Applied AI Engineer – Generative and Agentic AI

EXPERIENCE	PRIMARY PRACTICE	WORK AREAS
3–5 years | Lateral Hiring	Generative and Agentic AI Engineering	Product & Platform Development | Data & AI Consulting | Technology Services

Role Summary

Design, build, and productionize secure, governed enterprise Generative and Agentic AI solutions for Aptus products and customer engagements. The role spans use-case discovery, data preparation, RAG, agent orchestration, evaluation, integration, and secure cloud deployment, working across product development, consulting, and technology-services teams to deliver measurable business outcomes. This lateral role requires independent ownership of production deliverables and collaboration with enterprise stakeholders.

Key Responsibilities

- Work with customers, domain specialists, consultants, and product teams to translate business workflows, data constraints, risks, and success measures into practical AI solution designs.
- Build LLM-powered applications using Python, APIs, prompt and context engineering, embeddings, vector search, RAG, and structured and unstructured enterprise data.
- Develop tool-using and multi-agent workflows with function calling, MCP connectors, memory, human approvals, and AI gateway, model gateway, and agentic gateway controls for secure orchestration and enterprise integration.
- Create evaluation pipelines for relevance, groundedness, task completion, safety, latency, and cost; use automated tests and regression datasets to improve solution quality.
- Apply AI governance, AI security, AI ethics, and compliance controls across the solution lifecycle, including RBAC, PII and DLP protection, guardrails, audit trails, explainability, human oversight, and risk documentation.
- Prototype, containerize, deploy, monitor, troubleshoot, and document AI solutions on AWS or GCP; support demonstrations, user testing, production handover, and reusable Aptus accelerators.
- Deliver production RAG and multi-agent systems with tool permissions, MCP integration, evaluations, monitoring, and human-in-the-loop approvals.
- Conduct threat modelling, prompt-injection and data-leakage testing; optimize hallucination rates, latency, inference costs and task success.

Mandatory Skills

- 3–5 years of software/AI engineering experience, including deployed LLM, RAG or agentic applications and evaluation pipelines.
- Strong Python, SQL, Git, data structures, debugging, and REST API development, with the ability to write clean, testable, and maintainable code.
- Understanding of machine learning and NLP fundamentals, transformers, embeddings, prompting, vector search, RAG, agentic workflows, tool calling, and practical AI evaluation.
- Hands-on project experience with LLM APIs or open-source models, a vector database, and an orchestration framework such as LangChain, LlamaIndex, or Semantic Kernel.
- Working knowledge of FastAPI, Flask, or Node.js, cloud services, Docker, data integration, and AI gateway patterns for authentication, routing, fallbacks, rate limits, budgets, policies, and observability.

- Structured problem solving, ownership, clear technical communication, and the ability to convert business requirements into secure and usable AI solutions.
- Hands-on GitHub and Jira integration experience is mandatory: link Jira issues to branches, commits, pull requests, and CI/CD activity; maintain defect, review, and release traceability.

Education and Experience

- B.Tech/B.E. in Computer Science, IT or related engineering discipline; M.Tech/M.S. is an advantage. Relevant professional delivery experience, technical ownership, and demonstrable project outcomes are essential.
- 3–5 years of relevant full-time professional experience, including responsibility for production systems, engineering documentation, stakeholder communication and delivery outcomes.

Preferred Exposure and Selection

- Exposure to Hugging Face, PyTorch, LangChain, LlamaIndex, Semantic Kernel, vector databases, MCP, MLflow, Docker, Kubernetes, AWS, GCP, or Databricks.
- Familiarity with AI governance, AI security, AI ethics, regulatory and industry compliance, guardrails, PII and DLP controls, explainability, human oversight, AI observability, LLMOps, and responsible AI practices.
- Selection: relevant project screening, practical technical exercise, architecture/problem-solving discussion and delivery/ownership interview.

About Aptus Data Labs

Founded in 2014, Aptus Data Labs is an enterprise data and AI engineering company with more than 200 professionals across India, the United States, the United Kingdom, and South Africa. We deliver cloud data platforms, advanced analytics, Generative and Agentic AI, AI-powered business solutions, and AI-native and cloud-native product and platform development. Our experience spans more than 100 global projects and a portfolio of over 500 MCP connectors, APIs, and reference architectures.