

**Position:** LLM Engineer

**Location:** Bangalore, Karnataka

**Factspan Overview:**

Factspan is a pure play analytics organization. We partner with fortune 500 enterprises to build an analytics center of excellence, generating insights and solutions from raw data to solve business challenges, make strategic recommendations and implement new processes that help them succeed. With offices in Seattle, Washington and Bangalore, India; we use a global delivery model to service our customers. Our customers include industry leaders from Retail, Financial Services, Hospitality, and technology sectors.

**Job Description**

As an LLM (Large Language Model) Engineer, you will be responsible for designing, optimizing, and standardizing the architecture, codebase, and deployment pipelines of LLM-based systems. Your primary mission will focus on modernizing legacy machine learning codebases (including 40+ models) for a major retail client—enabling consistency, modularity, observability, and readiness for GenAI-driven innovation. You'll work at the intersection of ML, software engineering, and MLOps to enable seamless experimentation, robust infrastructure, and production-grade performance for language-driven systems.

This role requires deep expertise in NLP, transformer-based models, and the evolving ecosystem of LLM operations (LLMOps), along with a hands-on approach to debugging, refactoring, and building unified frameworks for scalable GenAI workloads.

**Responsibilities:**

Lead the standardization and modernization of legacy ML codebases by aligning to current LLM architecture best practices.

Re-architect code for 40+ legacy ML models, ensuring modularity, documentation, and consistent design patterns.

Design and maintain pipelines for fine-tuning, evaluation, and inference of LLMs using Hugging Face, OpenAI, or open-source stacks (e.g., LLaMA, Mistral, Falcon).

Build frameworks to operationalize prompt engineering, retrieval-augmented generation (RAG), and few-shot/in-context learning methods.

Collaborate with Data Scientists, MLOps Engineers, and Platform teams to implement scalable CI/CD pipelines, feature stores, model registries, and unified experiment tracking.

Benchmark model performance, latency, and cost across multiple deployment environments (on-premise, GCP, Azure).

Develop governance, access control, and audit logging mechanisms for LLM outputs to ensure data safety and compliance.

Mentor engineering teams in code best practices, versioning, and LLM lifecycle maintenance.

**Key Skills:**

Deep understanding of transformer architectures, tokenization, attention mechanisms, and training/inference optimization

Proven track record in standardizing ML systems using OOP design, reusable components, and scalable service APIs

Hands-on experience with MLflow, LangChain, Ray, Prefect/Airflow, Docker, K8s, Weights & Biases, and model-serving platforms.

Strong grasp of prompt tuning, evaluation metrics, context window management, and hybrid search strategies using vector databases like FAISS, pgvector, or Milvus

Proficient in Python (must), with working knowledge of shell scripting, YAML, and JSON schema standardization

Experience managing compute, memory, and storage requirements of LLMs across GCP, Azure, or AWS environments

**Qualifications & Experience:**

5+ years in ML/AI engineering with at least 2 years working on LLMs or NLP-heavy systems.

Able to reverse-engineer undocumented code and reimagine it with strong documentation and testing in mind.

Clear communicator who collaborates well with business, data science, and DevOps teams.

Familiar with agile processes, JIRA, GitOps, and confluence-based knowledge sharing.

Curious and future-facing—always exploring new techniques and pushing the envelope on GenAI innovation.

Passionate about data ethics, responsible AI, and building inclusive systems that scale

**Why Should You Apply?**

Grow with Us: Be part of a hyper- growth startup with ample number of opportunities to Learn & Innovate.

People: Join hands with the talented, warm, collaborative team and highly accomplished leadership.

Buoyant Culture: Regular activities like Fun-Fridays, Sports tournaments, Trekking and you can suggest few more after joining us 😊