collection and Grafana dashboards for cross-service monitoring
- Developed automated testing suite covering 15+ API endpoints with CI/CD-ready validation workflows

Smart Home Energy Management System

GitHub: github.com/yehao622/SmartHomeSimulator — **Demo:** smart-home-energy-demo.vercel.app

Tech: Docker, Vue.js/TypeScript, Node.js, Python, Go, Socket.IO, Kubernetes, Terraform, Reinforcement Learning

- Designed event-driven microservices architecture with 4 containerized services processing real-time IoT data from 11+ smart home devices (HVAC, EV charger, solar panels, battery storage)

- Deployed multi-environment infrastructure: Docker Compose for local development, Kubernetes manifests for orchestration, and Terraform templates for AWS EKS provisioning
- Built full-stack application with Vue.js frontend featuring real-time Chart.js visualizations, Node.js backend with Socket.IO event streaming, and Python microservice for reinforcement learning optimization
- Implemented RESTful device management API in Go with extensible service architecture for multi-home scalability
- Integrated PPO (Proximal Policy Optimization) reinforcement learning agent for intelligent energy scheduling based on time-of-use pricing and renewable energy availability
- Developed thermal modeling system for HVAC and water heater with physics-based simulations and CSV data export capabilities

Mathematical Computing Optimization System

GitHub: Entropy Inequalities — **Tech:** C, IBM CPLEX

- Integrated IBM CPLEX enterprise solver with C applications for large-scale constraint optimization problems with 1000+ variables and complex mathematical constraints
- Developed computational framework with memory-efficient algorithms for information theory applications
- Built automated validation systems for verifying solution correctness and optimality conditions

Applied AI/ML Techniques Across Kaggle Challenges

Tech: PyTorch, Hugging Face Transformers, XGBoost, LightGBM, CatBoost, Optuna, scikit-learn

- **LLM Response Classification:** Fine-tuned DeBERTa-v3-base transformer model for multi-class classification of LLM outputs, implemented custom training pipeline with TRL library, achieved weighted F1-score optimization through hyperparameter tuning
- **Risk Prediction System:** Built ensemble regressor combining gradient boosting models with advanced feature engineering including polynomial features and domain-specific risk multipliers, automated hyperparameter optimization using Optuna with 100+ trials
- **Bioinformatics Classification:** Developed multi-label protein function prediction system processing amino acid sequences, implemented k-mer feature extraction and Gene Ontology (GO) term hierarchy propagation for biological pathway classification

ACADEMIC EXPERIENCE

Computer Engineering MS Program — University of Tennessee — 2023-2025

- **Capstone Project:** Developed cloud-native smart home energy management system demonstrating microservices architecture, Kubernetes orchestration, infrastructure-as-code with Terraform, and multi-cloud deployment strategies
- Applied DevOps best practices: Docker containerization, CI/CD pipelines, automated testing, health monitoring, and production deployment on AWS with security hardening

Computer Engineering PhD Research — University of Tennessee — 2018-2023

- **Deep Learning & HPC Storage:** Researched Lustre distributed file system performance optimization using machine learning, developing automated benchmarking frameworks and predictive modeling pipelines
- **Distributed Systems:** Designed performance modeling frameworks for high-performance computing environments with focus on I/O optimization and system bottleneck analysis

Industrial Engineering MS Program — University of Tennessee — 2014-2016

- **Mathematical Optimization:** Specialized in large-scale linear and nonlinear optimization problems, integrating enterprise solvers (CPLEX, Gurobi) with custom C/C++ applications for industrial applications