365 DataScience

THE WORLD OF OPEN-SOURCE GEN AI

KEY AREAS	KEY PLAYERS	DESCRIPTION	ADVANTAGES	CONSIDERATIONS
FOUNDATION MODELS	LLaMA by Meta, LaMDA, BERT, MISTRAL AI_, ChatGPT, Hugging Face	Large, pre-trained models—publicly available as bases for further fine-tuning and task-specific adaptation.	• Cost-effective development • Domain-specific customizations • Community-driven improvements • Accelerated Innovation	• Complex setup • Licensing of data used for training • Data leakage risks • UX and Customer Support limitations
DATASETS	Hugging Face, kaggle, Snowflake, GitHub	Freely available for training, validating, and testing of AI models across diverse fields.	• Cost-effectiveness • Diversity and variety	• Complex data licensing • Risk of training with AI-generated data
VECTOR DATABASES	Weaviate, Milvus, Chroma, Qdrant, LlamaIndex, Hugging Face	Store and retrieve AI-generated vectors—enabling fast and scalable searches for similar items.	• Quick similarity searches • Allow LLMs to digest big data • Accelerated search through indexing	• Complex setup • High compute costs
ORCHESTRATION TOOLS	LangChain, Transformers, Langdock, FlowiseAI	Simplify the construction of LLM-powered apps by streamlining complex processes.	• Developer-friendly interfaces • Code consistency • Support for multiple LLMs • Workflow automation	• Security concerns • Feature limitations
EVALUATION TOOLS	DeepEval., ragas, deepchecks., arize	Assess AI model performance and reliability using AI as a judge.	• Cost-effective evaluation	• Validity of metrics can vary by use case.
COMMUNITY	Hugging Face, GitHub, kaggle, stackoverflow	A community of passionate developers, researchers, and users collaborating to advance AI technology and share best practices.	• Quick problem-solving • Shared knowledge • Innovation	• Quality control • Security risks • Legal compliance
ETHICS	AI for Good Foundation, IEET, Center for AI Safety	Ensure fairness, accountability, and transparency in AI systems to uphold ethical standards and societal values.	• Increase in public transparency • Diverse perspectives of the open-source community	• Unclear responsibility ownership • Inconsistent policy enforcement