

# LLM Apps

2023 AI Engineering Survey:  
Stack Used

# Orchestration frameworks

1. Langchain and LlamaIndex are practically tied.

# Other popular frameworks

1. Deepset Haystack
2. Vercel AI SDK
3. MS TypeChat
4. Guardrails
5. MS Guidance
6. MS Semantic Kernel

# Vector Databases

1. Pinecone
2. Pg\_embedding
3. PGvector (postgres)
4. Chroma
5. Supabase
6. Redis
7. Elastic
8. MongoDB vector search
9. Databricks Lakehouse
10. Faiss
11. Weaviate
12. LanceDB
13. Qdrant

# Monitoring / Observability

1. Arthur
2. Arize
3. Fiddler
4. Gantry
5. Helicone
6. Langsmith
7. Existing monitoring stack

# Model serving and hosting

1. HuggingFace
2. Modal
3. Replicate
4. Baseten
5. Anyscale
6. MosaicML
7. Runpod
8. Together.ai
9. OpenAI directly
10. Anthropic directly
11. AWS SageMaker
12. Google Vertex
13. OctoML
14. Amazon Bedrock
15. FireworksAI
16. Azure OpenAI
17. GCP

# Model training and fine-tuning

1. HuggingFace
2. Anyscale
3. MosaicML
4. OpenAI fine-tuning
5. AWS SageMaker
6. PyTorch Lightning
7. OctoML
8. Modal
9. DeepSpeed-chat

# Prompt Management

- Most AI Engineers have built an internal prompt management tool instead of opting for an external tool.
  1. Humanloop
  2. Honeyhive
  3. Promptlayer
  4. Scale Spellbook
  5. Weight & Biases Prompts
  6. LangChain / LangSmith Hub
  7. Hegel.ai prompttools
  8. Vellum