Avdeep Singh

|+91-7814869362 | avdeep2001@gmail.com| LinkedIn | GitHub

Professional Summary

AI Engineer with 2.5+ years of experience in Generative AI, LLMs, and RAG systems. My interest in AI sparked during college through ML projects and my first internship, where I realized data is the new gold. To accelerate learning in this fast-evolving field, I've pursued contract and freelance roles, gaining diverse hands-on exposure across multiple projects rather than long-term corporate assignments. Specialized in multilingual chatbots, agentic AI, and model fine-tuning, I've built solutions that automate processes, reduce costs, and improve efficiency using Python, TensorFlow, and cloud tools.

Professional EXPERIENCE

TechWise Digital Pvt Ltd | Artificial Intelligence Developer (Contract)   Jul 2025 – Present
(Bengaluru, Karnataka, India)

- Served as core developer for an Agentic AI product, building the entire backend system—including deployment, case management, and integration—to automate recruiter workflows.
- Designed 4 autonomous agents (1 supervisor and 3 sub-agents) with 10 custom tools, using LangChain, LangGraph, LangSmith, Qdrant DB for RAG, MongoDB for case management, and Google Gemini LLM.
- Enabled features to enhance job descriptions (JDs), match and score resumes against JDs, reducing pre-screening time from days to minutes through real-time agent collaboration.

Metadee | AI Engineer (Contract)   Jan 2025 - Jun 2025
(Remote)

- Developed a RAG-based chatbot for religious texts, delivering 90% accurate responses (per domain expert validation) across Hindi, Sanskrit, English, and Hinglish with seamless context switching.
- Fine-tuned a transformer-based TTS model for Sanskrit using 10+ hours of curated audio data, addressing phonetic challenges to achieve 90% pronunciation accuracy and 85% audio quality for audiobooks.
- Managed dataset preprocessing, including custom tokenization for ancient languages, to support high-fidelity model training.

Webchill | AI Engineer (Contract)   Oct 2024 - Dec 2024
(Remote)

- Implemented an LLM-powered customer support chatbot, cutting resolution time by 60% and boosting satisfaction through API integrations for tailored responses.
- Adapted generative models for automated content, creating personalized user experiences while ensuring compatibility with existing infrastructure.
- Handled deployment and testing to minimize integration issues, resulting in a seamless rollout.

Briskon Technologies Pvt Ltd | Machine Learning and Artificial Intelligence Engineer   Jan 2024 – Oct 2024
(Bengaluru, Karnataka, India)

- Created a transformer-based chatbot with GPT and Gemini LLMs, incorporating RAG for handling vast datasets from B2B and B2C clients, automating 30% of inquiries, enabling appointment booking, and increasing self-service by 40%.
- Engineered an AI SEO content writer using Gemini, streamlining a multi-step pipeline (topic input, keyword optimization, competitor analysis, content skeleton, and final generation) that significantly reduced creation time and costs, eliminating the need for a dedicated content writing team.
- Optimized models with techniques like quantization to enhance speed and relevance for high-volume SEO services.

Freelancer | Data Scientist (Contract)   Jul 2023 – Dec 2023
(Remote)

- Built a LangChain-based chatbot with RAG for client-specific FAQ handling and query resolution, focusing on straightforward user interactions without advanced features like scheduling.
- Refined prompts and integrated external APIs to improve response accuracy and adaptability for basic use cases.

TopGrep Tech | Machine Learning Developer   Jan 2023 – Jun 2023
(Bengaluru, Karnataka, India)

- Developed an 85% accurate sales forecasting model with time-series methods, improving pipeline efficiency by 30% and team productivity by 50% via interactive dashboards.

KEY PROJECTS

AI Recruitment Screening System   *Jul 2025 - Present*
- Complete product architecture with supervisor-agent orchestration pattern
- Automated JD enhancement, resume parsing, and intelligent matching algorithms
- Built for scale: Handles 10+ resumes per JD with parallel processing

Multilingual Religious Knowledge Base   *Jan 2025 - Jun 2025*
- Sanskrit TTS model with custom audio processing pipeline
- Cross-lingual semantic search across ancient texts
- Context-aware responses validated by domain experts

AI-Powered Content Writer with Generative AI (Gemini)   *Jul 2024 - Oct 2024*
- Launched a AI tool using LangChain and Gemini LLM for SEO-optimized content, incorporating autonomous agents for research and multi-source data gathering to achieve high-quality outputs (95% internal score).
- Included features for customizable parameters, competitor analysis, and graphic suggestions, optimizing for blog and website creation.

RAG Based Customer Support Chatbot   *Jan 2024 - Apr 2024*
- Deployed a GPT and Gemini-powered RAG system with LangChain, automating 30% of inquiries and reducing support costs by 35% through reranking and search algorithms.
- Added agent functionalities for proactive assistance, scheduling, and recommendations, reaching 92% user satisfaction via surveys.

TECHNICAL SKILLS

- **Programming Languages:** Python, SQL, HTML, CSS.
- **Machine Learning & AI Frameworks:** TensorFlow, PyTorch, Scikit-learn, Keras, Hugging Face Transformers, LangChain, OpenAI API, LangGraph, CrewAI
- **Machine Learning Techniques:** Classification, Regression, Clustering, Predictive Modeling, Time Series Forecasting, Model Evaluation, Cross-validation, Hyperparameter Tuning, Feature Selection.
- **Deep Learning & Neural Networks:** Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), Long Short-Term Memory (LSTM), Transformers, Transfer Learning, Fine-tuning, Model Optimization.
- **Natural Language Processing:** Large Language Models (LLMs), GPT, Gemini, Text-to-Speech (TTS), Sentiment Analysis, Named Entity Recognition (NER), Text Classification, Text Generation, Topic, Language Translation.
- ***Generative AI & LLM Operations:*** *Retrieval-Augmented Generation (RAG), Prompt Engineering, Prompt Optimization, Conversational AI, Chatbot Development, Multi-Agent Systems, AI Agent Development, LLM Fine-tuning, Content Generation.*
- ***Cloud Platforms & MLOps:***
  AWS: SageMaker, EC2, S3, Lambda, SageMaker Endpoints, SageMaker Training Jobs, CloudWatch
  Google Cloud Platform: Vertex AI, Cloud ML Engine, Cloud Storage, BigQuery, Cloud Functions, AI Platform
  MLOps: Model Versioning, Model Monitoring, Automated Retraining, CI/CD for ML, Model Registry.
- ***Development & Deployment Tools:*** *Git, GitHub, Docker, FastAPI, API Integration, Containerization, Version Control.*

Certification
Advanced MACHINE LEARNING | IIT Kanpur   *2022*

EDUCATION
Guru Nanak Dev Engineering College   **2020 - 2023**
B.tech in Information Technology (CGPA 8.3)
Guru Nanak Dev Polytechnic College   **2017 - 2020**
Diploma in Computer Science (73%)