

VINAYAK DHEMARE

Full Stack DevOps Engineer

+91 9403196949 • vinayak.supreme1@gmail.com • [linkedin.com](#) • [portfolio](#) • PUNE, India

Summary

Full-Stack & DevOps Engineer focused on building scalable, reliable, and production-ready systems. Experienced in modern web development, cloud infrastructure, CI/CD, and containerization, with an engineering mindset toward performance, maintainability, and practical problem-solving.

Skills

GitHub • React • Next.js • Node.js • Express • MongoDB • PostgreSQL • Prisma • Drizzle • REST APIs • WebSockets • Redis • Kafka • RabbitMQ • Microservices • Linux • GitHub Actions • Docker • Kubernetes • Terraform • AWS • EC2 • EKS • S3 • CloudFront • GCP • Prometheus • Grafana • ArgoCD • CI/CD • Automation • Container Orchestration • Cloud Infrastructure • System Design • Cloud Architecture • CI/CD Pipelines • Scalability • Performance Optimization • Leadership

Experience

Full Stack & DevOps Engineer	INDIA
Freelance	11/2022 - Present
- Designed and deployed scalable full-stack applications across education, productivity, and SaaS. - Built production applications using React, Node.js, Express, MongoDB, and PostgreSQL. - Automated testing and deployments using GitHub Actions, Docker, and CI/CD. - Managed cloud infrastructure on AWS/GCP, including Kubernetes-based deployments and monitoring.	
ORGANIZER	Pune, India
CORE TEAM	02/2026 - Present
- Organized AWS Builder Group events and technical workshops on campus. - Collaborated with technical communities to drive student participation and developer engagement. - Led outreach, event promotion, and on-ground coordination for technical initiatives.	

Projects

AI-Powered Chromium Browser	Personal Project
	01/2026 - Present
Built a Chromium-based AI browser with on-device LLM inference , enabling private webpage summarization, contextual Q&A, and AI-assisted browsing without transmitting webpage content to external AI APIs	
Integrated local LLM inference with Chromium to deliver privacy-focused, context-aware browsing features. Tech: Chromium, C++, JavaScript/TypeScript, Local LLM, AI/ML	
GPT-2 124M – Transformer Language Model from Scratch	Personal Project
	09/2026 - Present
Built a 124M-parameter GPT-2–style LLM from scratch in C++ , implementing the Transformer architecture and training infrastructure.	
- Recreated the core GPT-2 124M architecture without relying on a pre-trained GPT-2 model. - Hardcoded the complete Transformer architecture, including Multi-Head Self-Attention, MLP, LayerNorm, residual connections, and causal masking. https://github.com/ouroboros1X/GPT2_INFERENCE-ENGINE	

Key Achievements

 Finalist – Google Cloud Agentic AI Day Hackathon (2025) Top team among 200+ participating teams Finalist among 200+ teams for developing an AI-driven solution using Google Cloud's Agentic AI framework.	 Technical Mentor & Community Leader Guided peers in AI, Web Development, and DevOps best practices; organized tech events and knowledge-sharing sessions.	 Delegate – Model United Nations Represented an assigned country in MUN debates, researching global issues and negotiating resolutions with diverse delegates.
---	---	---