Job Description: Generative AI Engineer

Role & Responsibility

We are seeking a highly innovative and skilled Generative AI Engineer to join our dynamic team. In this role, you will be instrumental in researching, designing, developing, and deploying cutting-edge generative AI models across various applications. You'll work on transforming complex ideas into practical, scalable solutions, pushing the boundaries of what's possible with AI.

Your responsibilities will include:

- **Research and Development:** Stay abreast of the latest advancements in generative AI (GANs, VAEs, Transformers, Diffusion Models, LLMs, etc.) and evaluate their applicability to business problems.
- **Model Design & Implementation:** Design, build, and train novel generative AI models from concept to production, focusing on specific domains such as content generation, data synthesis, or complex system design.
- **Data Curation & Preprocessing:** Work with large, complex datasets, including structured, unstructured, and multimodal data, ensuring data quality and preparing it for model training.
- **Performance Optimization:** Optimize generative models for performance, scalability, and efficiency, including fine-tuning architectures, hyperparameter optimization, and deploying models to production environments.
- **Evaluation & Validation:** Develop robust evaluation metrics and methodologies for generative models, ensuring the quality, diversity, and fidelity of generated outputs.
- **Collaboration:** Work closely with cross-functional teams including product managers, domain experts, and MLOps engineers to integrate generative AI solutions into existing platforms and workflows.
- **Documentation:** Maintain clear and comprehensive documentation of models, methodologies, and experiments.

Skills & Experience

Experience (Mandatory)

- **3+ years** of professional experience in Machine Learning or Artificial Intelligence, with at least **1 year** specifically focused on Generative AI models.
- Proven experience in deploying AI/ML models into production environments.

- Demonstrated experience with large-scale data processing and machine learning pipelines.

Mandatory Skills :

- **Generative AI Architectures:** Strong theoretical and practical understanding of various generative models such as **GANs, VAEs, Diffusion Models, Transformers, and Large Language Models (LLMs)**.
- **Deep Learning Frameworks:** Expert proficiency in at least one major deep learning framework: **PyTorch or TensorFlow/Keras**.
- **Programming Languages:** Advanced proficiency in **Python**, including relevant libraries (NumPy, Pandas, Scikit-learn, etc.).
- **Mathematical Foundations:** Solid understanding of linear algebra, calculus, probability, and statistics.
- **Cloud Platforms:** Experience with cloud computing platforms like **Azure, AWS, or GCP**, particularly with their AI/ML services.
- **Model Evaluation:** Expertise in evaluating generative model outputs (e.g., FID, Inception Score, Perplexity, BLEU/ROUGE for text, human evaluation).
- **MLOps Principles:** Familiarity with MLOps practices for model lifecycle management, versioning, monitoring, and deployment.

Desirable Skills :

- Experience with **specific domains** like Natural Language Processing (NLP), Computer Vision, Bioinformatics, or Financial Modeling.
- Knowledge of **Reinforcement Learning** and its applications.
- Experience with **MLflow, Kubeflow, or other MLOps tools**.
- Familiarity with containerization technologies (**Docker, Kubernetes**).
- Strong communication and presentation skills, with the ability to explain complex technical concepts to non-technical stakeholders.
- A portfolio of personal projects or contributions to open-source generative AI initiatives.