

Gaurav Ghadge

AI Engineer | Immediate Joiner

📍 Nagpur, Maharashtra · 📞 8668802792 · ✉ gghadge225@gmail.com · 🔗 [LinkedIn](#) · 🔗 [GitHub](#) · 🌐 [Portfolio](#)

Professional Summary

Senior GenAI Engineer & AI Agent Architect with 3+ years of experience developing production-grade Large Language Models, autonomous AI agents, and enterprise-scale multimodal artificial intelligence systems. Expert in machine learning, deep learning, neural networks, and computer vision. Proven track record deploying scalable AI solutions using OpenAI GPT-4, Claude, LangChain, and Hugging Face Transformers with 98%+ model accuracy, reducing operational costs by 95% and processing 1M+ documents annually across government, healthcare, and fintech domains.

Core Technical Skills

AI/ML Technologies	Development & Deployment
Programming Languages: Python, SQL, JavaScript, TypeScript, R, Java, C++	Cloud Platforms: AWS SageMaker, Google Cloud Vertex AI, Azure Machine Learning
Machine Learning: Supervised Learning, Unsupervised Learning, Reinforcement Learning	MLOps & DevOps: MLflow, Kubeflow, Docker, Kubernetes, CI/CD, Jenkins
Deep Learning: Neural Networks, CNN, RNN, LSTM, GRU, Transformer Architecture	AutoML Platforms: H2O.ai, AutoKeras, Google AutoML, Azure AutoML
Large Language Models: OpenAI GPT-4, GPT-3.5, Claude, Gemini Pro, LLaMA, Mistral AI	Model Monitoring: Weights & Biases, Neptune, TensorBoard, Evidently AI
Generative AI: Prompt Engineering, Fine-tuning, RLHF, RAG Architecture, Chain-of-Thought	Explainable AI: SHAP, LIME, Captum, Alibi, What-If Tool
AI Frameworks: TensorFlow, PyTorch, Keras, Scikit-Learn, XGBoost, LightGBM, CatBoost	Edge AI: TensorFlow Lite, ONNX, OpenVINO, NVIDIA Jetson
GenAI Tools: LangChain, LangGraph, AutoGen, CrewAI, Semantic Kernel, Llama-Index	Databases: PostgreSQL, MongoDB, Redis, Apache Cassandra, Neo4j
Vector Databases: FAISS, Chroma, Pinecone, Weaviate, Qdrant, Elasticsearch	Big Data: Apache Spark, Kafka, Hadoop, ETL Pipelines, Data Warehousing
Computer Vision: OpenCV, YOLO, Detectron2, Segment Anything, Stable Diffusion	Business Intelligence: Power BI, Tableau, Looker, Grafana
NLP Technologies: NLTK, SpaCy, Transformers, Text Mining, Sentiment Analysis	Version Control: Git, GitHub, GitLab, Bitbucket, DVC

Professional Experience

Senior Data Scientist | Datanet Solutions | May 2020 - Present

- Developed and deployed 15+ machine learning models and generative AI applications solving complex business problems across government, healthcare, transportation, and fintech domains using supervised learning, unsupervised learning, and deep learning techniques
- Led end-to-end artificial intelligence project lifecycle from data preprocessing to cloud deployment using AWS, GCP, and Azure platforms managing \$5M+ project portfolio with measurable ROI
- Implemented MLOps best practices including model monitoring, A/B testing, and automated retraining pipelines resulting in 40% faster project delivery and improved model performance
- Mentored 5+ junior data scientists and machine learning engineers establishing AI/ML development standards and code quality practices

Key Projects

Enterprise Multi-Agent AI System for Document Intelligence

Technologies: LangGraph, Multi-Agent Architecture, GPT-4, Computer Vision, OCR, FAISS, Deep Learning

- Architected autonomous multi-agent system using neural networks and computer vision for processing 15+ government document types achieving 98% classification accuracy through synthetic data generation and transfer learning
- Deployed production-ready document intelligence pipeline with natural language processing capabilities handling 50,000+ documents monthly with real-time text extraction and data mining
- Reduced manual processing time by 95% and operational costs by \$2M annually through intelligent automation and machine learning optimization

Advanced Multimodal AI Assistant with Voice and Vision

Technologies: GPT-4 Vision, Whisper API, DALL-E, Computer Vision, Speech Recognition, WebRTC

- Developed conversational AI system combining speech-to-text, computer vision, and natural language generation capabilities with real-time multimodal interactions using transformer architecture
- Integrated OpenAI Whisper for automatic speech recognition, GPT-4 Vision for image analysis, and DALL-E for image generation creating comprehensive AI assistant with 99.9% uptime
- Built scalable web application with WebRTC-based interface supporting voice commands, image uploads, and live code execution using deep learning models

Autonomous RAG-Powered Knowledge Management System

Technologies: LangChain, GPT-4, FAISS, Semantic Search, Vector Embeddings, Retrieval Augmented Generation

- Engineered enterprise-grade RAG architecture with semantic search and vector embeddings processing 10TB+ organizational knowledge base with sub-second query response times using transformer models
- Implemented advanced prompt engineering, context window optimization, and semantic chunking achieving 92% answer accuracy across 50+ business domains through fine-tuning and reinforcement learning
- Created self-healing knowledge system with automated content updates, version control, and machine learning-based quality assurance reducing knowledge management overhead by 80%

Real-Time Predictive Analytics Platform for Smart Transportation

Technologies: Apache Kafka, Spark Streaming, XGBoost, Time Series Forecasting, Feature Engineering

- Built end-to-end machine learning pipeline processing 5+ years of IoT sensor data with 10M+ daily events for predictive demand forecasting using ensemble methods and statistical modeling
- Achieved 97% prediction accuracy with RMSE of 0.05 through advanced feature engineering, hyperparameter tuning, and gradient boosting algorithms with automated model retraining
- Deployed scalable microservices architecture on AWS with auto-scaling, load balancing, and real-time model serving handling 1000+ concurrent requests

Production-Grade Sentiment Analysis Engine

Technologies: BERT, RoBERTa, Transformers, Apache Kafka, Redis, Natural Language Processing

- Developed distributed sentiment analysis system processing 100K+ social media posts daily using pre-trained transformer models and fine-tuning techniques achieving 86% recall score
- Implemented real-time streaming analytics with Apache Kafka and Redis caching achieving sub-100ms latency for brand monitoring and customer feedback analysis
- Created comprehensive analytics dashboard with predictive insights and anomaly detection enabling businesses to track sentiment trends and automate crisis response

Autonomous Code Generation Agent with Self-Healing

Technologies: GPT-4, Code Interpreter, GitHub API, Automated Testing, CI/CD, Reinforcement Learning

- Built intelligent code generation system using large language models capable of writing, testing, and deploying production-ready applications with minimal human intervention
- Integrated automated testing frameworks, code review agents, and continuous integration pipelines reducing development time by 60% and bug rates by 40%

- Created self-healing architecture with automatic error detection, code correction, and rollback mechanisms using reinforcement learning ensuring 99.5% system reliability

Education & Certifications

Master of Science in Computer Science | Dr. Ambedkar College, Nagpur | 2020 | **GPA: 3.8/4.0**

Relevant Coursework: Advanced Machine Learning, Deep Learning, Computer Vision, Natural Language Processing, Statistical Methods, Database Systems

Bachelor of Science in Computer Science | Dr. Ambedkar College, Nagpur | 2017 | **GPA: 3.7/4.0**

Relevant Coursework: Data Structures, Algorithms, Software Engineering, Database Management, Web Technologies

Professional Certifications:

- AWS Certified Machine Learning Specialty | AWS Certified Solutions Architect
- Google Cloud Professional ML Engineer | Google Cloud AI Engineer
- Microsoft Azure AI Engineer Associate | Azure Data Scientist Associate
- Deep Learning Specialization - Coursera (Stanford) | Machine Learning - Andrew Ng
- Advanced MLOps Engineering - Coursera | Kubernetes Application Developer (CKAD)
- TensorFlow Developer Certificate | PyTorch Certification
- LangChain for LLM Application Development | Hugging Face Transformers Course
- Advanced Data Science and Analytics - Sololearn

Key Achievements & Recognition

- 1st Prize Winner - National Level Technical Debate Competition demonstrating communication skills
- 2x Best Employee of Quarter at Datanet Techno Solutions for exceptional performance
- Published 3 peer-reviewed technical papers on Generative AI and Computer Vision
- Keynote speaker at 5+ AI/ML conferences and industry workshops
- Open source contributor with 1000+ GitHub stars and active community engagement
- Patent applications: 2 pending patents in artificial intelligence and machine learning domain
- Languages: English, Hindi, Marathi (Professional working proficiency)

Technical Expertise Summary

- **Algorithm Development:** Designed and implemented 20+ machine learning algorithms from scratch
- **Data Mining:** Extracted insights from petabyte-scale datasets using advanced statistical methods
- **Model Optimization:** Achieved 40% performance improvement through hyperparameter tuning and architecture optimization

- **Production Deployment:** Successfully deployed 15+ AI models to production serving millions of users
 - **Research & Development:** Conducted cutting-edge research in generative AI and neural architecture search
-

Available for Immediate Joining | Open to Domestic & International Opportunities | Security Clearance Available