

Aryan Raj

+91-8287276911 | aryanraj2713@gmail.com | linkedin.com/in/aryanraj13 | github.com/aryanraj2713

EDUCATION

SRM Institute of Science and Technology

B.Tech in Computer Science and Engineering with specialization in Software Engineering

CGPA: 8.4/10

Chennai, IN

May 2021 – May 2025

EXPERIENCE

Software Engineer (AI/ML)

ValueLabs

Dec 2025 – Present

Hyderabad, IN

- Built 15+ deep learning and generative AI MVPs for client use cases, accelerating solution validation cycles by 30% and driving internal AI adoption across cross-functional teams
- Enhanced AiDE and subsidiary tools by integrating multiple productivity platforms and workflow automations, improving developer throughput and reducing repetitive engineering effort by 25%
- Architected analytics pipelines, data warehousing, and lakehouse infrastructure for 20+ KPI dashboards and executive reporting, enabling faster, data-driven decision-making across business units

Machine Learning Engineer

SEOstack (WT Softech)

Sep 2025 – Dec 2025

Hyderabad, IN

- Engineered production LLM orchestration workflows and multi-agent pipelines for business automation, reducing latency and failure rates by 20% across core use cases
- Deployed AWS serverless chatbot architecture (Lambda, API Gateway) handling high-volume customer interactions, lowering operational costs by 15% while improving response quality and engagement metrics
- Developed robust prompt evaluation and observability pipelines for LLM applications, improving response consistency and enabling faster model iteration across production releases

INTERNSHIP EXPERIENCE

Machine Learning Engineering Intern

HyperVerge

Aug 2024 – Jul 2025

Bengaluru, IN

- Fine-tuned domain-specific LLMs for KYC and fraud detection workflows, improving risk scoring precision and compliance automation for enterprise onboarding operations
- Evaluated LLMs on 1M+ records and deployed CV/NLP identity verification pipelines with CI/CD, model versioning, and monitoring, meeting production FAR/FRR performance targets
- Partnered with product and compliance teams to productionize document intelligence workflows, reducing manual review turnaround time by 25% for onboarding operations

Machine Learning Intern

EmendoAI

Jan 2024 – Aug 2024

California, USA (Remote)

- Developed and deployed 10+ generative AI microservices using AWS Lambda, API Gateway, and OpenTelemetry, enabling scalable client-facing deployments with end-to-end observability
- Implemented serverless retrieval-augmented generation (RAG) services and automated AI test pipelines, reducing production defects by 20% and improving release confidence
- Designed prompt templates, evaluation checks, and safety guardrails for client assistants, improving grounded response quality during pilot rollouts

Research Intern

Indian Institute of Technology Madras

Feb 2023 – Jul 2023

Chennai, IN

- Built an ML-based anti-collision system for unmanned surface vehicles (USVs), improving marine obstacle detection accuracy by 28% over baseline methods
- Implemented stereo-vision detection and object tracking on ESP32 edge hardware; validated reliability through wave-basin experiments and Unity simulations for real-time maritime navigation
- Documented experiment results and presented technical findings to research mentors, helping translate prototypes into actionable next-phase navigation improvements

PROJECTS

- Open-KYC** | *Next.js, OpenCV, TensorFlow, Tesseract, WebRTC* 2024
- Built an AI-powered KYC platform using OCR, facial authentication, and real-time liveness detection for digital onboarding
 - Achieved 98% verification accuracy and won 1st place at Standard Chartered Hackathon 2024 among 200+ teams
- Educative.AI** | *Python, FastAPI, React.js, LLMs, OCR* 2024
- Developed a full-stack AI learning assistant using OCR and fine-tuned LLMs for note extraction, MCQ generation, and Q&A
 - Supported 500+ learners with speech-to-text workflows and automated resource summarization features
- OneMed** | *Python, Next.js, LLMs, Pinecone, MongoDB, React Native* 2024
- Built an AI-enabled EHR platform with LLM summarization to reduce clinician documentation effort and improve care workflows
 - Reduced documentation time by 60%, enabled real-time emergency record sync across 15+ hospitals, and ranked Top 10 at MozoHack 2024
- PaperPilot** | *Scikit-learn, AWS, Python, React.js* 2023
- Implemented a TF-IDF + KNN recommendation engine to personalize academic paper discovery across research domains
 - Processed 5,000+ IEEE papers, achieved 85% relevance, and automated weekly recommendations for 1,000+ users
- AI-RoadGuard** | *TensorFlow, OpenCV, Flask, React.js* 2022
- Developed a CNN-based accident detection and emergency alerting platform for real-time road safety monitoring
 - Delivered 90%+ detection accuracy, reduced false alerts by 50%, improved emergency response time by 40%, and won Best Project at MLH MesoHack 2022

SKILLS

Programming: Python, JavaScript, Java, C/C++, SQL, HTML/CSS, R
AI/ML: LLMs, Generative AI, TensorFlow, PyTorch, Scikit-learn, OpenCV, Transformers, LangChain, LlamaIndex
Backend/Frontend: FastAPI, Flask, Node.js, React.js, Next.js, REST APIs, TailwindCSS
Cloud/DevOps: AWS, Azure, GCP, Docker, Kubernetes, Terraform, CI/CD, OpenTelemetry
Data: PostgreSQL, MongoDB, Redis, Pinecone, Data Warehousing, Lakehouse, Pandas, NumPy

PUBLICATIONS

Structured Relevance Assessment for Robust Retrieval-Augmented Language Models
ITC4SD 2025, Springer Nature — arXiv: 2507.21287

LEADERSHIP AND ACHIEVEMENTS

Hackathons & Awards

- 1st place at Standard Chartered Hackathon 2024 for OpenKYC, an AI-powered KYC verification solution
- 2nd place at Hack Nova 2024 with Educative.AI, selected to represent at Innverve 2024, AIT Pune
- Best Project Award in Open Innovation at MLH Meso Hack 2022 for AI-Roadguard accident detection system
- Top 10 finalist at MozoHack 2024 for OneMed, an AI-powered EHR platform for hospitals

Open Source & Technical Writing

- Contributed to Dify open-source project with 100,000+ GitHub stars
- Authored technical articles on AI and machine learning for DataX Journal on Medium platform

Research & Leadership

- Researcher at Next Tech Lab, Norman and McCarthy Labs (2021-2025), specializing in deep learning and computer vision applications
- Technical Director at Data Science Community SRM (2022-2023), organized DS Hack 2.0 and led technical workshops for 500+ students