# Vageesh Dwivedi

AI and ML Engineering Leader | +1-(408) 679-6025 | vageesh.dwivedi@gmail.com

## SUMMARY

Results-driven engineering leader with 20+ years of experience in AI/ML, Conversational AI, Cloud Infrastructure, and Enterprise Software Development. Proven track record of building and scaling teams, driving AI innovation, and optimizing cloud costs. Strong expertise in LLM-powered solutions, RAG-based Q&A systems, and Virtual Assistant platforms. Adept at leading cross-functional teams to deliver high-impact AI solutions at scale.

## CORE SKILLS

- **AI/ML & Generative AI:** RAG-based Q&A, LangChain, AutoGen, LlamaIndex, LLM Inference (vLLM, LiteLLM)
- **Conversational AI & NLP:** Virtual Assistants, Intent Classification, Embedding Models
- **Software Development:** JavaScript, Node.js, React, Redux, Kubernetes, CI/CD Pipelines
- **Cloud & MLOps:** Kubernetes, MLFlow, Prompt Management, Cloud Cost Optimization (GCP, Azure)
- **Enterprise Engineering:** Business Messaging, UI Frameworks, Accessibility, Security & Compliance

## EXPERIENCE HIGHLIGHTS

### Director, Software Development | [24]7.ai | Campbell, CA | Nov 2015 – Present

- **Led a 24-engineer global team** across three geographies to develop **enterprise-scale AI-powered Conversational AI solutions**, serving millions of users.
- **Drove AI/ML innovation**, implementing **RAG-based Q&A** and **Microsoft Autogen-powered agentic flows**, reducing onboarding time by **30%** and AHT by **10%**.
- **Optimized cloud infrastructure**, reducing costs by **40%** through containerization, Kubernetes migration, and CI/CD automation.
- **Revolutionized virtual assistant development** with a **low-code, designer-focused self-service tool**, cutting development effort by **75%** and deployment time by **60-80%**.
- **Spearheaded digital transformation**, migrating **legacy IVR to digital chat**, reducing customer support costs by **60%** while maintaining **70% CSAT**.
- **Developed a React Native-based SDK** for iOS/Android chat integrations, reducing customer launch time by **70%**.
- **Built a reusable UI component library** using React.js, improving development efficiency and UI consistency.

### Manager, Software Development | [24]7.ai | Bengaluru | Dec 2010 – Nov 2015

- **Took ownership of two critical frontend frameworks**, leading cross-geographical transition and optimization.
- **Built and mentored a high-performing engineering team**, overseeing the complete product lifecycle.
- **Implemented agile development processes**, ensuring release stability and rapid iteration.
- **Developed cross-browser solutions** for enterprise applications, optimizing for **IE7/8 and legacy mobile OS**.
- **Launched a successful internship program**, mentoring junior developers into high-impact engineers.

### Technology Lead | Infosys | Bengaluru | May 2008 – Dec 2010

- **Led QA operations for Yahoo India's ad contracts system**, ensuring guaranteed ad delivery through rigorous testing and automation.
- **Designed advanced testing methodologies** including stability, latency, deployment, and end-to-end integration.
- **Developed a high-performance ad tracking platform**, implementing Apache server applications and **C/C++ Apache modules**.
- **Engineered robust logging infrastructure** using **log4cpp and Boost on Linux**, improving ad performance monitoring.

### Application Developer | IBM | Bengaluru | Jan 2005 – May 2008
- Developed order management and invoicing systems for **Thomson Reuters**, gaining expertise in **OpenVMS, ProC, and Shell scripting**.
- Strengthened client communication, negotiation, and software development best practices.

### Programmer Analyst | Cognizant | Chennai | Jun 2004 – Jan 2005
- Built test tools for auction industry products using **XML, XSLT, Perl, and automation frameworks**.

---

## PATENTS

- System for handling multi-party interactions with agents of an enterprise and method thereof.

  (**WO2021144723A1**)

---

## EDUCATION

### National Institute of Technology, Tiruchirappalli, India

Bachelor of Technology (Electrical & Electronics), 2004

---

## KEY PROJECTS

### Autogen Agentic flows — *[24]7 Concierge*

Designed **Autogen-powered agentic flows**, seamlessly orchestrating **bot-human collaboration**, reducing cognitive load for users, and improving task completion efficiency.

### Flowize.ai Agent Assistant and trainer — *[24]7 Copilot*

Automated contact center agent training, reducing onboarding time by **30%** and decreasing AHT by **10%** through **real-time response recommendations**.

### LLM proxy hosting — *Prompt Management and LLM Inference/Serving*

Built **on-prem hosting** for LLM inference and prompt management tools, providing **cost attributions, model flexibility, and better version control**.

### Virtual Assistant Builder Tools — *[24]7 Engagement Cloud*

Developed a **low-code self-service tool**, reducing engineering effort by **75%** and deployment time by **60-80%**, empowering non-technical teams to iterate faster.

**RAG-based Information Retrieval** — *[24]7 Answers*

Implemented a **RAG-powered knowledge retrieval system**, enabling instant and accurate self-service for customers and agents. This system auto-updates from PDFs, text files, and HTML, handling **20-25% of intents**, reducing agent workload, and increasing automation containment rate by **10-15%**.

**Cloud Cost Optimization** — *GCP, Own Data Center and Azure*

Migrated **Java and Node.js workloads** to **Docker Swarm and Kubernetes**, reducing cloud costs by **30-40%** and improving deployment efficiency by **50-70%**.

**Business Messaging Platform** — *[24]7 Messaging*

Led development of **Unified Messaging**, partnering with **Apple, Google, and Twilio**, enabling **60% operational cost reduction** while maintaining **>70% CSAT**.

**Open Channel API** — *[24]7 Messaging*

Designed and implemented **Open Channel API**, facilitating seamless integration of messaging platforms with external systems via webhook payloads and OAuth2 authentication.

**Adaptive Content Framework** — *[24]7 Messaging*

Developed a **unified content framework**, allowing brands to specify channel-specific responses automatically formatted for **Apple Messages for Business, SMS, RCS, Web, and Voice channels**. This streamlined integration, reducing customer launch time by **80%**.

**Android and iOS SDK** — *[24]7 Chat*

Built a third-party SDK, enabling brands to embed chat experiences in native apps, cutting launch time in half.

**Web SDK** — *[24]7 Chat*

Developed a **React.js-based Web SDK**, enabling brands to embed chat experiences in websites and native apps, improving chat UI consistency and security compliance.