VIKNESH S

ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING PROFESSIONAL

viknesh.datasci@gmail.com | Ph:+91 9940705356 | github.com/VIKNESH1211 | linkedin.com/in/viknesh-s-2a20a9218

EDUCATION

SRM INSTITUTE OF SCIENCE AND TECHNOLOGY — **Chennai, India**
Bachelor of Technology — **Sept 2021 - Jun 2025**

Major in Computer Science and Engineering
Specialization in Artificial Intelligence and Machine Learning
Cumulative CGPA: 8.3/10
Relevant Coursework: Statistical Machine Learning, Deep Learning, Computer Vision, Natural Language Processing, Operating Systems, DBMS, Data Structures, Fuzzy Logic.

PROFESSIONAL EXPERIENCE

Avkalan.AI, London (Remote) — **Feb 2024 - Dec 2024**
AI Engineer

- Developed a Retrieval Augmented Generation (RAG) AI agent to improve fashion recommendations by combining vector search with Qdrant and personalized content generation using the Claude 3.5 Sonnet model.
- Utilized AWS services such as Bedrock, Textract, and EC2 to create a chatbot-based solution for extracting medical reports and prescription information efficiently.
- Implemented an information retrieval tool using YOLOV8, Pytesseract (OCR), and Open-CV's thresholding, designed to extract structured data from government IDs for streamlined processing.

Rheo AI Solutions Pvt. Ltd., Chennai — **July 2024 - Sept 2024**
Systems Engineering Intern

- Integrated AWS Bedrock & ElasticSearch and populated an ElasticSearch database with high-quality embeddings using advanced LLMs from Cohere and Amazon Titan, optimizing existing retrieval in an RAG-based chatbot.
- Conducted in-depth performance analysis of Claude 3 Opus, Titan Text Premier, and Command R+ models within the RAG process, which helped the company scale up the product using AWS Bedrock, furthermore increasing overall efficiency by more than 30%
- Employed ReRanking and RAG Fusion to optimize retriever efficiency in the RAG workflow, significantly enhancing the chatbot's performance and aiding in a breakthrough to use the bot in high-stakes situations.

Samsung Research Institute Bangalore, PRISM Program — **May 2023 - Nov 2023**
Research Intern

- Collected over 2000 image samples and applied image augmentation techniques to expand the dataset for key-point detection tasks, improving overall dataset quality and diversity.
- Acquired expertise in key-point and bounding box annotations by utilizing CVAT and mastering Samsung's proprietary annotation tool, ensuring precise and efficient image labelling.
- Performed an extensive survey of over 30 research papers and implemented YOLOv8-pose detection, while also developing a key-point detector using Keras open-source frameworks.

Renault Nissan Automotive India Pvt Lt, Chennai — **Feb 2023 - May 2023**
AI/ML Developer and RPA Developer

- Worked under the Director of Supply Chain Management, to identify repetitive and time-consuming tasks requiring precision, and developed Robotic Process Automation (RPA) solutions that reduced operational costs and time.
- Developed advanced RPA solutions using Renault Nissan's proprietary tool IAUTOMATE to optimize supply chain workflows, enhancing time efficiency and reducing human error.
- Designed a GUI-based Time Series Forecasting tool for the Out-Bound Logistics Department using Temporal Convolution Networks (TCN) and XG-Boost Regressor, enhancing predictive accuracy and providing insights that significantly improved shipment schedule optimization and logistics decision-making.

RESEARCH EXPERIENCE

Research Collaborator – National Institute of Technology Tiruchirapalli — **Dec 2024 - Present**
Research in Computer Vision and LLM Optimization

- Researching Deep Image Prior-based segmentation techniques, focusing on single-image training models for efficient and accurate segmentation.
- Exploring LLM quantization methods to reduce model size & computational load while preserving performance.
- Working on infrastructure-level optimization techniques for scalable and cost-effective deployment of Large Language Models (LLMs) across resource-constrained environments.

HyperVerge Nexus program - Spring 2024 **Feb 2024 - Jun 2024**
(Link)

- Developed a Deepfake detection system using LipSync analysis with the Audio Visual Hidden Unit BERT (AV-HUBERT) model, where I introduced architectural modifications to enhance detection accuracy.
- Chosen as a fellow for the Nexus Program.
- Conducted in-depth exploration of Deepfake generation and detection technologies.

Deep Learning Research Collaboration with UAH **Jan 2022 - Aug 2023**
Undergraduate Research in AI for Earth Science

- Analyzed satellite images from VIRS and MODIS data provided by NASA-IMPACT, manually annotating over 300 images using Make-Sense-AI for experimental purposes.
- Developed and trained a custom Convolutional Neural Network (CNN) architecture for detecting Transverse Cirrus Bands (TCB) in satellite imagery, and fine-tuned models like EfficientNet-V2, MobileNet-V2, VGG-16, and ResNet, achieving 93% accuracy. Alongside, fine-tuned Vision Transformers and Swin Transformers, reaching up to 96% accuracy.

CONFERENCE & PUBLICATIONS

1. Co-authored Publication @ **IEEE Xplore** – *Detection of Cracks Using LeNet-5 with Mish Activation: A Comparative Analysis with Swin Transformers*, presented at the 5th International Conference on Electronics and Sustainable Communication Systems (ICESC 2024).
2. *A Manually Annotated Dataset of Transverse Cirrus Bands for Object Detection in Satellite Imagery*, presented at ICMSEE 2025, **Debre Tabor University, Ethiopia**, to be published in **Taylor & Francis Conference Proceedings**, indexed in **Scopus**.
3. Upcoming Publication @ **Springer CCIS** - *Deep Image Prior-Based Segmentation: A Study and Proposal of Deeper U-Net*, accepted at the NCVPRIPG 2025 Jointly organised by **IIT Jammu , NIT Srinagar, NIT Jalandhar and IUPRAI**
4. Upcoming Publication @ **Science Direct Journal** - *Comparative Analysis on deep learning models focused on Detection of Transverse Cirrus Bands using Satellite imagery* (In Review)

ACADEMIC PROJECTS

Government-Funded Project on Machine Learning Applications in Quantum Cryptography

Contributed to a Central Government-funded project on quantum cryptography under the guidance of the Ministry of Electronics and Information Technology (MeitY), Government of India. Utilized AWS Braket services to explore basic machine learning applications integrated with quantum computing for cryptographic purposes.

Legal Text Summarizer and Translator

Developed a full-stack application for summarizing and translating legal text. The application employs the facebook/bart-large-cnn model for summarization and the facebook/mbart-large-50-one-to-many-mmt model for translation, enabling efficient processing of complex legal documents.

SKILLS SUMMARY

Technical Skills: Python, C, C++, Machine Learning, Deep Learning, Computer Vision, NLP, Swin Transformers, Robotic Process Automation, YOLO Versions, AWS Services, Time Series Analysis, Information Retrieval
Web development: HTML/CSS/Javascript/ React.Js, App Development: Flutter.
Certifications & Training: IBM Machine Learning Professional Certification, IBM Statistics 101, IBM Introduction to Machine Learning - Specialization, Introduction to Machine Learning IITM (Silver Badge), Python for Data Science IITM (ELITE).
Software Known: Excel, Powerpoint, LaTex, TensorFlow, PyTorch, Arduino IDE.

PROFESSIONAL ASSOCIATIONS

Next Gen AI SRMIST **Aug 2023 - Aug 2024**
President
During my tenure as President of Next Gen AI SRM, I led 200 students and focused on fostering a collaborative learning environment across 5 domains. Furthermore, I organized hackathons and conducted boot camps that empowered students through hands-on experiences in AI and Machine Learning.

Data Science Community SRM **May 2023 - Jun 2024**
Research Director
As the Research Director of the Data Science Community at SRM, I led a team of 30 members, guiding them through the intricacies of conducting research and authoring scientific papers. My role was instrumental in developing our group's capabilities and fostering a collaborative environment for Research and Development.