

AKSHAY S G

 SG-Akshay10 |  Akshay S G |  akshay10sg@gmail.com |  +91 9176090049

Education

Shiv Nadar University Chennai

B.Tech Artificial Intelligence and Data Science

Sep. 2021 – July 2025

CGPA: 8.5/10.0

Experience

Digital.ai

Intern

Feb 2025 - Present

Bengaluru, India

- Built a **Retrieval Augmented Generation (RAG)** based chatbot using **LangChain**, **FastAPI**, and **PGVector** to automate responses for documentation-related queries, leading to a **40%** reduction in monthly customer support inquiries and significantly decreasing reliance on the services team.
- Designed **AI-Quick-Insight**, a **multi-agent NL-to-visualization system** leveraging **LlamaIndex**, **PGVector**, **Snowflake**, and **FastAPI** to generate interactive charts and insights from user queries in under **30** seconds.
 - Implemented table retrieval over 100+ schema objects using RAG for top-5 context selection.
 - Created a conversational refinement pipeline using an LLM-judge to resolve ambiguous or incomplete queries dynamically.
 - Achieved **63%** accuracy in auto-generated visualizations comparable to in-house data analyst output, significantly accelerating time-to-insight.
- Developed a vulnerability detection system for decompiled binaries that flags risky code sections and recommends security guardrails.
 - Built an automated vulnerability detection system leveraging deep learning models trained on security datasets to identify vulnerabilities in Smali bytecode, achieving **64%** prediction accuracy.
 - Implemented **LLM-as-judge** with RAG-enhanced on Smali documentation to validate neural network predictions and determine appropriate security guardrail recommendations.

Matrice.ai

Machine Learning Engineering Intern

June 2024 - Jan 2025

Remote

- Built a drag-and-drop **no-code computer vision platform** allowing users to **train**, **annotate**, **export**, and **deploy** models without writing code.
- Developed cloud-based data ingestion pipeline with parallel batch processing for efficient customer uploads, achieving **75%** reduction in upload time by replacing sequential zip extraction with multi-threaded folder processing.
- Built **AutoML** annotation system integrating **SAM (Segment Anything Model)** for automated image classification, bounding box detection, and pixel-level segmentation, for labeling unlabeled datasets.
- Engineered automated training pipeline with configurable parameters (epochs, learning rate, model selection) and integrated multiple state-of-the-art computer vision architectures including **YOLO**, **EVA** and **DINO**.
- Implemented model export functionality supporting multiple formats (**ONNX**, **TensorRT**) with seamless deployment pipeline using hosted inference servers.

Sports Mechanics India Pvt Ltd

Data Science Intern

Dec 2023 - May 2024

Chennai, India

- Developed a **real-time notification engine** using Python to analyze live match data from cricket databases and detect key events using a rule-based system and delivered notifications to users through the platform app.
- Fine-tuned a **Mistral-7B** language model (quantized on GCP) for generating **AI cricket commentary** using enriched ball-by-ball metadata.
- Integrated the notification engine with the commentary system to **dynamically enrich live commentary**, reducing the manual effort for human commentators by **80%**.
- Trained a **YOLO-based video analytics tool** to extract bowler metrics (ball grip, delivery type, line, length, speed) from match footage, enabling performance analysis for coaching and scouting use-cases.

Projects

Chat-with-PDF | Python, PyMuPDF, Gemini API, Langchain, FastAPI, React.js

[Link to Github](#)

- Full-stack RAG QA system with source citation over multiple document uploads.

TweetGuardian | Python, Tensorflow, Pytorch, Scikit-learn, Hugging Face

[Link to Github](#)

- Developed twitter tweet abuse classifier using BERT+MLP; achieved **F1 score: 0.91**.

Technical Skills

Languages : Python, C, JAVA, JavaScript, HTML/CSS, React.js

Developer Tools : Tableau, SQL, Figma, GitHub, Hugging Face, PyTorch, TensorFlow, Docker