VINOD PATIL

Bengaluru, Karnataka, India | +91 8971308022 | vinod0x01@gmail.com | linkedin.com/in/vinod-a-patil | github.com/vinod0x01

PROFESSIONAL SUMMARY

Senior Software Engineer (SDE3) with 5 years building AI agent platforms and search infrastructure at Lowe's. Architect the Agent2Agent product discovery agent behind the MyLow chatbot — Google ADK, FastAPI, MCP — and cut its LLM token consumption 65%. Led Lowes.com's Elasticsearch-to-OpenSearch migration to semantic and hybrid retrieval across a 3M-document catalog serving 500–2,000 QPS at peak, improving NDCG by about 20% over the keyword baseline.

TECHNICAL SKILLS

Languages: Python, Java, SQL

AI Agents & LLM: Google ADK, Agent2Agent (A2A), Model Context Protocol (MCP), multi-agent orchestration, agent memory optimisation, RAG, LLM integration, prompt engineering, natural-language query understanding

Search & Retrieval: OpenSearch, Elasticsearch, semantic & hybrid search, vector search, embeddings, relevance tuning, query DSL, ML ingest pipelines

Backend & Data: FastAPI, asyncio, Pydantic, Spring Boot, REST APIs, microservices, Redis, MongoDB, PostgreSQL, Kafka, Apache Beam/Dataflow, Airflow

Cloud & Platform: GCP, Kubernetes/GKE, Docker, Istio, NGINX, Jenkins, Ansible, Prometheus, Grafana

PROFESSIONAL EXPERIENCE

Lowe's Companies, Inc.

Senior Software Engineer (SDE3) | Jul 2026 – Present

- Architected and led the Product Discovery Agent, an Agent2Agent (A2A) service on Google ADK, FastAPI and MCP with Redis-backed agent memory and MongoDB transcript persistence, sustaining about 5 req/s of agent traffic through business hours.
- Cut LLM token consumption 65% through a memory-consolidation scheme, dynamic graph workflows and separation of responsibilities across ADK workflows.
- Shipped AI Overview on Search and Product Listing pages, synthesising catalog and review data into a single answer; impact A/B test in flight.
- Built natural-language filtering end to end: an LLM extracts filter intent from plain-text queries and maps it onto Lowe's internal filter taxonomy.

Software Development Engineer | Jun 2024 – Jun 2026

- Led Lowes.com's Elasticsearch-to-OpenSearch migration to semantic and hybrid retrieval across a 3M-document catalog serving 500–2,000 QPS at peak, improving NDCG by about 20% over the keyword baseline for retail and Pro storefronts.
- Cut search p95 latency from 50–100 ms to 20–30 ms cluster-side and 400–500 ms to 300–400 ms end to end; a custom Elasticsearch compression plugin contributed 11 ms of that and reduced index storage 65%.
- Operated two OpenSearch clusters across us-central and us-east — 20 nodes each steady-state, scaled to 35 for Black Friday/Cyber Monday — with zero-downtime blue/green refreshes, snapshot automation and synonym/stopword deployment.
- Built the Apache Beam/Dataflow indexing pipeline and Airflow DAGs behind full and incremental index updates across both regions, including a custom OpenSearch sink where no official Beam connector existed.
- Productionised ML-driven retrieval: ingest pipelines, neural and hybrid index templates, embeddings and vector indexing, and linguistic analysers, validated across 200+ search terms.
- Diagnosed a semantic-search defect in OpenSearch, reported it upstream and worked with maintainers through to the fix; resolved an NGINX healthy-node routing failure affecting cluster traffic.
- Led MyLow's migration from thread-heavy WebSockets to HTTP/2 streaming, built the multimodal image pipeline behind visual product assistance, and ported Agent Services to Lowe's internal ADK — a pattern since adopted by several teams.
- Mentored and onboarded 6+ engineers across search and agent services.
- Received the Lowe's India Team Excellence Award for shipping semantic search on Lowes.com; the platform was subsequently named a finalist for Best Use of AI in Retail at the Modern Retail Awards.

EdgeVerve Systems (Infosys)

Product Developer | Oct 2021 – May 2024

- Designed Java/Spring Boot REST microservices for Finacle core banking — Spring Cloud, JPA, Spring Security and observability instrumentation — holding key APIs under 300 ms at p95.
- Automated Kubernetes deployment with Ansible templates across five targets (on-prem, OpenShift, GCP, AWS, Azure), cutting Finacle delivery time 70%; reduced service-mesh resource consumption 30% through Istio tuning.
- Built a universal data-interface microservice using the decorator pattern, collapsing point-to-point integrations into a single interface.

EDUCATION

M.Sc. Computer Science, Liverpool John Moores University | Awarded Mar 2025

Post Graduate Programme in Software Development (Cloud Backend Development), IIIT-B & upGrad | Jun 2022 – Aug 2023

B.E. Computer Science and Engineering, KLS Gogte Institute of Technology | Aug 2015 – Jun 2019