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 Assistant Development, 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 (in three geographies) developing enterprise-scale AI-powered Conversational AI solutions, managing end-to-end product lifecycle including feature planning, technical design and security compliance for a SaaS suite serving millions of users across digital and voice channels.
- Pioneered AI/ML initiatives implementing **RAG-based Q&A** systems and **Microsoft Autogen**-powered agentic flows, optimizing contact center operations through innovative solutions like **Flowize.ai** for agent training and seamless bot-human orchestration reducing onboarding time by **30%** and AHT by **10%**
- Reduced cloud infrastructure costs by **40%** through strategic containerization initiative, migrating legacy applications to Kubernetes while implementing robust CI/CD pipelines and security frameworks across environments.
- Revolutionized virtual assistant development by implementing a low-code, designer-focused self-service tool, reducing development effort by 75% and accelerating deployment time by 60-80% across various complexity levels.
- Drove digital transformation for major enterprises, achieving 60% cost reduction in customer support through migration from legacy IVR to digital chat platforms while maintaining 70% customer satisfaction SLA, partnering with industry leaders like Apple, Google, and Twilio for enterprise messaging solutions.
- Developed a **React Native-based SDK** for iOS/Android chat integrations and a Web SDK using React.js to reduce launch time for customers by 70%.
- Spearheaded the **TFS-UI** Reactjs-based reusable component library, reducing UI development efforts.

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

- Spearheaded the transition and ownership of two critical frontend frameworks at [24]7.ai, combining hands-on development with strategic optimization across global teams.
- Built and led a high-performing engineering team from scratch, overseeing the full product lifecycle from feasibility analysis to feature rollout.
- Established agile development processes with regular release cycles and feedback loops, ensuring production stability while rapidly advancing from IC to leadership.
- Evolved JavaScript expertise from Prototype.js and Scriptaculous to native implementations, advocating for

framework-independent solutions.

- Specialized in cross-browser compatibility, delivering robust solutions for enterprise applications on IE7/8 and legacy mobile OS.
- Launched and led 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 advertising contracts system, ensuring guaranteed ad delivery through rigorous testing and automation.
- Designed and implemented diverse testing methodologies—Continuous Operation, Stability, Network Latency, Deployment, Code Coverage, Memory Leak, and End-to-End Integration—leveraging Shell and Perl scripting for automation.
- Developed a high-performance ad tracking platform, building full-stack solutions with Apache server applications and custom C/C++ Apache modules.
- Engineered a robust logging infrastructure using log4cpp and Boost on Linux, enabling comprehensive ad performance monitoring across Yahoo and partner networks.

Application Developer | IBM | Bengaluru | Jan 2005 – May 2008

- Worked with Thompson-Reuters on a legacy Order Management, Reporting, and Invoicing system. OpenVMS, ProC and shell scripts were the main learnings.
- This role helped me build effective client communication, negotiation skills, understanding of onshore-offshore, technical analysis, design strategy, software development, and estimation.

Programmer Analyst | Cognizant | Chennai | Jun 2004 – Jan 2005

- Built test tools for products in auction industry using **XML, XSLT, Perl, and automation testing 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

NOTABLE PROJECTS

Autogen Agentic flows — *[24]7 Concierge*

Designed and implemented **Microsoft's Autogen-based agentic flows** for a **user concierge application**, enabling seamless **orchestration between traditional bots and human agents** to efficiently complete end-user tasks. By leveraging **multi-agent collaboration**, the system dynamically determined when to escalate complex queries to human agents while automating routine interactions. This solution is expected to take the cognitive and thinking load from the user by acting on behalf of the user with custom support interfaces (living up to the true meaning of “User Agent”).

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

Implemented **Flowize.ai chatflows** to automate training for newly hired **contact center agents**, reducing onboarding time by 30%. Also provided human agents with a **real-time copilot** on the operations floor, **suggesting next-best responses**. This improved human agent efficiency, with **60-70% of responses being used as-is or modified**, leading to a **10% reduction in AHT (average handling time)** and a **boost in CSAT (customer satisfaction) scores**.

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

On-prem hosting of LLM tools like promptfoo and langfuse for prompt management and litellm and vllm for LLM inference and serving. This helped the data scientists an ecosystem to finetune the prompts, product owners with cost attributions and subscription key management, and application engineers with the flexibility to choose among LLM models and their versions.

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

Built tool for automated message workflows needed by Virtual Agents. This tool transformed virtual assistant application development by replacing **engineering-heavy processes** with a **low-code, self-service platform** for conversation designers. This **reduced development effort by up to 75%** an**d time to market/deploy by 60-80%**, depending on the complexity of the virtual assistant change request. The tool also significantly streamlined workflows, empowering designers to implement changes faster while freeing up engineering resources for higher-value tasks.

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

The **AI-powered unified knowledge base** enabled **instant and accurate self-service** for customers and agents by **auto-updating** with data from **PDFs, text files, and HTML content**. Leveraging a **RAG-based pipeline**, the system efficiently handled **20-25% of intents**, reducing agent workload and improving response times. This resulted in a **10-15% lift in automation containment rate** and a **10% increase in CSAT**, ensuring seamless access to up-to-date enterprise information while optimizing conversational AI performance.

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

Undertook a project to **reduce cloud costs** significantly. Theis initiative was executed in two phases. First, **standalone Java and Node.js applications** were containerized, and their runtime was migrated to **Docker Swarm**, optimizing resource utilization. A **declarative CI/CD framework** was implemented, streamlining release builds, artifact management, environment configurations, ACLs, and security filters.

In the second phase, **workloads** were transitioned from **Docker Swarm to Kubernetes**, enhancing scalability and operational efficiency. As a result, cloud costs were **reduced by 30% (and 40% in some cases) compared to deployment on VMs**, deployment time efficiency improved by **50-70%**, and security was strengthened, ensuring better resource management and resilience.

Common UI Controls Framework — *TFS UI*

I led the design and architecture of **TFS-UI**, an **internal reusable component library** designed to **streamline UI development** and **enhance application consistency**. Built from scratch, **TFS-UI** provides a standardized set of components that accelerate development, reduce redundancy, and improve maintainability. The library has been widely adopted across teams, significantly **reducing UI development efforts** and ensuring a **unified user experience** across multiple products.

Business Messaging Platform — *[24]7 Messaging*

Led the development of **Unified Messaging platform**, collaborating with partners such as **Apple** (Apple Messages for Business), **Google** (Google Business Messaging), and **Twilio** (Programmable SMS for businesses). With a team of **12 engineers** and product owners. This helped the **Digital Transformation** efforts for big Enterprise companies and this transition resulted in a **60% reduction in operational costs** while maintaining a **customer satisfaction rate above 70%** in alignment with the SLA.

Open Channel API — *[24]7 Messaging*

Led the development of Messaging Open Channel API which included an ingress/egress specification using swagger doc, implementation of webhook payload generation and implementation of OAuth2 and jwt security handshakes. This was deployed for a large US retailer which doubled customer revenue and tripled messaging volume for the customer.

Adaptive Content Framework — *[24]7 Messaging*

Led the development of a **unified content framework** for **internal data, so that customers can specify channel-specific responses which will be automatically formatted to adapt to any consumer channel.** Designed and implemented an **internal data schema specification**, ensuring seamless adaptation across **Apple Messages for Business, SMS, RCS, Facebook Business Messaging**, **Web, and Voice** channels. Built **middleware for content transformation**, enabling automatic conversion into **channel-specific payloads**. Developed **rendering libraries** for channel-specific user interfaces, streamlining integration and reducing customer launch time by **almost 80%**.

Android and iOS SDK — *[24]7 Chat*

We offered iOS and Android SDKs as part of our platform for customers to embed messaging experience within their native apps. To effectively manage these channels my team came up with a breakthrough approach of creating a **third-party Mobile SDK** for iOS and Android using **React Native**. The experiment was successful, leading to significant **consolidation of UI components** across the two platforms. This helped our customers launch native channels with messaging in less than half the time.

Web SDK — *[24]7 Chat*

Led the **concept-to-launch** of **247.ai's user-facing frontend UI framework**, which included a **ReactJS-Redux web single-page application** for use by brands and businesses as a third-party JS integration on websites or native apps. My team went through an extensive evaluation of external frameworks and selected React.js. This SDK was the basis of providing a rich frontend chat experience, including Multi-Tab continuity, Audio/Visual Notifications, user data security, and WCAG-compliant accessibility.

Flash app to HTML conversion — *[24]7 WowPx*

For the first time, I experimented with **JavaScript, CSS, HTML**, and **SVG** through projects like building a **mobile bill estimator slider,** which was to **rebuild an existing flash-based application** using modern **HTML, CSS, and JavaScript**. These experiences helped me improve user interfaces and optimize performance across browsers. This helped clients offer UX, which was fully supported by browsers, without the need for the installation of plugins like Flash. The usage of self-service widgets increased by 30%.

Contract Tracking — *Y! Ad systems*

Managed QA activities for a key component in **online ad systems**, specifically related to **Yahoo advertising contracts** and **guaranteed advertisements**. Gained experience in various testing techniques, including **stability**, **network latency**, **deployment**, **memory leak**, and **end-to-end integration tests**, alongside traditional functional and regression tests using **Shell/Perl scripts**. Additionally, worked on a platform for **tracking and logging ads** on Yahoo and partner sites, involving an **Apache server application**, **C/C++ Apache modules**, and utilities like **log4cpp** and **Boost** in a **Linux environment**.