JOHN DOE

Senior AI/ML Engineer | LangChain Specialist | RAG Expert

CONTACT INFORMATION

Address: 123 Tech Avenue, New York, NY 10001, USA
Phone: +1 555-123-4567 (Work) | +1 555-987-6543 (Mobile)
Email: john.doe@example.com
LinkedIn: linkedin.com/in/johndoe | GitHub: github.com/johndoe

PROFESSIONAL SUMMARY

Innovative AI/ML Engineer with 6+ years of experience designing production-grade ML systems, specializing in NLP, LLMs, and RAG pipelines. Proven track record of transforming complex AI research into scalable enterprise solutions with measurable business impact.

TECHNICAL EXPERTISE

AI/ML: LangChain, LlamaIndex, Hugging Face, PyTorch, TensorFlow, Scikit-learn
LLMs: GPT-4, Claude, Gemini, Llama 2/3, BERT, SBERT, T5, Mistral
Vector DBs: FAISS, Pinecone, ChromaDB, Weaviate, Milvus
Languages: Python, JavaScript, SQL, Bash, Go
Cloud: AWS (SageMaker, Bedrock), GCP (Vertex AI), Azure OpenAI
Tools: Docker, Kubernetes, Git, MLflow, FastAPI, Streamlit

PROFESSIONAL EXPERIENCE

Senior AI Engineer | Example Tech Corporation   **Jan 2023 - Present**

  - Architected enterprise RAG systems processing 150,000+ documents daily
  - Developed LangChain systems reducing customer query response time by 65%
  - Built semantic search using FAISS with 96% retrieval accuracy at 50ms latency
  - Led team of 5 engineers establishing ML best practices
  - Implemented A/B testing for LLM prompt optimization

ML Engineer | Innovation Labs Inc., San Francisco   **Jun 2020 - Dec 2022**

 - Automated insurance claim classification using BERT with 94% accuracy
 - Created NLP pipeline for sentiment analysis on 2M+ customer reviews
 - Deployed ML models on AWS SageMaker with automated retraining
 - Built document understanding system reducing manual review by 70%
 - Reduced model inference costs by 45% through optimization

Data Scientist | DataDriven Analytics, Boston   **Jul 2018 - May 2020**

 - Developed predictive models for customer churn reducing attrition by 18%
 - Built recommendation system increasing cross-sell conversion by 23%
 - Created dashboards using Python, SQL, and Tableau for reporting

KEY PROJECTS

Enterprise Document Intelligence Platform
  Built RAG system using LangChain, FAISS, GPT-4 for querying 10M+ documents

Multi-Modal AI Assistant
  Conversational AI with text, image, document understanding for 50K+ users

Semantic Code Search Engine
  SBERT-based code search enabling 5x faster code discovery

Automated Report Generation
  LLM-powered report generation reducing analyst workload by 60%

EDUCATION

Master of Science in Computer Science   **2016 - 2018**
Stanford University | ML & NLP Specialization | GPA: 3.9/4.0

Bachelor of Technology in Computer Engineering   **2012 - 2016**
Example Institute of Technology | First Class with Distinction

CERTIFICATIONS

- AWS Certified Machine Learning - Specialty (2023)
- Google Cloud Professional ML Engineer (2023)
- DeepLearning.AI - LangChain for LLM Application Development (2023)
- DeepLearning.AI - Building Systems with ChatGPT (2023)
- Coursera - NLP Specialization (2022)
- AWS Solutions Architect - Associate (2021)

PUBLICATIONS & TALKS

- Scaling RAG Systems for Enterprise - AI Engineering Summit 2024
- Efficient Vector Search at Scale - Medium (50K+ views)
- Building Production LLM Applications - PyData Conference 2023
- From BERT to GPT: A Practical Guide - Towards Data Science (100K+ readers)

OPEN SOURCE CONTRIBUTIONS

- LangChain - Custom document loaders and retriever implementations
- Hugging Face Transformers - Bug fixes and documentation
- FAISS - Performance optimization patches
- Personal Projects: 15+ repos with 2,000+ GitHub stars

AWARDS & RECOGNITION

- Best AI Innovation Award - Example Tech Corporation (2024)
- Top Contributor - LangChain Community (2023)
- Hackathon Winner - AI for Good Challenge (2022)
- Employee of the Year - Innovation Labs (2021)
- Graduate Research Fellowship - Stanford (2017)

LANGUAGES

English (Native) | Hindi (Native) | Spanish (Conversational)

INTERESTS

AI Research, Open Source, Technical Writing, Hiking, Chess, Photography

REFERENCES

Available upon request