

Vjaylakshman K

vjaylakshman15@gmail.com | linkedin.com/in/vjay15 | github.com/vjay15 | [Website](#) | Chennai

EXPERIENCE

DevOps Engineer Intern

Nov 2025 – Present

SurveySparrow

Chennai

- Undergoing hands-on training in core DevOps concepts, tools, and workflows
- Learning version control (Git), Linux basics, and scripting to build foundational skills
- Exploring CI/CD, containerization, and cloud fundamentals under the guidance of the DevOps team

AI Engineer Intern

Jun 2025 – Jul 2025

Greatify

Chennai

- Applied supervised fine-tuning via OpenAI API, Vertex AI, HuggingFace, and LLaMA Factory
- Fine-tuned quantized LLaMA 3 8B (4-bit, LoRA) for rubric-based answer evaluation
- Developed and deployed Django + vLLM APIs for model inference

Database Designer Intern

Aug 2024 – Nov 2024

Nabhaas Cloud Consulting

Chennai

- Designed normalized Oracle 19c schema for 50K+ daily water usage records across 20 city zones
- Collected, processed, and indexed datasets for faster queries on 20K+ record tables

EDUCATION

B.S. Abdur Rahman Crescent Institute of Science and Technology

Oct. 2022 – 2026

Bachelor of Technology in Computer Science and Engineering

CGPA: 9.24

Indian Institute of Technology Madras

Sep. 2022 – Present

Bachelor of Science in Data Science and Applications

CGPA: 8.25

PROJECTS

PDFChat – RAG-based PDF QA Web App | *Python, Langchain, Streamlit*

Nov 2024

- Developed a retrieval-augmented generation web app capable of querying 50+ page PDFs with high accuracy
- Integrated Mistral 7B LLM with Cohere embeddings, enabling sub-2 second responses for 1K+ token queries
- Designed an interactive Streamlit UI supporting multiple simultaneous queries with real-time feedback

Quiz Master Application | *Python, Flask, SQLite (SQLAlchemy), JavaScript, ChartJS*

Mar 2025

- Built a role-based quiz app with structured content management and secure user authentication
- Implemented timed quizzes with auto-submit and real-time analytics using ChartJS

AI Evaluation API System | *Django, Python, OpenAI, vLLM, Cloudflare*

Jul 2025

- Developed a secure Django-based API with API key authentication, SHA256 request validation, and activity logging
- Implemented dynamic subdomain routing with database switching to serve fine-tuned LLaMA 3 and GPT models
- Created a Streamlit-based tester for model selection, live API queries, and response visualization via Cloudflare tunnels

LLaMA 3.1 8B QA Fine-tuning | *Python, Unsloth, HuggingFace, bitsandbytes*

Jul 2025

- Fine-tuned LLaMA 3.1 8B (4-bit LoRA) using Unsloth's optimized training pipeline on a custom QA-rubric dataset
- Enhanced rubric-based answer evaluation accuracy and consistency, aligning closely with grading standards
- Leveraged memory-efficient training to run on a single GPU without compromising performance

TECHNICAL SKILLS

Languages: Python, Java, C/C++, JavaScript, SQL, HTML/CSS, PHP

Tools & Technologies: Git, Azure, AWS, Linux, VS Code, Hadoop, XAMPP, GCP, OpenAI

Libraries & Frameworks: pandas, NumPy, scikit-learn, Langchain, Flask, PyQt, Chart.js, Langgraph, Huggingface, vLLM, LLaMaFactory, Django