SRIHARSHA VELICHETI

srih8991@gmail.com | Github | LinkedIn | Portfolio | Phone:+91 8309012139

SUMMARY

II am Sriharsha Velicheti with close to 2 years of experience in AI/ML engineering, with expertise in Machine Learning, Deep Learning, Python, NLP, and Generative AI, including LLMs, RAG, and agentic workflows. I've built production-grade ML pipelines, multimodal systems, and scalable cloud-based solutions, and I'm passionate about solving real-world data challenges while delivering reliable, high-impact AI systems. I continuously strive to refine my craft, adapt to emerging technologies, and build AI solutions that create meaningful business value.

SKILLS

- Core: Generative AI, LLM Orchestration, NLP, Prompt Engineering
- Frameworks & Libraries: LangChain, Gemini AI, OpenAI API, Hugging Face Transformers,
- Cloud & DevOps: Azure AI, GCP (Vertex AI), Docker
- Programming: Python, SQL, [Typescript, (vibe code)]

EXPERIENCE

TenderGenie (Document intelligence application) | (Datasmithai PUNE) **Nov 2024- Present**

- Built TenderGenie, an document-intelligence Agentic chatbot for manufacturing tenders, enabling natural conversational querying with chat-history awareness, context retention, and intelligent summarization.
- Designed a RAG-based multimodal pipeline (Gemini AI + Qdrant + FastAPI) capable of understanding tenders, datasheets, compliance forms, and technical tables, delivering accurate answers and structured summaries.
- Engineered a high-performance Tree (Trie Data structure) based Search for keyword extraction engine achieving sub-10-second processing speeds, resulting in ~100× efficiency improvement over manual workflows.
- Developed multi-service, multi-stage CI/CD pipelines using Azure DevOps with Dockerized services deployed through ACR → Azure App Services; added automated chatbot evaluation in CI by generating question sets from source documents for RAG performance benchmarking.
- Led the end-to-end architecture, optimization, and production deployment, achieving ~89.77% accuracy and reducing tender review time by 10×

Autonomous Court Case Automation in MHADA | DatasmithAI,PUNE **Nov 2024- Present**

- Built a complete browser automation system for MHADA (Maharashtra Housing and Area Development Authority) to handle case-query operations across dynamically changing court portals (Supreme Court, High Court, Civil, and Magistrate Courts), enabling reliable data extraction, navigation, and workflow automation with full end-to-end ownership.
- Impact: Delivered 80% less manual work and **3× faster case-query** processing through fully automated multi-court data retrieval
- Tech Stack: Python, Playwright, Async Automation Pipelines, Dynamic Page Handling, Error-Resilient Scraping & Orchestration.

Go-Bundle Agentic Automation Engine | (DatasmithAI, PUNE)

- Built a fully agentic browser **automation system** using Gemini Computer Use, enabling autonomous multi-service subscription signups with intelligent page detection, reasoning-driven actions, and adaptive retries.
- Designed a modular, database-driven bundle automation engine with encrypted credential management, parallel executions, webhook tracking, and production-grade orchestration for large-scale automated signup flows.
- Built using Browserbase, Stagehand (Gemini Computer Use), TypeScript, and Supabase with Prisma for database and credential management.

Multimodel RAG application (Research intern @ Siemens Bangalore) **Jan 2024- May 2024**

- Researched end-to-end RAG architectures, including complex PDF ingestion, structured data extraction, schema aware parsing, and workflow automation with LLMs acting as validation "judges" for cleaner pipelines.
- Designed and evaluated multiple vector store ingestion strategies, optimizing embeddings, retrieval, and relevance scoring across different RAG setups.
- Implemented multi-LLM response generation and benchmarking for accuracy, consistency, and factual grounding.
- Contributed to a project integrating PID diagrams into knowledge graphs, applying Graph-RAG concepts for real-time plant monitoring and emergency surveillance.

PROJECTS

Resume Filtering system for HR |

- Created a Generative AI-powered resume management system that automates the filtering, ranking, and classification of resumes through agentic classification and scoring based on cosine similarity.
- Conducted NLP-based candidate analysis, which included skill extraction, technical proficiency scoring, and interactive visualizations, along with assessing strengths and weaknesses by employing a large language model as a judge.
- Integrated AI-driven functionalities such as generating role-specific interview questions, conducting market salary analyses, and automating acceptance/rejection emails via SMTP.
- Developed a modular Streamlit application using Python, spaCy, scikit-learn, PyPDF2, Qdrant, and Groq AI, ensuring a scalable, maintainable, and comprehensive workflow management system.

EDUCATION

Jain University | Bengaluru | Bachelor's degree **Aug 2020 - June2024**
Bachelor of Technologies CSE (Data Science) - CGPA 8.756
Honors in Data science (specialization)

ACHIEVEMENTS / AWARDS

- **Achievements:**
  - TenderGenie Product Deployed and 3 business clients in less than a year of joining in DatasmithAI .
  - Secured 6th individual position out of 159 participants in MACHINE HACK hackathon with an accuracy of 92.6% (during college)
- **Awards/Activities:**
  - Successful Workshop regarding Generative AI in I2IT Pune College, **representing DatasmithAI.**
  - **President** of Data Science Student Club for 10 months.
  - Data Science Facilitator at ***Google Developer Student Club*** for a year and mentored over 100 students in the field of AI/DS/ML
- Certifications: NPTEL certification on Python for Data Science with 69% score in the Exam conducted by IITM