

Generative AI Engineer Job Profile

Company Information

Company: TechInnovate Solutions

Department: Artificial Intelligence & Machine Learning

Location: San Francisco, CA / Remote

Employment Type: Full-time

Experience Level: Mid to Senior Level (3-7 years)

Position Overview

We are seeking a highly skilled Generative AI Engineer to join our cutting-edge AI team. The successful candidate will be responsible for developing, implementing, and optimizing generative AI models and systems that power our next-generation products and services. This role involves working with large language models, computer vision, and multimodal AI systems.

Key Responsibilities

Model Development & Implementation

- Design and develop generative AI models including LLMs, diffusion models, and GANs
- Fine-tune and customize pre-trained models (GPT, BERT, Stable Diffusion, etc.) for specific use cases
- Implement and optimize model architectures using PyTorch, TensorFlow, and Hugging Face Transformers
- Develop prompt engineering strategies and techniques for optimal model performance

Research & Innovation

- Stay current with latest research in generative AI, NLP, and computer vision
- Experiment with novel architectures and training techniques
- Contribute to research publications and open-source projects
- Prototype new AI applications and proof-of-concepts

Production & Deployment

- Deploy generative AI models to production environments using cloud platforms (AWS, GCP, Azure)
- Implement MLOps pipelines for model training, validation, and deployment
- Optimize models for inference speed and resource efficiency

- Monitor model performance and implement continuous improvement strategies

Collaboration & Leadership

- Work closely with product managers, data scientists, and software engineers
 - Mentor junior team members and contribute to technical documentation
 - Present findings and recommendations to stakeholders and leadership
 - Participate in code reviews and maintain high code quality standards
-

Required Skills & Qualifications

Technical Skills

- **Programming Languages:** Python (expert level), experience with C++ or Rust is a plus
- **ML Frameworks:** PyTorch, TensorFlow, Hugging Face Transformers, Weights & Biases
- **Generative AI:** Experience with GPT, BERT, T5, Stable Diffusion, DALL-E, or similar models
- **Cloud Platforms:** AWS SageMaker, Google Cloud AI Platform, or Azure ML
- **MLOps:** Docker, Kubernetes, CI/CD pipelines, model versioning
- **APIs:** FastAPI, Flask, REST APIs, GraphQL
- **Databases:** Vector databases (Pinecone, Weaviate), PostgreSQL, MongoDB

Core Competencies

- **Deep Learning:** Advanced understanding of neural networks, attention mechanisms, transformers
- **NLP:** Text processing, tokenization, embeddings, semantic search
- **Computer Vision:** Image generation, style transfer, multimodal models
- **Prompt Engineering:** Advanced prompting techniques, chain-of-thought, few-shot learning
- **Model Optimization:** Quantization, pruning, distillation, ONNX conversion

Soft Skills

- Strong analytical and problem-solving abilities
 - Excellent communication and presentation skills
 - Collaborative mindset and team-oriented approach
 - Adaptability to rapidly evolving AI landscape
 - Strong project management and organizational skills
-

Preferred Qualifications

Education

- Master's or PhD in Computer Science, Machine Learning, AI, or related field
- Bachelor's degree with exceptional practical experience will be considered

Experience

- 3+ years of hands-on experience with generative AI models
- Experience with large-scale model training and distributed computing
- Previous work on production AI systems serving millions of users
- Contributions to open-source AI projects or research publications

Specialized Knowledge

- Experience with multimodal AI systems (text-to-image, image-to-text)
 - Knowledge of reinforcement learning from human feedback (RLHF)
 - Familiarity with AI safety and alignment research
 - Experience with edge deployment and model compression
 - Understanding of AI ethics and responsible AI practices
-

Technical Challenges You'll Work On

- **Custom Model Development:** Build domain-specific generative models for unique business requirements
 - **Multimodal Systems:** Develop AI systems that understand and generate text, images, and audio
 - **Real-time Generation:** Optimize models for low-latency, high-throughput applications
 - **Personalization:** Create AI systems that adapt to individual user preferences and contexts
 - **Safety & Alignment:** Implement safeguards to ensure responsible AI deployment
-

What We Offer

Compensation & Benefits

- Competitive salary range: \$150,000 - \$300,000 based on experience
- Significant equity package with high growth potential
- Comprehensive health, dental, and vision insurance
- \$5,000 annual learning and development budget
- Flexible PTO and sabbatical opportunities

Work Environment

- Hybrid work model with flexible schedules
- Access to cutting-edge AI infrastructure and compute resources
- Collaborative, research-oriented culture
- Regular tech talks, conferences, and learning opportunities
- State-of-the-art development tools and hardware

Growth Opportunities

- Clear career progression path to Principal/Staff Engineer or AI Research Lead
 - Opportunity to lead AI initiatives and build teams
 - Speaking opportunities at major AI conferences
 - Support for publishing research and contributing to open source
-

Application Process

Required Materials

1. **Resume/CV** highlighting relevant AI and ML experience
2. **Portfolio** showcasing generative AI projects and implementations
3. **GitHub Profile** with examples of your code and contributions
4. **Cover Letter** explaining your interest in generative AI and our mission

Interview Process

1. **Initial Screening** (30 minutes) - General background and motivation
2. **Technical Phone Screen** (60 minutes) - ML fundamentals and coding
3. **Take-Home Assignment** - Generative AI implementation project
4. **On-site/Virtual Interviews** (4-5 hours total):
 - Deep technical discussion with AI team
 - System design for AI applications
 - Code review and pair programming
 - Culture fit and leadership discussion
 - Final interview with hiring manager

Timeline

- Application review: 1-2 weeks
 - Complete interview process: 2-3 weeks
 - Decision and offer: 1 week
 - Start date: Flexible, typically 2-4 weeks after offer acceptance
-

About Our AI Team

Our Generative AI team is at the forefront of applied AI research and development. We work on challenging problems that impact millions of users worldwide, from creative AI tools to intelligent automation systems. Our team values:

- **Innovation:** We encourage experimentation and pushing the boundaries of what's possible
- **Collaboration:** We believe the best AI solutions come from diverse perspectives working together
- **Impact:** We focus on building AI systems that create real value for users and society
- **Learning:** We invest heavily in continuous learning and staying current with the field

Join us in shaping the future of AI and building the next generation of intelligent systems!

Contact Information

Hiring Manager: Dr. Sarah Chen, VP of AI Engineering

Email: careers@techinnovate.com

LinkedIn: [Company AI Careers Page]

Application Portal: jobs.techinnovate.com/genai-engineer

Questions? Feel free to reach out to our team at ai-hiring@techinnovate.com

TechInnovate Solutions is an equal opportunity employer committed to diversity and inclusion. We welcome applications from all qualified candidates regardless of race, gender, age, religion, sexual orientation, or disability status.

Last Updated: May 2025

Job ID: GENAI-ENG-2025-001