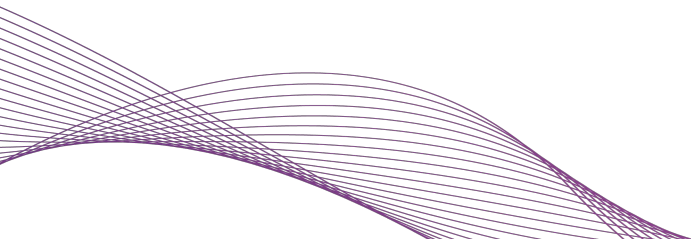


(Python) Knjižnice

Za uporabo LLM-jev

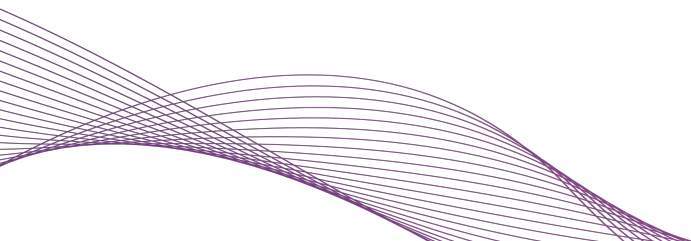


Knjižnice:

- Huggingface (“nizko” nivojska),
- LangChain (python, TypeScript, Agent -> Vse),
- LlamaIndex (Python, TypeScript, RAG -> Vse),
- DSPy (experimentalno, stanfordnlp, honorable mention),
- ...

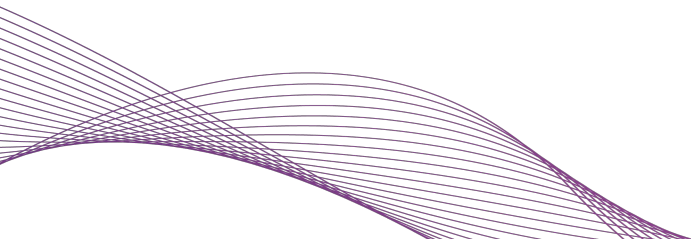
LangChain :

- Framework za razvoj aplikacij, katere poganjajo LLM-ji
- LangChain, LangChain-Community, LangChain-Core
- Ecosystem built around it...



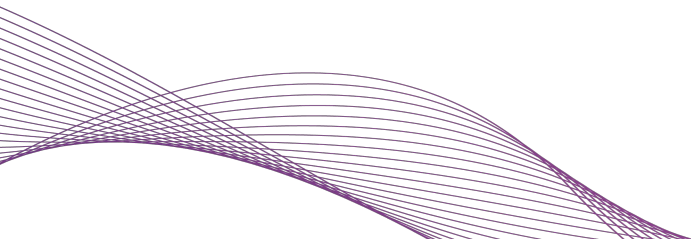
LangServe :

- Pomaga nuditi LangChain objekte,
- Automatična zaznava in definicija endpointov,
- Parcelizacija,
- Debug endpoint,
- Integracija z LangSmith.



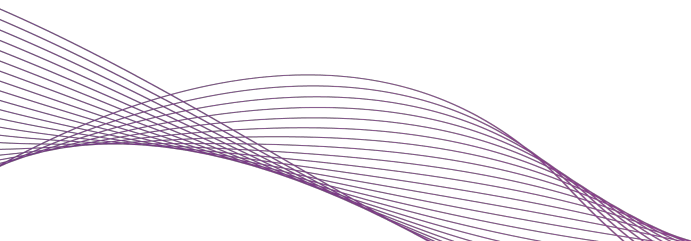
LangSmith:

- MLOPs platform,
- Deluje tudi z drugimi knjižnicami,
- Trace logging, evalvacije, prompt versioning.



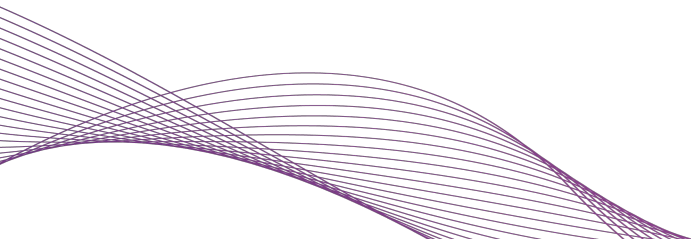
LangGraph :

- Stateful aplikacije z več agenti,
- DAG - Usmerjeni aciklični grafi = LangChain
- LangGraph omogoča dodajanje ciklov.



LCEL (LangChain Expression Language)

- Preprost način za sestavljanje verig,
- Cilj: Prototipi v produkcijo brez sprememb,
- Veriga je asinhrona, ima podporo za paralelno poganjanje,



Observability

LangSmith

Deployments

LangServe

Chains as Rest APIs

Python

Templates

Reference Applications

Python

Cognitive
Architectures

LangChain

Python

JavaScript

Chains Agents Retrieval Strategies

LangChain-Community

Python

JavaScript

Model I/O

Model
Prompt
Example Selector
Output Parser

Retrieval

Retriever
Document Loader
Vector Store
Text Splitter
Embedding Model

Agent Tooling

Tool
Toolkit

Integrations
Components

LangChain-Core

Python

JavaScript

LCEL - LangChain Expression Language

Parallelization Fallbacks Tracing Batching Streaming Async Composition

Protocol

Debugging

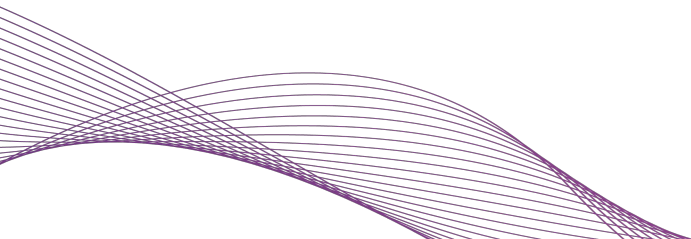
Playground

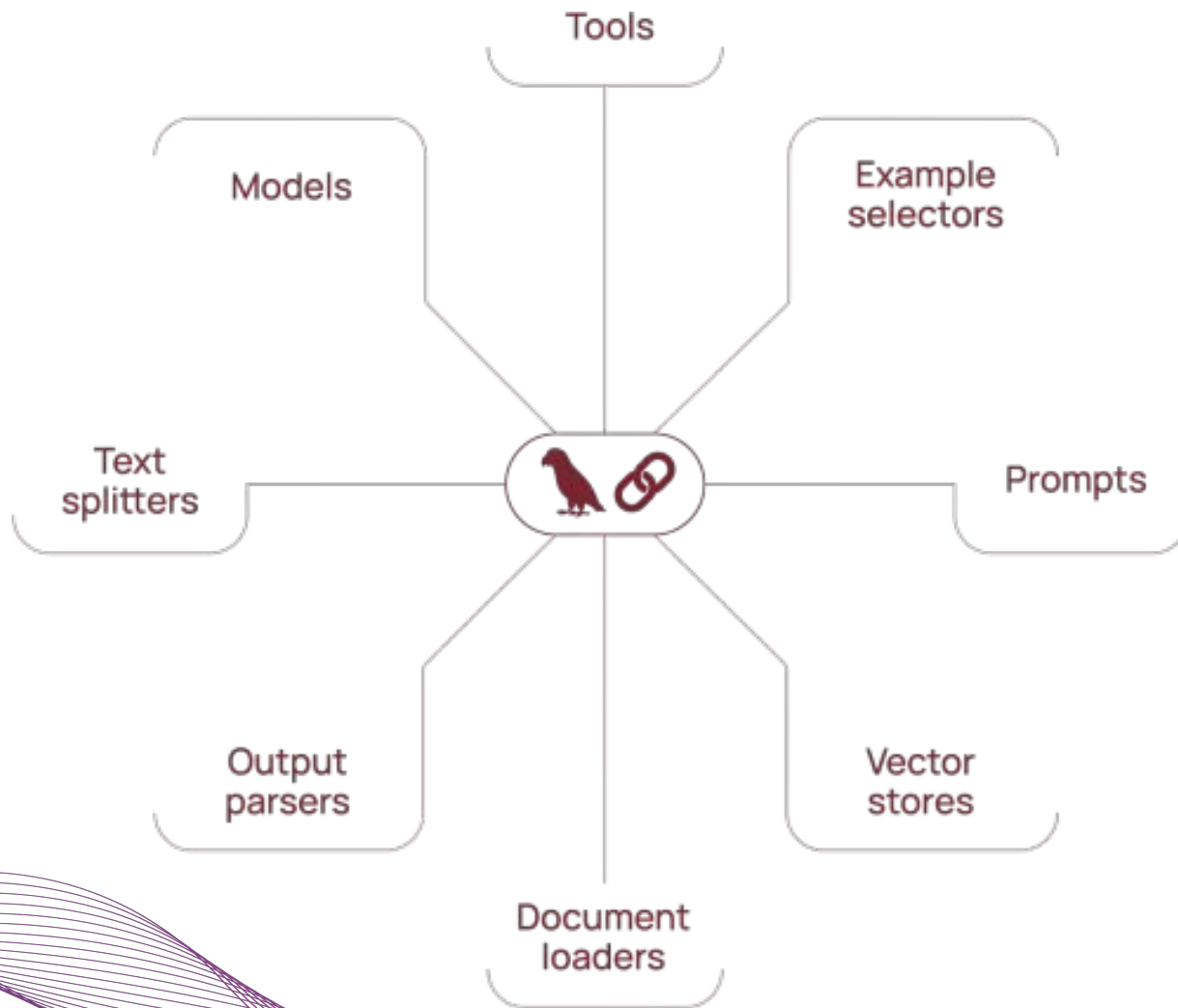
Evaluation

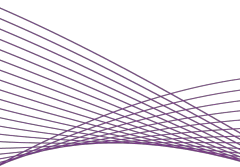
Annotation

Monitoring

Componente

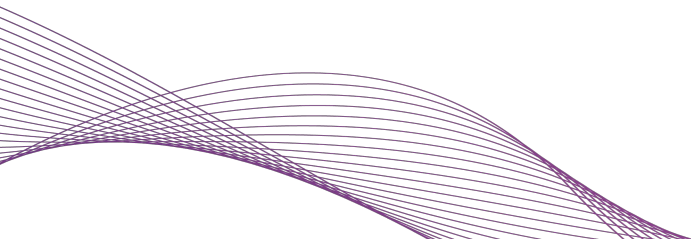






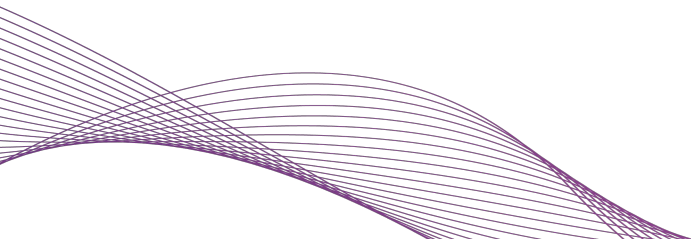
Modeli

- Na voljo na različnih service-ih
- Enoten I/O za vse modele ne glede na to kje jih poganjamo,
- 2 tipa:
 - Completions,
 - Chat models.



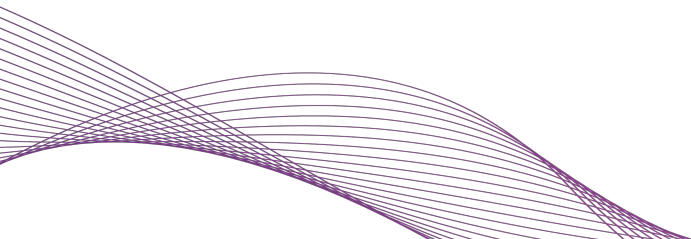
Pozivi

- Enoten sistem za gradnjo pozivov,
- Enoten sistem za pakiranje.



Retrieval:

- Nalaganje dokumentov,
- Deljenje dokumentov,
- Generacija vektorskih vložitev (embeddingov),
- Povezava na vektorske podatkovne baze (dodajanje, branje).

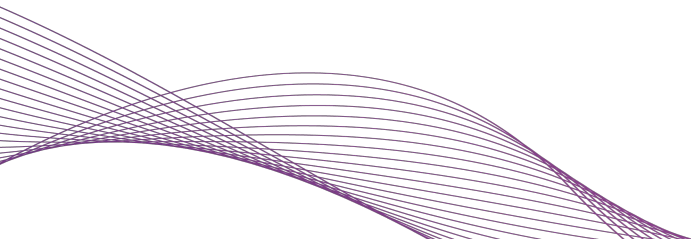


Orodja:

- Enoten vmesnik za orodja,
- Splošna orodja za pogoste integracije:
 - Github,
 - Gmail,
 - Web,
 - ...

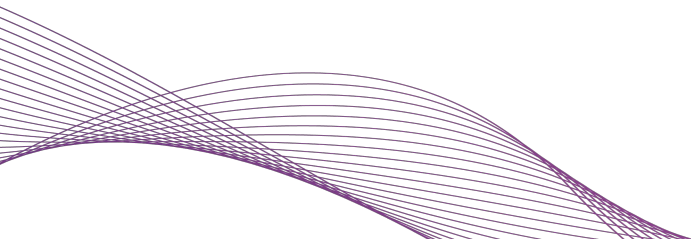
Agenti:

- Implementacije standardnih ogrodji za agente (ReAct),
- Skupaj s pozivi (v Angleščini),
- Preposto dodajanje orodij, ...



Verige:

- Povezovanje vseh komponent v verige,
- Standardiziran I/O omogoča generiranje vrig komponent s poljubnimi elementi.



Spomin :

- Ohranjanje stanja med klici verige,
- Chat History, Summarized chat history, ...

