Vaibhav Vanage

Pune, Maharashtra | 8888642508 | vaibhav.vanage@gmail.com | LinkedIn | GitHub

PROFESSIONAL SUMMARY

Software Engineer with a B.Tech in Computer Science (Data Science) and a published NLP research paper (96%+ accuracy, hybrid transformer pipeline), currently building production ML systems for information extraction, entity resolution, and data enrichment from structured and unstructured sources. Designing NLP, transformer, and LLM-driven pipelines that ingest, normalize, and enrich high-volume data across logs, code, and documents, and translating research into scalable, production-ready systems spanning agentic AI platforms, knowledge graphs, and LLM safety layers.

TECHNICAL SKILLS

Machine Learning & NLP: Large Language Models (LLMs), Generative AI, Transformer Models, Named Entity Recognition (NER), Entity Resolution, Text Classification, Knowledge Graphs, Multi-Agent Orchestration
ML Frameworks & Libraries: PyTorch, Hugging Face Transformers, scikit-learn
Programming: Python, SQL, TypeScript, JavaScript, Java
Data Pipelines & Engineering: ETL, Log Ingestion & Normalization, Data Enrichment, Error Fingerprinting & Clustering, AST Parsing, Tree-sitter, SCIP
Databases: PostgreSQL, MySQL, MongoDB, Vector DBs (Qdrant, Pinecone, VespaDB)
MLOps & Cloud: Docker, AWS, Azure, Azure DevOps, ArgoCD, ELK Stack, Grafana, Prometheus, Ollama
AI Safety & Governance: Prompt-Injection Defense, LLM Output Validation, Data Loss Prevention (DLP), Responsible AI

PROFESSIONAL EXPERIENCE

Bajaj Finserv Health *January 2025 – Present*
Associate Software Development Engineer *Pune, India*

Agentic RCA (Root Cause Analysis) — ML-Driven Data Pipeline

- Architected ML-driven pipeline ingesting production logs from ELK, Azure, and monitoring systems, normalizing heterogeneous unstructured telemetry into structured records for downstream analysis.
- Engineered error-clustering system using log fingerprinting to group similar production issues at scale, eliminating duplicate analysis on recurring incidents and accelerating mean time to diagnosis.
- Built entity-correlation layer linking clustered errors to deployment events (ArgoCD), code changes (Azure DevOps), and infrastructure signals (Grafana/Prometheus) to enrich raw issues with actionable context.
- Integrated LLM-based reasoning layer to generate structured RCA reports identifying root cause, affected services, suspected commits, and remediation steps—converting raw telemetry into reliable, enriched data assets.
- Designed continuously updated knowledge base of resolved incidents, improving freshness and consistency of RCA outputs and eliminating re-analysis of known failure patterns.

AI Cortex — Enterprise Multi-Agent AI Platform

- Designed multi-agent orchestration platform embedded across the SDLC, serving developers, PMs, QA, and TPMs through a unified conversational interface.
- Built agent workflows that extract structured information from unstructured sources (code, logs, APIs, system metadata) and enrich it via LLMs to power documentation generation, requirement refinement, and debugging.
- Engineered tool-use and retrieval layer that connects agents to repositories, logs, APIs, and internal services in real time, grounding LLM outputs in organizational context to reduce hallucinations and keep responses domain-accurate.
- Contributed to architecture decisions on scalable, maintainable AI data infrastructure, aligning pipelines with reliability, latency, and throughput requirements for production use.

Backend & Platform Engineering

- Delivered production-grade document-storage service that has processed 200K+ file uploads and served 500K+ retrieval requests through high-throughput ingestion, storage, and retrieval APIs.
- Built dynamic data-masking middleware for MySQL and PostgreSQL that intercepted queries at runtime and applied role-based visibility rules—identifying sensitive fields, enforcing access controls, and maintaining audit trails to meet compliance-driven backend requirements.
- Integrated DevSecOps tooling into CI/CD pipelines—including SBOM generation, dependency tracking, and vulnerability monitoring—automating security validation across the SDLC and reducing manual review effort.

PERSONAL PROJECTS

KodeAtlas — Open-Source Code Intelligence & Knowledge Graph Platform

- Built continuous code-analysis engine using AST parsing and tree-sitter to extract entities (functions, APIs, services, dependencies) from multi-repo codebases and construct a unified knowledge graph with line-level provenance on every edge.
- Implemented cross-repository entity resolution detecting service interactions via HTTP calls, shared libraries, and message queues—enabling impact analysis, dependency tracking, and system-wide visibility.
- Integrated LLM-based enrichment layer producing hierarchical summaries from function to service granularity, transforming raw code into living, continuously updated documentation.
- Combined deterministic parsing with AI-driven enrichment over scalable pipelines to demonstrate end-to-end extraction of structured data from unstructured sources.

Zephyron — Open-Source Adversarial Red Teaming Framework for LLMs

- Built open-source Python framework implementing 34 adversarial attack techniques against LLMs, translating recent NLP research (Nature Communications 2026, ICLR 2025, NeurIPS 2025) into modular, production-ready attack, judge, and mutator components.
- Designed multi-stage LLM evaluation pipeline—deterministic refusal detection, LLM-as-judge classification, 10-assertion weighted signal voting, and hard-override rules—achieving **0% false-positive rate** on a 24-case golden validation dataset.
- Engineered adaptive Thompson-sampling orchestrator dynamically allocating scan budget across attack categories by observed bypass rate, shipped with live dashboard, SARIF/HTML reports, and automated compliance mapping (OWASP LLM Top 10, MITRE ATLAS, NIST AI RMF, EU AI Act).

LLM Guard & LeakGuard — AI Safety & Security Systems

- Designed multi-stage LLM middleware combining rule-based heuristics (regex), classification models, and semantic techniques to detect prompt injection, jailbreaks, and sensitive-data exposure, producing explainable allow/flag/block decisions suited for enterprise auditability.
- Built output-validation pipeline to flag hallucinated content, PII leakage, and policy violations before responses reach end users, operationalizing responsible AI practices.
- Developed LeakGuard, a browser-based DLP extension intercepting user input locally and applying regex heuristics to block API keys, tokens, and PII from reaching external AI interfaces—processing fully on-device with zero data transmission.

RESEARCH & PUBLICATIONS

Hybrid Context-Aware Fake News Detection (HCA-FND) — Published Research Paper

- Designed two-stage NLP pipeline: TF-IDF + logistic regression for fast candidate filtering, followed by DistilBERT transformer for deep semantic classification—balancing accuracy, latency, and interpretability for real-time use cases.
- Incorporated Named Entity Recognition (NER) and real-time fact-checking to validate extracted claims, achieving **96%+ accuracy** across multiple evaluation metrics.
- Demonstrated production-oriented approach to information extraction and text classification by combining classical ML with modern transformer-based NLP.

EDUCATION

SVKM's Dwarkadas J. Sanghvi College of Engineering, Mumbai, India — *2022 – 2025*
B.Tech in Computer Science and Engineering (Data Science) — *CGPA: 8.1/10*

Dr. Babasaheb Ambedkar Technological University, Raigad, India — *2019 – 2022*
Diploma in Computer Engineering — *Percentage: 92.33%*