Shailesh Pranav Rajendran

Frisco, Texas | +1 240 - 733 - 6077 | shaileshpranav.r@gmail.com
https://www.linkedin.com/in/shailesh-pranav-r | https://shaileshpranav.github.io/

Results-driven Software Engineer specializing in Generative AI and machine learning with proven expertise in designing enterprise-grade solutions. Track record of optimizing performance and enhancing data privacy while driving technological innovation across multiple domains.

Experience

WorldLink US – Data / ML Engineer (Nov 2023 - Present)

Technologies: MeloTTS, Whisper, TensorRT, FastAPI, Docker, Kubernetes

- **Real-Time Speech Interaction:** Implemented real-time speech-to-text and text-to-speech capabilities using open-source models. Designed an on-premises system ensuring data privacy, security, and compliance with enterprise standards.
- **LLM Optimization with TensorRT:** Optimized LLM inferencing using TensorRT, reducing inference latency and increasing throughput for real-time chatbot applications.
- **Chatbot Containerization:** Containerized the chatbot application using Docker & Kubernetes, enabling seamless deployment, scalability, and maintainability.

Unleashed Brands – Phase I (Contract)

Technologies: FastAPI, PostgreSQL, Redis, Milvus, TensorRT

- **Enterprise-Grade Chatbot Deployment:** Deployed a secure on-premises chatbot leveraging Generative AI. Designed and implemented a Retrieval-Augmented Generation (RAG) architecture using Milvus (vector database), enhancing retrieval accuracy and reducing response latency by 25-30%.
- **Data Pipeline Design:** Built a scalable data pipeline integrating FastAPI and vector databases, optimizing chatbot performance and data retrieval efficiency.

Unleashed Brands – Phase II (Contract)

Technologies: FastAPI, Redis, JWT, PostgreSQL

- **Secure Document Management System:** Developed a secure backend for document management using Python, PostgreSQL, and Redis.
- **RBAC & Authentication:** Implemented custom Role-Based Access Control (RBAC) with JWT authentication and row-level security, automating document workflows and reducing manual processing time by 20 hours per week.
- **Logging & Compliance:** Integrated logging and audit mechanisms to ensure traceability and regulatory compliance.

New York University Tandon School of Engineering – Research Intern (Feb 2020 - May 2020)

Technologies: Embedded Systems, C++, Python, SolidWorks, Arduino

- Designed and fabricated a rehabilitation device for stroke patients with partial arm maneuverability loss under the guidance of Dr. Vikram Kapila.
- Developed embedded software for medical rehabilitation devices using C++ & Python.
- Implemented real-time sensor data processing and motor control systems using Arduino.
- Created modular, reusable software components, reducing development time by 40%.
- Designed an automated testing framework for device calibration, improving accuracy and efficiency.

Skills

- **Programming & Scripting:** Python, C, C++, C#, Java
- **ML & AI:** TensorFlow, PyTorch, Llama-Index, Langchain, Keras, Scikit-Learn
- **DevOps & Infrastructure:** Jenkins, GitHub Actions, AWS, Azure, GCP, Raspberry Pi, Git, Docker
- **Databases:** Neo4j, Chroma DB, Milvus, PostgreSQL, Redis, Neo4j
- **Software Architecture:** Distributed Systems, System Design, API Design

Education

May 2023 **University of Maryland** College Park, MD, USA
Master of Engineering: Robotics Engineering

May 2020 **PSG College of Technology** Coimbatore, TN, India
Bachelor of Engineering: Robotics and Automation Engineering

Leadership Experience

TIDES Conference September 2016
• Volunteered as an organizer for the TIDES Leadership Summit, conducted by the Confederation of Indian Industry (CII).

ENEXT Conference April 2016
• Lead the development team at the E-NEXT Conference hosted by the Entrepreneurs Club of PSG College of Technology