

Abhinay Cherupally

Data Science — AI/ML Engineer — Full-Stack ML Developer

✉ cherupallyabhi@gmail.com | 📞 +91-6302323643 | 🌐 <https://github.com/Theagentvikram> | 🌐 [Linkedin.com/in/AbhiCherupally](https://www.linkedin.com/in/AbhiCherupally) |

📍 Hyderabad, Telangana, India | 📄 Portfolio

SUMMARY

AI/ML Engineer with 3+ years' experience developing scalable machine learning and generative AI solutions, including Retrieval-Augmented Generation (RAG), JSON-based data interchange, and data processing pipelines. Built an NLP fake news detection system at Intel improving accuracy by 15%. Published YOLOv8-DeepSORT retail theft research and engineered **ResuMatch** SaaS using FastAPI, LangChain, vector databases, and NoSQL.

EDUCATION

Bachelor's in Artificial Intelligence & Machine Learning | Sreenidhi Institute of Science and Technology (SNIST)

Dec 2021 - Aug 2025 · Hyderabad, India

Top 5% of department

- Applied machine learning frameworks (TensorFlow, PyTorch, Keras) to optimize models, boosting accuracy by 20% in capstone projects.
- Built GenAI and LLM-driven applications using Transformers and LangChain, showcasing hands-on expertise in generative AI.

WORK EXPERIENCE

Machine Learning Engineer | Syncd Analytics

Aug 2025 - Present · Freelance, United States

- Contributed to the design and deployment of an AI-powered Quality Control platform for appraisal document processing, helping achieve 60-80% reduction in manual review time.
- Developed microservices using Flask REST APIs with JSON payloads and integrated Next.js frontend components, working with ChromaDB for vector storage and implementing Dockerized deployments on AWS infrastructure.
- Built ML pipelines integrating OpenAI GPT-4 and AWS models, implementing DSPy for document parsing workflows, QC rule verification logic, and multi-model fallback mechanisms.
- Implemented authentication flows using JWT, created real-time dashboard components for monitoring, optimized parallel document processing and data processing pipelines, and ensured compliance with MISMO/FHA/UAD standards.

Machine Learning Intern | Intel Corporation

May 2023 - Jul 2023 · Remote

- Built comprehensive fake news detection system using advanced NLP pipelines, RAG techniques, NoSQL databases, and classification models, demonstrating strong problem-solving skills to address complex data challenges
- Implemented end-to-end ML workflows from CSV-based data ingestion and preprocessing to model evaluation and deployment on GCP infrastructure, applying problem-solving techniques to optimize performance
- Collaborated with cross-functional teams on real-world AI applications with production-level requirements
- Achieved high accuracy in text classification using transformer-based models and traditional ML techniques

Founder & Lead | Techabhee

Feb 2020 - Dec 2023 · Hyderabad, India

- Founded and led **Techabhee**, a technology blogging brand focused on laptops, mobiles, and consumer tech reviews, consistently publishing content for nearly 2 years.
- Grew the blog to over **100,000+ visitors** through high-quality, SEO-optimized content that ranked for multiple tech review keywords on Google.
- Executed **cold email campaigns** and **Meta (Facebook/Instagram) ads** to drive targeted traffic and build a recurring reader base.
- Implemented and managed **affiliate marketing** workflows, integrating product links and tracking performance to monetize content.
- Used tools such as **MailChimp**, **SendSense**, **Google Search Console**, **Google Analytics**, **Jetpack**, and **WordPress** to manage campaigns, monitor SEO performance, and optimize site engagement.

PROJECTS

Jobeasy - AI-powered Hiring Assistant | Personal Project

Mar 2025 - Present

- Built JobEasy as an end-to-end full-stack SaaS platform for job seekers, owning architecture, backend services, frontend UI, and cloud deployment.
- Implemented an AI-powered ATS scanner and My Resumes module that evaluate resumes against job descriptions and help users manage multiple tailored versions.
- Developed a Career Desk and Job Tracker to centralize job opportunities, track application stages, and maintain notes, streamlining the job search workflow.
- Created an in-app AI assistant for generating and refining resumes, cover letters, and outreach messages, along with an AutoApply feature that automates applications on supported platforms.

Nexus – Second Brain | Domain Project

Jan 2025 – Present

- Orchestrated an end-to-end RAG-powered “Second Brain” knowledge management framework using Large Language Models to automate document categorization and deliver context-aware semantic search with over 95% retrieval relevance on benchmarked queries.
- Architected a scalable Python backend pipeline for data ingestion, vector database integration (ChromaDB), AI-driven tagging, and relationship mapping, enabling graph-based knowledge workflows that reduced manual data organization time by 60%.
- Engineered a high-performance context retrieval engine combining LLMs, Transformers, and dynamic querying to synthesize multi-source datasets into daily AI-generated briefings, boosting user task completion speed by 30% in internal evaluations.

MeetNote | SaaS Project

Aug 2025 – Present

- Developed MeetNote as a cross-platform solution, extending it from a SaaS Chrome extension into a native Swift-based macOS app for seamless desktop meeting capture and note-taking.
- Implemented real-time audio capture and transcription pipelines to generate accurate, searchable meeting transcripts and structured notes with minimal user intervention.
- Built an automated summarization and highlight-generation workflow that condenses long meetings into concise action items, decisions, and key takeaways to speed up post-meeting review.
- Redesigned the architecture and tech stack to support both browser and desktop clients, exposing robust APIs for syncing meetings, transcripts, and summaries across platforms.

PUBLICATIONS

Retail Theft Detection | IJSREM Journal

<https://ijsrem.com/download/retail-theft-detection-system-using-yolov8-and-deepsort/>

- Developed real-time surveillance system for shoplifting behavior detection using computer vision
- Implemented object tracking algorithms with high accuracy in retail environment scenarios

SKILLS

- **Programming Languages:** Python, JavaScript, Java, C/C++, SQL, R
- **ML/AI Frameworks:** TensorFlow, PyTorch, Scikit-learn, Keras, YOLOv8, OpenCV, DsPy, NLP, Vision Models
- **GenAI & LLM Tools:** LangChain, Transformers, Ollama, Mistral, Fine-tuning, LoRa, QLoRa, LLMs
- **Web Development:** React.js, FastAPI, Node.js, HTML/CSS, RESTful APIs
- **Cloud & DevOps:** AWS (Lambda, Elastic Beanstalk, S3, EC2, Fargate), Docker, Git, Linux, OCI
- **Data Science Tools:** Pandas, NumPy, Matplotlib, Seaborn, Jupyter, Data extraction
- **Databases:** MySQL, PostgreSQL, MongoDB, ChromaDB, Supabase
- **Soft Skills:** Teamwork, Problem Solving, Communication, Critical Thinking

CERTIFICATIONS

Oracle Certified GenAI Professional | Oracle | Oct 2025

[Public Link to verify](#)

AWS Trained Data Engineer | AWS | Oct 2023

[Public Link to verify](#)

Intel Machine Learning Internship Certificate | Jul 2023

IBM SkillsBuild – Data Science Foundations | Jan 2023

Oracle Certified Data Science Professional | Oracle | Oct 2025

Public link to verify

ACHIEVEMENTS

- Achieved All India Ranks of 4800 and 6000 in GATE 2025 and GATE 2024 respectively, by applying advanced data science and artificial intelligence methodologies, proving high-level analytical and technical proficiency.
 - Earned the PMKVY Certificate of Skill Competency as an AI Machine Learning Engineer after completing 480 hours of intensive training at the Engineering Staff College of India (ESCI), Hyderabad, equipping with industry-relevant skills
-