Birbal Kumar
Data Scientist

Jaipur, Rajasthan | +91 7220972090 | Gmail: Birbalk99@gmail.com
LinkedIn: www.linkedin.com/in/birbalk99 | GitHub: https://github.com/Birbalk99

SUMMARY

Data Scientist with 3.6+ years of experience in AI, Generative AI, LLMs and agentic workflows, grounded in a strong Computer Science background. I have led end-to-end development of LLM-powered chatbots, autonomous agents and ML systems, including designing HLDs and LLDs for scalable production-grade solutions.
Skilled in multi-step reasoning, tool use and memory-based agents using models like GPT and Llama. Proficient in Python, SQL, Django, Flask, AWS, with hands-on experience in ML pipelines, vector search and LLM fine-tuning. Passionate about building impactful, AI-driven products that solve real-world problems.

CORE COMPETENCIES

- Artificial Intelligence, Machine Learning, Deep Learning
- NLP, Computer Vision, Generative AI, LLMs, Prompt Engineering
- Retrieval-Augmented Generation (RAG), Agentic AI Systems, Real-time AI Systems
- MLOps, LLMOps, Data Engineering, Model Deployment

TECHNICAL SKILLS

- **Programing & Scripting:** Python, SQL, Bash
- **ML & DL Libraries:** Scikit-learn, TensorFlow, PyTorch, Keras, XGBoost, LightGBM, Hugging Face, Transformers
- **Data Wrangling & Visualization:** Pandas, NumPy, Matplotlib, Seaborn, Plotly, OpenCV, PIL, PyTesseract
- **NLP & LLM Stack:** LangChain, Llama Index, Transformers, GPT-3.5/4/4o, Ollama, Sentence Transformers
- **Agent Frameworks:** CrewAI, LangGraph, AutoGen, SuperAGI
- **Computer Vision & OCR Tools:** YOLOv5–YOLOv12, OpenCV, OCR, Detectron2, Roboflow, MediaPipe
- **Frameworks & APIs:** Flask, FastAPI, Django, Celery, Redis, WebSocket, LangChain, LlamaIndex, Streamlit
- **Vector DBs & RAG Stack:** FAISS, Pinecone, ChromaDB, MongoDB Atlas
- **Cloud Platforms: AWS** (EC2, S3, Lambda, IAM), **GCP** (Vertex AI), **Azure**, Git, GitHub, Docker, Replit, Streamlit
- **IDEs & Platforms:** VS Code, Jupiter, Google Colab, Tableau, Excel
- **CI/CD & Monitoring:** GitLab CI/CD, GitHub Actions, SonarQube (Code Quality), Jenkins
- **OS Environments:** Ubuntu, MacOS, Windows

PROFESSIONAL EXPERIENCE

MANGAL   **Data Scientist**
Jaipur, Rajasthan, IN   Dec 2025 – Present

❖ **Metadrob Private Limited**

Led and developed an AI-powered voice and chat assistant leveraging Generative AI for seamless customer interactions over Twilio calls and real-time website chat. Integrated OpenAI GPT-4o mini and Llama 3 for intelligence response generation, significantly enhancing user engagement and support efficiency. Designed a multi-modal data ingestion pipeline to process and enable an LLM-driven Retrieval-Augmented Generation (RAG) system for context-aware responses. Developed high- performance vector search using FAISS and MongoDB, optimizing retrieval speed. Built a scalable, event-driven architecture using FastAPI, WebSocket, Celery, and Redis, ensuring low-latency AI interactions and parallel processing for real-time status tracking. Integrated AWS Polly & Google TTS for human-like speech synthesis, enhancing voice-based AI interactions. Consistently pushed the boundaries of Generative AI innovation, delivering impactful LLM-powered automation solutions.

❖ **Rams Creative Technologies Pvt. Ltd.**

Leading the AI team overseeing the design and development of innovative AI solutions using LLMs, RAG pipelines, YOLO, CrewAI and LangChain. Responsible for setting technical direction, mentoring team members, and delivering high-impact use cases across legal, infrastructure and manufacturing domains. Successfully delivered multiple AI-driven projects and POCs including road pothole detection, legal document assistants, metal sheet defect detection systems and Excel-integrated chatbots and graph generators leveraging Python, MongoDB and state-of-the-art AI/ML techniques.

PAI SOLUTIONS   Data Scientist
Delhi, IN   Mar 2023 – Dec 2024

Data Scientist at Pai Solutions, specializing in Generative AI, I led the development of an AI Support Chatbot that enhanced customer support efficiency. This project involved integrating advanced AI models, including OpenAI GPT-3.5-turbo and Llama and deploying on AWS for scalability. I also developed a dynamic content capture system for extracting images from PDFs and videos, integrated into our Flask application. Additionally, I contributed to various other Generative AI projects, continuously pushing the boundaries of AI innovation.

CAMP AUTOMATION PRIVATE LIMITED   Data Scientist
Bengaluru, Karnataka, IN   Mar 2023 – Dec 2024

Product Developer at Camp Automation Private Limited, I played a key role in enhancing product functionality and user experience. I designed and implemented an improved dashboard with a refined UI, making product monitoring and management more efficient. Leveraging machine learning techniques like K-Means and DBSCAN, I optimized the Facebook Ads Manager by enabling predictive modeling, automated audience segmentation, and sentiment analysis to improve ad performance. I utilized Python and library's, ensuring high-quality insights.

EDUCATION

- RAJASTHAN TECHNICAL UNIVERSITY   JAIPUR, RAJASTHAN, IN
  B.Tech in Computer Science | 64 %   Jun - 2021
- SECONDARY EDUCATION (BSEB)   SIWAN, BIHAR, IN
  Rana Pratap H/S cum inter collage | 63 %   May - 2015

CERTIFICATION

- Artificial Intelligence Engineer | Simplilearn | July 2023
  Certificate URL: https://success.simplilearn.com/df110e96-6ab3-4f2f-9f38-3cbb0078b9a1#acc.Xws57tyD
  Certificate ID: 78309242
- Data Science Certification | Simplilearn | Nov 2021
  Certificate URL: https://success.simplilearn.com/f595d641-9812-49a0-aec8-4e2f32591f68
  Certificate ID: 61935232
- Neo4j Certified Professional | Neo4j Graph Academy | May 2025
  Certificate URL: https://graphacademy.neo4j.com/c/7f55bca8-af44-4977-9fbd-ad7919cbdb6d
  Certificate ID: 7f55bca8-af44-4977-9fbd-ad7919cbdb6d
- Credly Badges | google Cloud, Mongo DB, Others
  Certificate URL: https://www.credly.com/users/birbalk99/badges

PROJECTS

- **Telephone Conversation Agent | Tech Stack:** Generative AI, MongoDB, AWS, FAISS, Fast API, Twilio, Celery

  Engineered an LLM powered voice assistant capable of real-time conversations over Twilio calls, leveraging multi-modal data sources for contextual responses. Integrated OpenAI GPT-4o for intelligent response generation and AWS Polly & Google TTS for human like, high fidelity speech synthesis. Designed a scalable event driven architecture using FastAPI + WebSocket, enabling low latency, bidirectional Twilio call streaming. Optimized asynchronous task execution with Celery & Redis, ensuring parallel processing of large datasets and real time status tracking. Architected a distributed vector storage system using FAISS + MongoDB, facilitating fast semantic retrieval across large knowledge bases.
  Key Skills: LLM, Generative AI, RAG, FAISS, LangChain, OpenAI GPT-4o, QLoRa, WebSocket, Twilio Stream API, AWS Polly, Prompt Engineering, MongoDB, Redis, Celery, Async Processing, Parallel Computing.
- **Knowledge Based Chatbot (RAG) | Tech Stack:** Python, LLM, Generative AI, Llama

Created an LLM powered chatbot to provide technicians with quick access to site manuals, significantly reducing troubleshooting time and enhancing on-site productivity. Enabled significant reductions in troubleshooting time and enhanced on-site productivity through intelligent query handling and context-aware responses.

Key Skills: LLM, Generative AI, LangChain, AWS Document DB, LLAMA, OLLAMA, Prompt Engineering.

- **Telco Feedback Sentiment Analysis | Tech Stack:** Python, Deep Learning

  Designed an NLP-based sentiment analysis system to analyses customer feedback, enabling actionable insights that drive targeted improvements in products and services.

  Key Skills: Sentiment Analysis, Text Pre-processing, Word Embeddings (Word2Vec, TF-IDF, etc.) Classification Models, XGBoost, Hyperparameter Tuning, Grid Search, Accuracy Evaluation.

- **Automatic Document Creation Agent | Tech Stack:** Python, LLM, LangChain, Llama

  Developed an AI-driven system designed to automate the creation of project documentation by processing diverse data inputs such as media files, meeting recordings, audio, text and other resources. The agent intelligently organizes and generates professional, comprehensive documents tailored to project requirements, significantly reducing manual effort and enhancing productivity in project management workflows.

  Key Skills: LLM, LangChain, MongoDB atlas, LLAMA 3.2, Prompt Engineering.

- **Advanced Image Processing and Analysis System | Tech Stack:** Python, OpenCV, PIL, NumPy, PyTesseract, Matplotlib

  Developed an advanced computer vision project that integrates image preprocessing, edge detection, color extraction and object identification using diverse machine learning and image processing techniques. The system leverages OpenCV and machine learning algorithms to automate tasks like background removal, text extraction, color analysis and contrast ratio assessment. Enhanced performance through multi-step pipeline optimization for edge cases and gradient detection, enabling accurate object segmentation and visual data analysis.

  Key Skills: Image Processing, Computer Vision, Color Detection, Edge Detection, Text Recognition, Data Visualization, Algorithm Optimization.

ADDITIONAL

LANGUAGES: English, Hindi

SOFT SKILLS:
- Problem Solving & Critical Thinking
- Leadership & Team Management
- Communication & Collaboration
- Strategic Thinking & Decision Making
- Time Management & Multitasking
- Creativity & Innovation