

Atharva Gadad

+1 (281) 935-6240 | gadadatharva123@gmail.com | Houston, TX

SUMMARY

Applied AI professional with strong fundamentals and proven experience in Artificial Intelligence and Machine Learning projects in fast-paced startup environments, demonstrating rapid career progression from intern to Senior Data Scientist within 2 years. Expertise in designing and building scalable data pipelines and distributed systems for large-scale data processing using PySpark, Databricks, and Apache Spark on AWS EMR. Strategic thinker with excellent communication skills and technical expertise in AI models and LLMs (ChatGPT, Llama2, Claude), cloud computing platforms (AWS, Azure), and container technologies (Docker, Kubernetes). Proven track record of cross-functional collaboration, client presentations, and partnership management with AWS. Skilled in data analysis, DevOps practices, software development, and working with data science tools. Motivated problem solver with published research and hands-on experience developing production-grade AI/ML applications, including RAG-based AI systems, AI-driven OCR solutions, and natural language AI agents for automation.

WORK EXPERIENCE

Senior Data Scientist (formerly Data Scientist and Data Science Intern)

Aug 2023 – Present

Petrabytes

Houston, TX

- Promoted to Data Scientist in June 2024 and Senior Data Scientist in September 2025, demonstrating ability to lead and reliability in delivering results in a fast-paced startup environment
- Represented the company at industry conferences presenting AI/ML products and solutions, and participated in networking events to build strategic partnerships
- Collaborated directly with AWS partnership team to architect and implement cloud-based AI solutions, leveraging Bedrock, SageMaker, and other AWS services
- Presented technical solutions and product demonstrations to prospective clients, contributing to business development efforts
- Developed an OSDU Manifest Validation Engine to validate manifests against OSDU Schema rules, enhancing data integrity, compliance practices, and platform compliance
- Worked with DAS and DTS data to extract information within specific timestamp ranges and link related data across multiple files
- Worked on the development of AI Language Agents using AWS Bedrock Agents, LangGraph and Ollama for intelligent automation, collaborating with cross-functional team members
- Built natural language AI-powered RDBMS (SQL) agents for database querying, improving customer experience through intuitive interfaces and optimizing database design
- Developed LLM RAG based domain specific conversational AI agents using various AI models - ChatGPT, Llama, Claude, with iterative experiments to improve performance
- Developed a natural language querying application using LLM AI agents to orchestrate tasks, query databases, and invoke pipelines, streamlining workflow automation
- Designed and built scalable Databricks data pipelines for large-scale distributed data processing, optimizing PySpark and Delta Lake code on AWS EMR to handle high data volumes with improved performance and efficiency
- Developed ETL pipelines using Apache Spark and PySpark for processing and transforming large datasets in distributed computing environments
- Implemented DevOps practices and continuous improvement processes for AI/ML application deployment and monitoring
- Utilized Jupyter Notebook for data science experimentation, model development, and analysis workflows
- Worked with Docker containers and Kubernetes (EKS) for containerized application deployment on cloud platforms
- Worked on AI-based OCR applications, time series forecasting, data ingestion, development of Flask APIs for various AI applications
- Developed a semi-structured data parsing application leveraging AI and LLMs to extract data sections and parse header information, contributing to product engineering efforts
- Provided mentorship to interns and contributed to team building initiatives, fostering ethical practices, technical growth, and continuous learning culture

Research Assistant

Dec 2022 – Dec 2023

Center for Pattern Recognition and Machine Learning, PESU

Bangalore, India

- Worked on the development of an AI-powered Computer Vision and Machine Learning based solution for diagnosis and stage detection of Retinopathy of Prematurity, an ocular disease in the healthcare domain
- Developed multiple AI CNN Models using fundus images obtained from Narayana Nethralaya Hospital, Bangalore for diagnosis with accuracies between 87% and 93%

AI/ML Summer Intern

Jun 2023 – Aug 2023

Pronisi

Bangalore, India

- Worked on AI Prompt Engineering using ChatGPT and development of an AI Recommendation Engine for budding sports players based on numerical data points generated from user videos using Python
- Developed a fully overhauled version of the codebase with full ChatGPT AI integration, enabling a more engaging user experience and 75% reduction in runtimes

Software Engineering Intern

Jun 2022 – Jul 2022

Kognosdata

Bangalore, India

- Worked on Data Analysis using Python libraries, development of front-end using Vaadin (a Java web application development platform), and Java back-end programming for Petroleum Engineering software products
- Performed data analysis on Excel files and other data sources for information systems

SKILLS

Programming Languages: C, C++, Java, Python, HTML, CSS, JavaScript, MySQL

Cloud Computing: AWS (SageMaker, Bedrock, EC2, RDS, S3, EMR, EKS), Azure (Blob Storage)

Big Data & Distributed Systems: Apache Spark, PySpark, Databricks, Delta Lake, Hadoop, AWS EMR, Large-scale data processing

AI/ML Technologies: ChatGPT, Llama2, Claude, LangChain, TensorFlow, Keras, Hugging Face, AI Agents, RAG Systems

Data Science Tools: Jupyter Notebook, Data Analysis, Data Visualization

DevOps & Containers: Docker, Kubernetes, DevOps practices, Continuous Integration, Continuous Improvement

Other Technologies: Tkinter, MySQLConnector, PyTesseract, MongoDB, Express, React, NodeJS, Vaadin, Linux, Excel, Information Systems

PUBLICATIONS

Accelerating Unstructured and Semi-Structured Data Insights with Generative AI - EAGE Digital 2025

Authors: Sashi Gunturu, Atharva Gadad, Yuriy Gubanov, Dmitriy Tishechkin

Deep Learning for Classifying Stages of Retinopathy of Prematurity - AMATHE 2024

Authors: Ayush Satish Govind, Atharva Gadad, Dhruv Jyoti Garodia, Ayush Gupta, Anand Vinekar, Gowri Srinivasa

A Pugnacious Comparative Study of Data Analysis Techniques for Wine Quality Prediction - I2CT 2024

Authors: Dhruv Jyoti Garodia, Atharva Gadad, Aaditya Vikram

EDUCATION

PES University

Bangalore, India

BS in Computer Science and Engineering (Information Systems); GPA: 9.81/10

2020 – 2024

- Graduated in the top 1% of the class (ranked 9th amongst 1,320 students)
- Awarded the CNR Scholarship for consistent academic excellence, ranking in the top 1% of the student body every semester, covering 40% of tuition fees

I am a US Citizen and will not require visa sponsorship. I am willing to relocate.