Introduction to LLMOPS

what is LLMOps ?

* LLMOPS involves best practise, techniques, and tools used for operationalizing LLMs in production environments.

* LLMOPS enables practitioners to apply best practises when building, deploying and maintaining LLMs.

LLMOps Benefits

* Reduces time and cost to build and deploy LLMs by streamlining the process

* Improves performances and reliability of LLMs through best practises.

* Enables effectively managing, scaling and maintaining LLMs

* Addresses catastrophic failure via monitoring and maintenance.

LLMOPS vs MLOPS

* MLOPS is board whereas LLMOps focuses on LLM's

* LLMOPS requires dealing with different

→ data requirements   → evaluation strategies
→ experimentation   → costs and latency.

LLMS considerations

* Tuning base Models which requires high-quality data and efficient techniques.

* Evaluating & debugging LLMS efficiently which depends on the task and business objective.

* One unique thing to Large Language model (LLM) is prompt Engineering. Improving results via prompt Engineering and efficient prompt management, tracking and optimization.

* Building Pipelines that requires complex chaining of LLM calls and operations

* optimizing for cost and latency.

* versioning and maintaining models to address safety, cost and quality

* collecting human feedback to improve LLMS.

Overview of LLMops - life cycle

```
┌─────────────────────────┐
│   Data preparation      │
│   & Exploration         │
└─────────────────────────┘
            ↓
┌─────────────────────────┐
│   pretraining           │
│   LLM.                  │
└─────────────────────────┘
            ↓
┌─────────────────────────┐
│   Model fine tuning     │
│   & prompt              │
│   Engineering           │
└─────────────────────────┘
            ↓
┌─────────────────────────┐
│   Model Evaluation      │
│   & Debugging           │
└─────────────────────────┘
            ↓
┌─────────────────────────┐
│   Model                 │
│   Deployment            │
└─────────────────────────┘
            ↓
┌─────────────────────────┐
│   Model                 │
│   Monitoring &          │
│   Maintenaces.          │
└─────────────────────────┘
```