

Role: AI/ML Data Scientist

Location: Bellevue, WA

Interview Mode: Video Interview

Role Summary:

We are seeking an experienced AI/ML Data Scientist to design, develop, and deploy advanced machine learning and generative AI solutions. This role requires a blend of deep technical proficiency, curiosity for cutting-edge research, and practical experience in production-grade AI/ML systems, including knowledge of LLMs, AI agents, and vector databases.

Key Responsibilities:

- Build and deploy ML and GenAI models to solve complex business problems across structured and unstructured data
- Leverage LLMs (e.g., GPT-4, LLaMA, Claude, Gemini) for use cases such as summarization, classification, chat agents, document understanding, and more
- Design and prototype AI agents capable of reasoning, memory recall, task planning, and multi-step workflows
- Work with vector databases (e.g., Pinecone, FAISS, Chroma) for semantic search and retrieval-augmented generation (RAG) systems
- Collaborate with product, engineering, and data teams to integrate models into real-world applications
- Conduct model evaluation, fine-tuning, and prompt engineering for optimization
- Stay current with the latest advancements in generative AI, foundation models, and open-source tooling
- Create clear documentation and visualizations for communicating technical solutions and results to stakeholders

Required Qualifications:

- Degree in Computer Science, Data Science, Engineering, or a related field
- 3+ years of experience in machine learning or AI/ML engineering roles
- Hands-on experience with GenAI frameworks and APIs (e.g., OpenAI, HuggingFace Transformers, LangChain, LlamaIndex)
- Experience building and scaling LLM-based applications (RAG pipelines, chatbots, agents, etc.)
- Familiarity with vector databases (e.g., Pinecone, Weaviate, FAISS)
- Solid programming skills in Python and ML libraries (e.g., PyTorch, TensorFlow, Scikit-learn)
- Understanding of ML lifecycle, including data processing, training, evaluation, and deployment
- Experience with cloud platforms (AWS/GCP/Azure) and containerization tools (Docker, Kubernetes)

Preferred Qualifications:

- Experience designing and deploying AI agents (e.g., Auto-GPT, BabyAGI, CrewAI, OpenAgents)
- Exposure to open-source LLMs (LLaMA 3, Mistral, Falcon, etc.) and fine-tuning techniques (LoRA, QLoRA, PEFT)
- Familiarity with MLOps tools and workflows (e.g., MLflow, DVC, SageMaker, Vertex AI)
- Strong communication and stakeholder engagement skills
- Publications or open-source contributions in GenAI/ML is a plus