SAKTHIVEL SENTHILKUMAR

+91 63697 18696
sakthivelnunnarilabs.com
Coimbatore, India
Sakthivel Senthilkumar

Summary

Sakthivel is a dedicated Machine Learning Engineer at Nunnari Labs with expertise in a wide range of AI domains, including Machine Learning (ML), Natural Language Processing (NLP), Deep Learning (DL), Computer Vision (CV), Generative AI, and AI agents. He excels in creating scalable, end-to-end AI solutions that tackle complex, real-world challenges. His approach combines advanced AI techniques with a deep understanding of industry needs, ensuring impactful and practical solutions.

A strong advocate for AI governance, Sakthivel is committed to developing ethical, transparent, and trustworthy AI systems. He prioritizes fairness, accountability, and inclusivity in all his work, ensuring that AI systems align with societal values and deliver meaningful outcomes. His dedication to Responsible AI drives him to shape systems that not only meet technical objectives but also adhere to ethical standards, fostering trust among users.

Beyond his technical expertise, Sakthivel is passionate about mentoring the next generation of AI professionals. He actively conducts workshops, technical training, and mentorship programs, bridging the gap between theoretical knowledge and real-world AI applications. At Nunnari Labs, he has contributed to advanced AI systems such as Topic Modelling, RAG, hybrid search, LLM fine-tuning, and knowledge graph-based applications. He also has hands-on experience with Amazon Bedrock for deploying large language models and Amazon Lex for building conversational AI applications, ensuring reliable and scalable deployment in production environments.

Professional Experience

Machine Learning Engineer | Nunnari Labs Pvt. Ltd, India — Apr 2023 – Present

- Actively contribute to projects involving machine learning, focusing on solving real-world challenges in areas such as Computer Vision, Natural Language Processing (NLP), Edge AI and Generative AI.
- Develop and deploy ML models tailored to specific project requirements, ensuring efficient integration with production environments.

Huawei Intern Developer | Huawei Services, Hong Kong (Remote) — Jan 2022 – Jul 2022

- Participated in the Huawei Intern Program as an Android Development Intern, gaining hands-on experience in Android application development.
- Contributed to the completion of various development tasks, collaborating with senior developers to enhance app functionality and optimize performance.

Technology Summary

Language Expertise: C, C++, Java, Python.
ML Frameworks/Tools: Langchain, Anthropic, Cohere, Openai, LangGraph, Autogen, Milvus, Chroma, Qdrant, Arize, Haystack PyTorch, HuggingFace, OpenCV, scikit-learn,ultralytics, TensorFlow, TFLite, Pandas, Numpy, Matplotlib, Seaborn, Plotly, Spacy, Gensim,Deepstream, TensorRT, Onnx, Nvidia AI Enterprise Stack, Nvidia Nemo, Llama Factory, Qualcomm AIMET.
Cloud ML: Google ML Engine, Amazon Bedrock, Amazon lex, IBM Watson, Azure ML, Oracle OCI.
Web Frameworks, Tools, Methodologies & Environments: Django, Flask, FastAPI, Docker, Kubernetes, REST, Agile, VS Code, STS, git, postman, Bruno swagger,Clickup, trello, Slack, Jira, Confluence, Sonar Qube.
Protocol Expertise: HTTP, WebSockets, REST, MQTT.
IoT Cloud Platforms: Google Cloud Platform, Bosch IoT Hub, Azure IoT Hub, Ubidots, IBM Watson
Databases: MySQL, MongoDB, Firebase
Operating Systems: Mac, Linux, Windows

Education

B.Tech Information Technology
Anna University, First Class, Completed: 2023

Certifications

Machine Learning with Python
August 2022
View Certificate

Fundamentals of Accelerated Computing with CUDA C/C++
July 2022
View Certificate

Fundamentals of Accelerated Computing with CUDA Python
July 2022
View Certificate

Projects

1. **AI-Powered Assistive Vision System:**
   - **Technologies**: Custom YOLO, QCS6490 Edge AI, Computer Vision, NLP, TTS/STT, OCR, Face Recognition, Temporal Mapping, UPI Integration, Android Development, PyTorch/ONNX.
   - **Description**: Developed an AI-powered wearable assistive system for visually impaired individuals featuring a custom YOLO model, real-time obstacle avoidance, face recognition, indoor navigation with return guidance, currency identification, UPI payment integration, and multi-language text-to-speech.
2. **Sparzaai - Topic Modeling**
   - **Technologies**: FastAPI, MongoDB, Ollama, HDBSCAN, PyMongo, Azure Blob Storage, NLTK, Pandas.
   - **Description**: Developed an enterprise-grade topic discovery system that processes communications from multiple sources (chat, email, tickets). The system employs LLM-based keyword extraction with Gemma 3:27B via Ollama, combined with hierarchical clustering using HDBSCAN. Features include incremental learning, version-controlled topic management, and a comprehensive API for topic refinement and document mapping.
3. **Rudo - Report Analyzer**
   - **Technologies:** FastAPI, Google Cloud Storage, Claude 3.5 Sonnet, Docling, OCR, EasyOCR, ROUGE, BLEU, BERT Score.
   - **Description:** Developed a healthcare-focused system that automates the extraction, summarization, and querying of medical documents.
4. **AI Interviewer**
   - **Technologies:**LangChain, FastAPI, Ollama, MongoDB, Google Cloud.
   - **Description:**Developed an AI-based interviewer that automates the generation of interview questions and evaluates candidate responses based on skill sets.
5. **Real-Time AI Welding Assistance**
   - **Technologies:**Opencv, Ultralytics, TensorRT, Onnx.
   - **Description:** Developed a real-time AI-powered welding assistant system using computer vision and deep learning. The system detects welding defects and provides feedback to welders during the process, improving quality control.
6. **Invoice Analyzer**
   - **Technologies:** Python, FastAPI, Azure Form Recognizer, REST API.
   - **Description:** Developed an Invoice Analyzer application using Azure Form Recognizer, built with FastAPI. The system provides an API that allows users to upload invoice images or base64-encoded data for analysis.
7. **Traffic Flow - Real-Time Vehicle Monitoring System**
   - **Technologies:**YOLOv5, DeepSORT, Computer Vision, Object Detection.
   - **Description:** Developed a real-time vehicle monitoring system using YOLOv5 for object detection and DeepSORT for vehicle tracking.

8. **Early Stage Diabetic Retinopathy Detection**
    - **Technologies:** Convolutional Neural Networks, Image Classification, Deep Learning.
    - **Description:** Developed an early-stage diabetic retinopathy detection system using Convolutional Neural Networks (CNNs) for image classification.

9. **Yoga Assistance Using Mediapipe**
    - **Technologies:** MediaPipe, Computer Vision, Pose Detection, Machine Learning
    - **Description:** Built a yoga assistance system using MediaPipe for pose detection. The application leverages computer vision and machine learning algorithms to recognize various yoga poses performed by individuals.

Volunteering Experience

1. **AI Tamilnadu Community**

    - Where we fellow members organize local meetups, community events and workshops encouraging free AI Education.

    - Performed continual pre-training and fine-tuning of Gemma 3 1B and Qwen 3 4B models using Tamil Wikipedia dataset for pre-training and Tamil AAYA dataset for supervised fine-tuning. Published the fine-tuned models on Hugging Face for community use.

2. **TensorFlow User Group Coimbatore**

    - A TensorFlow user group and forum focusing on democratizing AI in the community sponsored by Google India.

3. **Google Developer Community**

    - GDG is a community which organizes meetups, webinars and fests! It is a place to connect, learn and grow. We are also in the pursuit of identifying the hurdles in our community that can be solved by code.

Interests

- Mathematics, AI, Mentoring, Physics, Sci-fi movies, Bikes, Music and Basketball.

Activities & Participation

- Winner of the CITY JS & Wikimedia Hackathon 2022.
- Delivered a talk on "Gemma 2: Advancing Power, Safety, and Transparency in AI" at TFUGCBE LLM Unplugged 2024.
- Delivered multiple talks on Responsible AI, AI Ethics, Classical Machine Learning, and related topics at various conferences and community events.