[Your Name]

Generative AI Engineer
[Your Email] | [Your LinkedIn] | [Your GitHub]

Objective

Energetic and innovative Generative AI Engineer eager to leverage a profound background in AI, specializing in vector databases, LLM frameworks, and cloud deployments. Aims to apply technical skills and creativity in developing state-of-the-art AI solutions within a dynamic team setting.

Education

Bachelor of Science in Computer Science
[University Name], [Graduation Year]
Relevant Coursework: Machine Learning, Artificial Intelligence, Data Structures and Algorithms, Natural Language Processing

Skills

- **Programming Languages**: Python
- **Frameworks & Tools**: TensorFlow, PyTorch, LangChain, LlamaIndex,Streamlit,Flask
- **Generative AI Technologies**: Open-source and paid LLM models (Llama2, Mistral,OpenAI,Google Gemini Pro)
- **Vector Databases**: ChromaDB, Pinecone
- **Database Management**: Experience with DataStax Cassandra DB in production environments, Mysql, Mongodb
- **Deployment Platforms**: AWS Bedrock, AWS (EC2, Lambda), Azure Functions, Hugging Face Spaces
- **AI/ML Techniques**: Fine-tuning with custom data, vector embedding, NLP, neural network optimization,MLOPS, Dockers,Kubenetes
- **Soft Skills**: Analytical thinking, problem-solving, teamwork, effective communication

Professional Experience

Roles and Responsibilities

- **Managed and optimized vector databases** such as ChromaDB and Pinecone, enhancing AI applications' data retrieval capabilities.

- **Handled DataStax Cassandra DB in production environments**, ensuring optimal performance and reliability for AI-driven applications.
- **Developed scalable AI solutions** using LangChain and LlamaIndex frameworks, demonstrating expertise in generative AI technologies.
- **Implemented and fine-tuned both open-source and paid LLM models**, customizing solutions to meet specific project requirements and performance goals.
- **Leveraged AWS Bedrock for deploying AI models**, utilizing cloud services to ensure scalability and reliability of AI applications.

Projects

ATS Resume LLM App

- **Overview**: Developed an application optimizing resumes for ATS using OpenAI APIs, enhancing match rates for job applications.
- **Technologies**: OpenAI APIs, Python, AWS, Google Gemini Pro.
- **Outcome**: Achieved a 50% improvement in ATS match rates, increasing users' chances of securing interviews.

Text to SQL LLM App using Llama2

- **Overview**: Created an app to convert natural language queries into SQL commands using Llama2, improving database accessibility for non-technical users.
- **Technologies**: Llama2, Python, AWS Bedrock.
- **Outcome**: Reduced query formulation time by over 60%, enhancing productivity for users.

Advanced Q&A Chatbots with DataStax Databases and Vector Embedding

- **Overview**: Engineered chatbots that provide accurate, context-aware answers by integrating DataStax databases with vector embedding techniques.
- **Technologies**: DataStax Databases, vector embedding algorithms, Hugging Face Spaces.
- **Outcome**: Achieved a 40% reduction in the need for human intervention in customer service inquiries, improving response accuracy and user satisfaction.

Achievements

- **Innovative AI Solutions**: Successfully developed and deployed cutting-edge AI applications, demonstrating a strong impact in enhancing user engagement and operational efficiency.
- **Technical Leadership**: Led teams and projects to pioneer the use of generative AI technologies in practical, real-world applications.

Certifications

- Certified TensorFlow Developer (TensorFlow)
- Machine Learning with Python (Coursera)