# THE WORLD OF OPEN-SOURCE GEN AI



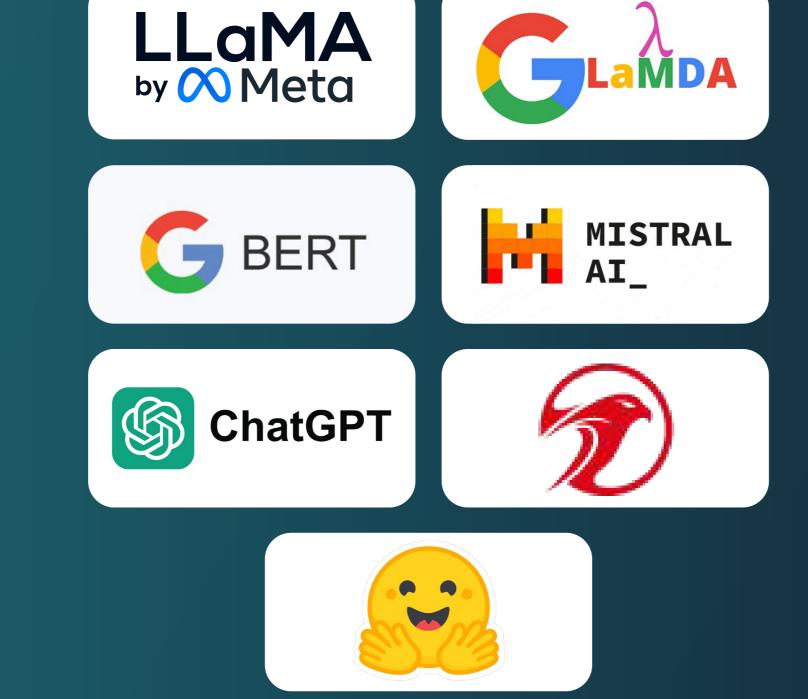








#### FOUNDATION MODELS

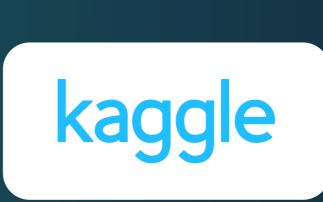


Large, pre-trained models—publicly available as bases for further fine-tuning and task-specific adaptation.

- 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



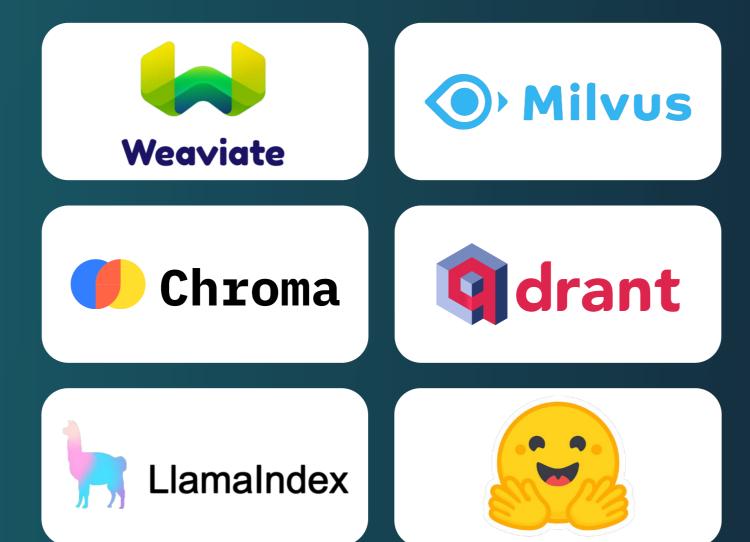




Freely available for training, validating, and testing of Al models across diverse fields.

- Cost-effectiveness
- Diversity and variety
- Complex data licensing
- Risk of training with Al-generated data

### VECTOR DATABASES



Store and retrieve Algenerated vectors enabling fast and scalable searches for similar items.

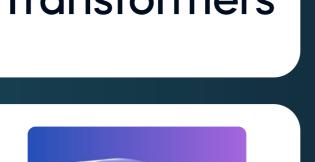
- Quick similarity searches
- Allow LLMs to digest big data
- Accelerated search through indexing
- Complex setup
- High compute costs

#### ORCHESTRATION TOOLS









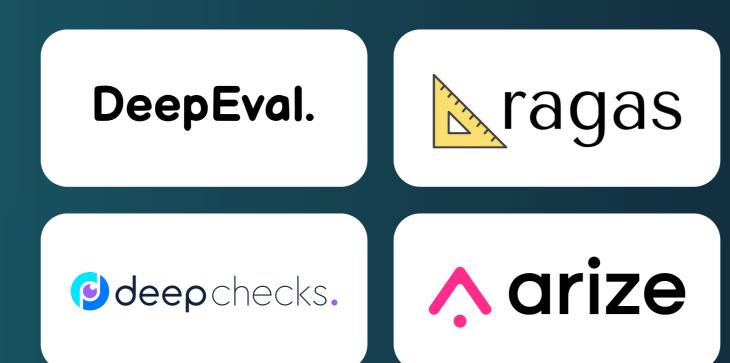
construction of LLMpowered apps by streamlining complex processes.

Simplify the

- Developer-friendly interfaces
- Code consistency
- Support for multiple LLMs
- Workflow automation
- Security concerns

Feature limitations

EVALUATION TOOLS



Assess Al model performance and reliability using Al as a judge.

- Cost-effective evaluation
- Validity of metrics can vary by use case.

COMMUNITY



kaggle







A community of passionate developers, researchers, and users collaborating to advance AI technology and share best practices.

- Quick problemsolving
- Shared knowledge
- Innovation
- Quality control
- Security risks
- Legal compliance

## ETHICS



Ensure fairness, accountability, and transparency in Al 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