

title  
AWS S3 File |  Langchain  
Amazon Comprehend Moderation Chain |  Langchain  
AWS S3 Directory |  Langchain  
SageMaker |  Langchain  
Bedrock |  Langchain  
BedrockChat |  Langchain  
SageMakerEndpoint |  Langchain  
Streaming |  Langchain  
Quick Start |  Langchain  
Amazon API Gateway |  Langchain  
Bedrock |  Langchain  
Custom LLM |  Langchain  
Caching |  Langchain  
Tracking token usage |  Langchain  
Tracking token usage |  Langchain  
Caching |  Langchain  
Streaming |  Langchain  
Quick Start |  Langchain  
Time-weighted vector store retriever |  Langchain  
Contextual compression |  Langchain  
Long-Context Reorder |  Langchain  
CacheBackedEmbeddings |  Langchain  
Split by character |  Langchain  
Split by tokens |  Langchain  
Semantic Chunking |  Langchain  
Split code |  Langchain  
HTMLHeaderTextSplitter |  Langchain  
PDF |  Langchain  
JSON |  Langchain  
File Directory |  Langchain  
Markdown |  Langchain  
CSV |  Langchain  
HTML |  Langchain  
Neptune Open Cypher QA Chain |  Langchain  
GraphSparqlQACChain |  Langchain  
Memgraph QA chain |  Langchain  
NebulaGraphQACChain |  Langchain  
NetworkX |  Langchain  
Neo4j DB QA chain |  Langchain  
FalkorDBQACChain |  Langchain  
KuzuQACChain |  Langchain  
HugeGraph QA Chain |  Langchain  
ArangoDB QA chain |  Langchain  
Diffbot Graph Transformer |  Langchain  
Page Not Found |  Langchain  
Page Not Found |  Langchain  
Page Not Found |  Langchain  
Page Not Found |  Langchain  
Page Not Found |  Langchain  
OpenAI metadata tagger |  Langchain  
ChatAnthropic |  Langchain

Chains |  Langchain  
Chains |  Langchain  
Anthropic Functions |  Langchain  
Chains |  Langchain  
Llama.cpp |  Langchain  
GPT4All |  Langchain  
JSONFormer |  Langchain  
Chat Messages |  Langchain  
MultiVector Retriever |  Langchain  
Text Splitters |  Langchain  
MultiQueryRetriever |  Langchain  
Chroma |  Langchain  
Vector store-backed retriever |  Langchain  
OpenAI |  Langchain  
Grobid |  Langchain  
Source Code |  Langchain  
MarkdownHeaderTextSplitter |  Langchain  
Recursively split by character |  Langchain  
WebBaseLoader |  Langchain  
CSV |  Langchain  
SQL Database |  Langchain  
Pandas Dataframe |  Langchain  
CSV |  Langchain  
Pandas DataFrame |  Langchain  
Token counting |  Langchain  
Tools as OpenAI Functions |  Langchain  
LangSmith Walkthrough |  Langchain  
OpenAI functions |  Langchain  
Timeouts for agents |  Langchain  
Access intermediate steps |  Langchain  
Cap the max number of iterations |  Langchain  
Handle parsing errors |  Langchain  
Running Agent as an Iterator |  Langchain  
Returning Structured Output |  Langchain  
Streaming |  Langchain  
Custom agent |  Langchain  
Tools |  Langchain  
Agent Types |  Langchain  
Quickstart |  Langchain  
Concepts |  Langchain  
YAML parser |  Langchain  
Ensemble Retriever |  Langchain  
Self-querying retriever |  Langchain  
Parent Document Retriever |  Langchain  
Indexing |  Langchain  
Output Parsers |  Langchain  
Quickstart |  Langchain  
Prompts |  Langchain  
Concepts |  Langchain  
[Beta] Memory |  Langchain  
Using tools |  Langchain  
Managing prompt size |  Langchain

Adding moderation | 🐱 ⚡ Langchain  
Adding memory | 🐱 ⚡ Langchain  
Routing by semantic similarity | 🐱 ⚡ Langchain  
Code writing | 🐱 ⚡ Langchain  
Agents | 🐱 ⚡ Langchain  
Querying a SQL DB | 🐱 ⚡ Langchain  
Multiple chains | 🐱 ⚡ Langchain  
Prompt + LLM | 🐱 ⚡ Langchain  
RAG | 🐱 ⚡ Langchain  
Add message history (memory) | 🐱 ⚡ Langchain  
Inspect your runnables | 🐱 ⚡ Langchain  
Stream custom generator functions | 🐱 ⚡ Langchain  
Add fallbacks | 🐱 ⚡ Langchain  
Create a runnable with the `@chain` decorator | 🐱 ⚡ Langchain  
Configure chain internals at runtime | 🐱 ⚡ Langchain  
Bind runtime args | 🐱 ⚡ Langchain  
RunnableBranch: Dynamically route logic based on input | 🐱 ⚡ Langchain  
RunnableParallel: Manipulating data | 🐱 ⚡ Langchain  
RunnableLambda: Run Custom Functions | 🐱 ⚡ Langchain  
RunnablePassthrough: Passing data through | 🐱 ⚡ Langchain  
Faiss | 🐱 ⚡ Langchain  
Tavily Search API | 🐱 ⚡ Langchain  
xml-agent | 🐱 ⚡ Langchain  
vertexai-chuck-norris | 🐱 ⚡ Langchain  
summarize-anthropic | 🐱 ⚡ Langchain  
stepback-qa-prompting | 🐱 ⚡ Langchain  
sql-pgvector | 🐱 ⚡ Langchain  
sql-research-assistant | 🐱 ⚡ Langchain  
sql-ollama | 🐱 ⚡ Langchain  
sql-llamacpp | 🐱 ⚡ Langchain  
solo-performance-prompting-agent | 🐱 ⚡ Langchain  
sql-llama2 | 🐱 ⚡ Langchain  
Langchain - Robocorp Action Server | 🐱 ⚡ Langchain  
skeleton-of-thought | 🐱 ⚡ Langchain  
self-query-qdrant | 🐱 ⚡ Langchain  
self-query-supabase | 🐱 ⚡ Langchain  
retrieval-agent | 🐱 ⚡ Langchain  
rewrite\_retrieve\_read | 🐱 ⚡ Langchain  
research-assistant | 🐱 ⚡ Langchain  
rag-vectara | 🐱 ⚡ Langchain  
rag-vectara-multiquery | 🐱 ⚡ Langchain  
rag-weaviate | 🐱 ⚡ Langchain  
RAG with Timescale Vector using hybrid search | 🐱 ⚡ Langchain  
rag\_supabase | 🐱 ⚡ Langchain  
rag-timescale-conversation | 🐱 ⚡ Langchain  
rag-self-query | 🐱 ⚡ Langchain  
rag-singlestoredb | 🐱 ⚡ Langchain  
rag-pinecone | 🐱 ⚡ Langchain  
rag-semi-structured | 🐱 ⚡ Langchain  
rag-pinecone-rerank | 🐱 ⚡ Langchain  
rag-redis | 🐱 ⚡ Langchain  
rag-pinecone-multi-query | 🐱 ⚡ Langchain

rag-ollama-multi-query | 🐱 ⚡ Langchain  
rag-opensearch | 🐱 ⚡ Langchain  
rag-multi-modal-mv-local | 🐱 ⚡ Langchain  
rag-multi-modal-local | 🐱 ⚡ Langchain  
RAG with Multiple Indexes (Routing) | 🐱 ⚡ Langchain  
RAG with Multiple Indexes (Fusion) | 🐱 ⚡ Langchain  
rag-mongo | 🐱 ⚡ Langchain  
rag-matching-engine | 🐱 ⚡ Langchain  
rag-momento-vector-index | 🐱 ⚡ Langchain  
rag-gpt-crawler | 🐱 ⚡ Langchain  
rag-gemini-multi-modal | 🐱 ⚡ Langchain  
rag-google-cloud-sensitive-data-protection | 🐱 ⚡ Langchain  
rag-elasticsearch | 🐱 ⚡ Langchain  
rag-google-cloud-vertexai-search | 🐱 ⚡ Langchain  
rag-fusion | 🐱 ⚡ Langchain  
rag-conversation | 🐱 ⚡ Langchain  
rag-conversation-zep | 🐱 ⚡ Langchain  
rag-codellama-fireworks | 🐱 ⚡ Langchain  
rag-chroma | 🐱 ⚡ Langchain  
rag-chroma-private | 🐱 ⚡ Langchain  
rag-chroma-multi-modal | 🐱 ⚡ Langchain  
rag-chroma-multi-modal-multi-vector | 🐱 ⚡ Langchain  
rag-aws-kendra | 🐱 ⚡ Langchain  
rag-aws-bedrock | 🐱 ⚡ Langchain  
python-lint | 🐱 ⚡ Langchain  
rag-astradb | 🐱 ⚡ Langchain  
openai-functions-tool-retrieval-agent | 🐱 ⚡ Langchain  
pirate-speak | 🐱 ⚡ Langchain  
pirate-speak-configurable | 🐱 ⚡ Langchain  
plate-chain | 🐱 ⚡ Langchain  
propositional-retrieval | 🐱 ⚡ Langchain  
pii-protected-chatbot | 🐱 ⚡ Langchain  
openai-functions-agent | 🐱 ⚡ Langchain  
OpenAI Functions Agent - Gmail | 🐱 ⚡ Langchain  
nvidia-rag-canonical | 🐱 ⚡ Langchain  
neo4j-generation | 🐱 ⚡ Langchain  
neo4j-parent | 🐱 ⚡ Langchain  
neo4j-semantic-layer | 🐱 ⚡ Langchain  
neo4j-vector-memory | 🐱 ⚡ Langchain  
neo4j\_cypher | 🐱 ⚡ Langchain  
neo4j-cypher-memory | 🐱 ⚡ Langchain  
neo4j-advanced-rag | 🐱 ⚡ Langchain  
neo4j-cypher-ft | 🐱 ⚡ Langchain  
mongo-parent-document-retrieval | 🐱 ⚡ Langchain  
llama2-functions | 🐱 ⚡ Langchain  
hyde | 🐱 ⚡ Langchain  
Hybrid Search in Weaviate | 🐱 ⚡ Langchain  
guardrails-output-parser | 🐱 ⚡ Langchain  
extraction-openai-functions | 🐱 ⚡ Langchain  
elastic-query-generator | 🐱 ⚡ Langchain  
extraction-anthropic-functions | 🐱 ⚡ Langchain  
csv-agent | 🐱 ⚡ Langchain

cohere-librarian | 🐱 ⚡ Langchain  
Chat Bot Feedback Template | 🐱 ⚡ Langchain  
cassandra-synonym-caching | 🐱 ⚡ Langchain  
Chain-of-Note (Wikipedia) | 🐱 ⚡ Langchain  
cassandra-entomology-rag | 🐱 ⚡ Langchain  
Bedrock JCVD 🗃️ | 🐱 ⚡ Langchain  
basic-critique-revise | 🐱 ⚡ Langchain  
anthropic-iterative-search | 🐱 ⚡ Langchain  
Contribute Code | 🐱 ⚡ Langchain  
Contribute Integrations | 🐱 ⚡ Langchain  
Frequently Asked Questions | 🐱 ⚡ Langchain  
Testing | 🐱 ⚡ Langchain  
Contribute Documentation | 🐱 ⚡ Langchain  
langchain | 🐱 ⚡ Langchain  
Callbacks | 🐱 ⚡ Langchain  
Logging to file | 🐱 ⚡ Langchain  
langchain-core | 🐱 ⚡ Langchain  
WandB Tracing | 🐱 ⚡ Langchain  
Safety | 🐱 ⚡ Langchain  
Model comparison | 🐱 ⚡ Langchain  
Pydantic compatibility | 🐱 ⚡ Langchain  
Run LLMs locally | 🐱 ⚡ Langchain  
Fallbacks | 🐱 ⚡ Langchain  
Activeloop Deep Lake | 🐱 ⚡ Langchain  
Evaluation | 🐱 ⚡ Langchain  
Deployment | 🐱 ⚡ Langchain  
Stores | 🐱 ⚡ Langchain  
Adapters | 🐱 ⚡ Langchain  
Chat loaders | 🐱 ⚡ Langchain  
Callbacks | 🐱 ⚡ Langchain  
Memory | 🐱 ⚡ Langchain  
Tools | 🐱 ⚡ Langchain  
Agents and toolkits | 🐱 ⚡ Langchain  
Retrievers | 🐱 ⚡ Langchain  
Vector stores | 🐱 ⚡ Langchain  
Text embedding models | 🐱 ⚡ Langchain  
Document loaders | 🐱 ⚡ Langchain  
Document transformers | 🐱 ⚡ Langchain  
Chat models | 🐱 ⚡ Langchain  
LLMs | 🐱 ⚡ Langchain  
Microsoft | 🐱 ⚡ Langchain  
Components | 🐱 ⚡ Langchain  
OpenAI | 🐱 ⚡ Langchain  
Hugging Face | 🐱 ⚡ Langchain  
Google | 🐱 ⚡ Langchain  
AWS | 🐱 ⚡ Langchain  
Anthropic | 🐱 ⚡ Langchain  
LLMs | 🐱 ⚡ Langchain  
Chat Models | 🐱 ⚡ Langchain  
Retrievers | 🐱 ⚡ Langchain  
Vector stores | 🐱 ⚡ Langchain  
Text embedding models | 🐱 ⚡ Langchain

Text Splitters |   Langchain  
Document loaders |   Langchain  
Synthetic data generation |   Langchain  
Graph querying |   Langchain  
Code understanding |   Langchain  
Web scraping |   Langchain  
Tagging |   Langchain  
Summarization |   Langchain  
Extraction |   Langchain  
Interacting with APIs |   Langchain  
Using local models |   Langchain  
Using agents |   Langchain  
Streaming |   Langchain  
Per-User Retrieval |   Langchain  
Add chat history |   Langchain  
Quickstart |   Langchain  
Returning sources |   Langchain  
Welcome Contributors |   Langchain  
SQL |   Langchain  
Chatbots |   Langchain  
  LangGraph |   Langchain  
Q&A with RAG |   Langchain  
  LangServe |   Langchain  
LangSmith |   Langchain  
Chains |   Langchain  
Agents |   Langchain  
Retrieval |   Langchain  
Model I/O |   Langchain  
Modules |   Langchain  
Cookbook |   Langchain  
Why use LCEL |   Langchain  
How to |   Langchain  
Interface |   Langchain  
Get started |   Langchain  
LangChain Expression Language (LCEL) |   Langchain  
Security |   Langchain  
Quickstart |   Langchain  
Installation |   Langchain  
Get started |   Langchain  
YouTube videos |   Langchain  
Tutorials |   Langchain  
Templates |   Langchain  
Welcome Contributors |   Langchain  
Changelog |   Langchain  
  Package Versioning |   Langchain  
Debugging |   Langchain  
Providers |   Langchain  
Q&A with RAG |   Langchain  
Introduction |   Langchain  
Introduction |   Langchain

url  
https://python.langchain.com/docs/integrations/document\_loaders/aws\_s3\_file  
https://python.langchain.com/docs/guides/safety/amazon\_comprehend\_chain  
https://python.langchain.com/docs/integrations/document\_loaders/aws\_s3\_directory  
https://python.langchain.com/docs/integrations/text\_embedding/sagemaker-endpoint  
https://python.langchain.com/docs/integrations/text\_embedding/bedrock  
https://python.langchain.com/docs/integrations/chat/bedrock  
https://python.langchain.com/docs/integrations/lms/sagemaker  
https://python.langchain.com/docs/modules/model\_io/lms/streaming\_llm  
https://python.langchain.com/docs/modules/model\_io/lms/quick\_start  
https://python.langchain.com/docs/integrations/lms/amazon\_api\_gateway  
https://python.langchain.com/docs/integrations/lms/bedrock  
https://python.langchain.com/docs/modules/model\_io/lms/custom\_llm  
https://python.langchain.com/docs/modules/model\_io/lms/llm\_caching  
https://python.langchain.com/docs/modules/model\_io/lms/token\_usage\_tracking  
https://python.langchain.com/docs/modules/model\_io/chat/token\_usage\_tracking  
https://python.langchain.com/docs/modules/model\_io/chat/chat\_model\_caching  
https://python.langchain.com/docs/modules/model\_io/chat/streaming  
https://python.langchain.com/docs/modules/model\_io/chat/quick\_start  
https://python.langchain.com/docs/modules/data\_connection/retrievers/time\_weighted\_vectorstore  
https://python.langchain.com/docs/modules/data\_connection/retrievers/contextual\_compression  
https://python.langchain.com/docs/modules/data\_connection/retrievers/long\_context\_reorder  
https://python.langchain.com/docs/modules/data\_connection/text\_embedding/caching\_embeddings  
https://python.langchain.com/docs/modules/data\_connection/document\_transformers/character\_text\_splitter  
https://python.langchain.com/docs/modules/data\_connection/document\_transformers/split\_by\_token  
https://python.langchain.com/docs/modules/data\_connection/document\_transformers/semantic-chunker  
https://python.langchain.com/docs/modules/data\_connection/document\_transformers/code\_splitter  
https://python.langchain.com/docs/modules/data\_connection/document\_transformers/HTML\_header\_metadata  
https://python.langchain.com/docs/modules/data\_connection/document\_loaders/pdf  
https://python.langchain.com/docs/modules/data\_connection/document\_loaders/json  
https://python.langchain.com/docs/modules/data\_connection/document\_loaders/file\_directory  
https://python.langchain.com/docs/modules/data\_connection/document\_loaders/markdown  
https://python.langchain.com/docs/modules/data\_connection/document\_loaders/csv  
https://python.langchain.com/docs/modules/data\_connection/document\_loaders/html  
https://python.langchain.com/docs/use\_cases/graph/neptune\_cypher\_qa  
https://python.langchain.com/docs/use\_cases/graph/graph\_spqrql\_qa  
https://python.langchain.com/docs/use\_cases/graph/graph\_memgraph\_qa  
https://python.langchain.com/docs/use\_cases/graph/graph\_nebula\_qa  
https://python.langchain.com/docs/use\_cases/graph/graph\_networkx\_qa  
https://python.langchain.com/docs/use\_cases/graph/graph\_cypher\_qa  
https://python.langchain.com/docs/use\_cases/graph/graph\_falkordb\_qa  
https://python.langchain.com/docs/use\_cases/graph/graph\_kuzu\_qa  
https://python.langchain.com/docs/use\_cases/graph/graph\_hugegraph\_qa  
https://python.langchain.com/docs/use\_cases/graph/graph\_arangodb\_qa  
https://python.langchain.com/docs/use\_cases/graph/diffbot\_graphtransformer  
https://python.langchain.com/docs/modules/data\_connection/retrievers/web\_research  
https://python.langchain.com/docs/use\_cases/docs/use\_cases/extraction  
https://python.langchain.com/docs/use\_cases/docs/integrations/document\_transformers/html2text  
https://python.langchain.com/docs/use\_cases/docs/integrations/document\_loaders/async\_chromium  
https://python.langchain.com/docs/use\_cases/docs/integrations/document\_loaders/async\_html  
https://python.langchain.com/docs/integrations/document\_transformers/openai\_metadata\_tagger  
https://python.langchain.com/docs/integrations/chat/anthropic

<https://python.langchain.com/docs/modules/chains>  
<https://python.langchain.com/docs/modules/chains>  
[https://python.langchain.com/docs/integrations/chat/anthropic\\_functions](https://python.langchain.com/docs/integrations/chat/anthropic_functions)  
<https://python.langchain.com/docs/modules/chains>  
<https://python.langchain.com/docs/integrations/lms/llamacpp>  
<https://python.langchain.com/docs/integrations/lms/gpt4all>  
[https://python.langchain.com/docs/integrations/lms/jsonformer\\_experimental](https://python.langchain.com/docs/integrations/lms/jsonformer_experimental)  
[https://python.langchain.com/docs/modules/memory/chat\\_messages/](https://python.langchain.com/docs/modules/memory/chat_messages/)  
[https://python.langchain.com/docs/modules/data\\_connection/retrievers/multi\\_vector](https://python.langchain.com/docs/modules/data_connection/retrievers/multi_vector)  
[https://python.langchain.com/docs/modules/data\\_connection/document\\_transformers/](https://python.langchain.com/docs/modules/data_connection/document_transformers/)  
[https://python.langchain.com/docs/modules/data\\_connection/retrievers/MultiQueryRetriever](https://python.langchain.com/docs/modules/data_connection/retrievers/MultiQueryRetriever)  
<https://python.langchain.com/docs/integrations/vectorstores/chroma>  
[https://python.langchain.com/docs/modules/data\\_connection/retrievers/vectorstore](https://python.langchain.com/docs/modules/data_connection/retrievers/vectorstore)  
[https://python.langchain.com/docs/integrations/text\\_embedding/openai](https://python.langchain.com/docs/integrations/text_embedding/openai)  
[https://python.langchain.com/docs/integrations/document\\_loaders/grobid](https://python.langchain.com/docs/integrations/document_loaders/grobid)  
[https://python.langchain.com/docs/integrations/document\\_loaders/source\\_code](https://python.langchain.com/docs/integrations/document_loaders/source_code)  
[https://python.langchain.com/docs/modules/data\\_connection/document\\_transformers/markdown\\_header\\_metadata](https://python.langchain.com/docs/modules/data_connection/document_transformers/markdown_header_metadata)  
[https://python.langchain.com/docs/modules/data\\_connection/document\\_transformers/recursive\\_text\\_splitter](https://python.langchain.com/docs/modules/data_connection/document_transformers/recursive_text_splitter)  
[https://python.langchain.com/docs/integrations/document\\_loaders/web\\_base](https://python.langchain.com/docs/integrations/document_loaders/web_base)  
<https://python.langchain.com/docs/integrations/toolkits/csv>  
[https://python.langchain.com/docs/integrations/toolkits/sql\\_database](https://python.langchain.com/docs/integrations/toolkits/sql_database)  
<https://python.langchain.com/docs/integrations/toolkits/pandas>  
[https://python.langchain.com/docs/integrations/document\\_loaders/csv](https://python.langchain.com/docs/integrations/document_loaders/csv)  
[https://python.langchain.com/docs/integrations/document\\_loaders/pandas\\_dataframe](https://python.langchain.com/docs/integrations/document_loaders/pandas_dataframe)  
[https://python.langchain.com/docs/modules/callbacks/token\\_counting](https://python.langchain.com/docs/modules/callbacks/token_counting)  
[https://python.langchain.com/docs/modules/agents/tools/tools\\_as\\_openai\\_functions](https://python.langchain.com/docs/modules/agents/tools/tools_as_openai_functions)  
<https://python.langchain.com/docs/langsmith/walkthrough>  
[https://python.langchain.com/docs/modules/agents/agent\\_types/openai\\_functions\\_agent](https://python.langchain.com/docs/modules/agents/agent_types/openai_functions_agent)  
[https://python.langchain.com/docs/modules/agents/how\\_to/max\\_time\\_limit](https://python.langchain.com/docs/modules/agents/how_to/max_time_limit)  
[https://python.langchain.com/docs/modules/agents/how\\_to/intermediate\\_steps](https://python.langchain.com/docs/modules/agents/how_to/intermediate_steps)  
[https://python.langchain.com/docs/modules/agents/how\\_to/max\\_iterations](https://python.langchain.com/docs/modules/agents/how_to/max_iterations)  
[https://python.langchain.com/docs/modules/agents/how\\_to/handle\\_parsing\\_errors](https://python.langchain.com/docs/modules/agents/how_to/handle_parsing_errors)  
[https://python.langchain.com/docs/modules/agents/how\\_to/agent\\_iter](https://python.langchain.com/docs/modules/agents/how_to/agent_iter)  
[https://python.langchain.com/docs/modules/agents/how\\_to/agent\\_structured](https://python.langchain.com/docs/modules/agents/how_to/agent_structured)  
[https://python.langchain.com/docs/modules/agents/how\\_to/streaming](https://python.langchain.com/docs/modules/agents/how_to/streaming)  
[https://python.langchain.com/docs/modules/agents/how\\_to/custom\\_agent](https://python.langchain.com/docs/modules/agents/how_to/custom_agent)  
<https://python.langchain.com/docs/modules/agents/tools/>  
[https://python.langchain.com/docs/modules/agents/agent\\_types/](https://python.langchain.com/docs/modules/agents/agent_types/)  
[https://python.langchain.com/docs/modules/agents/quick\\_start](https://python.langchain.com/docs/modules/agents/quick_start)  
<https://python.langchain.com/docs/modules/agents/concepts>  
[https://python.langchain.com/docs/modules/model\\_io/output\\_parsers/types/yaml](https://python.langchain.com/docs/modules/model_io/output_parsers/types/yaml)  
[https://python.langchain.com/docs/modules/data\\_connection/retrievers/ensemble](https://python.langchain.com/docs/modules/data_connection/retrievers/ensemble)  
[https://python.langchain.com/docs/integrations/retrievers/self\\_query/](https://python.langchain.com/docs/integrations/retrievers/self_query/)  
[https://python.langchain.com/docs/modules/data\\_connection/retrievers/parent\\_document\\_retriever](https://python.langchain.com/docs/modules/data_connection/retrievers/parent_document_retriever)  
[https://python.langchain.com/docs/modules/data\\_connection/indexing](https://python.langchain.com/docs/modules/data_connection/indexing)  
[https://python.langchain.com/docs/modules/model\\_io/output\\_parsers/](https://python.langchain.com/docs/modules/model_io/output_parsers/)  
[https://python.langchain.com/docs/modules/model\\_io/quick\\_start](https://python.langchain.com/docs/modules/model_io/quick_start)  
[https://python.langchain.com/docs/modules/model\\_io/prompts/](https://python.langchain.com/docs/modules/model_io/prompts/)  
[https://python.langchain.com/docs/modules/model\\_io/concepts](https://python.langchain.com/docs/modules/model_io/concepts)  
<https://python.langchain.com/docs/modules/memory/>  
[https://python.langchain.com/docs/expression\\_language/cookbook/tools](https://python.langchain.com/docs/expression_language/cookbook/tools)  
[https://python.langchain.com/docs/expression\\_language/cookbook/prompt\\_size](https://python.langchain.com/docs/expression_language/cookbook/prompt_size)

[https://python.langchain.com/docs/expression\\_language/cookbook/moderation](https://python.langchain.com/docs/expression_language/cookbook/moderation)  
[https://python.langchain.com/docs/expression\\_language/cookbook/memory](https://python.langchain.com/docs/expression_language/cookbook/memory)  
[https://python.langchain.com/docs/expression\\_language/cookbook/embedding\\_router](https://python.langchain.com/docs/expression_language/cookbook/embedding_router)  
[https://python.langchain.com/docs/expression\\_language/cookbook/code\\_writing](https://python.langchain.com/docs/expression_language/cookbook/code_writing)  
[https://python.langchain.com/docs/expression\\_language/cookbook/agent](https://python.langchain.com/docs/expression_language/cookbook/agent)  
[https://python.langchain.com/docs/expression\\_language/cookbook/sql\\_db](https://python.langchain.com/docs/expression_language/cookbook/sql_db)  
[https://python.langchain.com/docs/expression\\_language/cookbook/multiple\\_chains](https://python.langchain.com/docs/expression_language/cookbook/multiple_chains)  
[https://python.langchain.com/docs/expression\\_language/cookbook/prompt\\_llm\\_parser](https://python.langchain.com/docs/expression_language/cookbook/prompt_llm_parser)  
[https://python.langchain.com/docs/expression\\_language/cookbook/retrieval](https://python.langchain.com/docs/expression_language/cookbook/retrieval)  
[https://python.langchain.com/docs/expression\\_language/how\\_to/message\\_history](https://python.langchain.com/docs/expression_language/how_to/message_history)  
[https://python.langchain.com/docs/expression\\_language/how\\_to/inspect](https://python.langchain.com/docs/expression_language/how_to/inspect)  
[https://python.langchain.com/docs/expression\\_language/how\\_to/generators](https://python.langchain.com/docs/expression_language/how_to/generators)  
[https://python.langchain.com/docs/expression\\_language/how\\_to/fallbacks](https://python.langchain.com/docs/expression_language/how_to/fallbacks)  
[https://python.langchain.com/docs/expression\\_language/how\\_to/decorator](https://python.langchain.com/docs/expression_language/how_to/decorator)  
[https://python.langchain.com/docs/expression\\_language/how\\_to/configure](https://python.langchain.com/docs/expression_language/how_to/configure)  
[https://python.langchain.com/docs/expression\\_language/how\\_to/binding](https://python.langchain.com/docs/expression_language/how_to/binding)  
[https://python.langchain.com/docs/expression\\_language/how\\_to/routing](https://python.langchain.com/docs/expression_language/how_to/routing)  
[https://python.langchain.com/docs/expression\\_language/how\\_to/map](https://python.langchain.com/docs/expression_language/how_to/map)  
[https://python.langchain.com/docs/expression\\_language/how\\_to/functions](https://python.langchain.com/docs/expression_language/how_to/functions)  
[https://python.langchain.com/docs/expression\\_language/how\\_to/passthrough](https://python.langchain.com/docs/expression_language/how_to/passthrough)  
<https://python.langchain.com/docs/integrations/vectorstores/faiss>  
<https://python.langchain.com/docs/integrations/retrievers/tavily>  
<https://python.langchain.com/docs/templates/xml-agent>  
<https://python.langchain.com/docs/templates/vertexai-chuck-norris>  
<https://python.langchain.com/docs/templates/summarize-anthropic>  
<https://python.langchain.com/docs/templates/stepback-qa-prompting>  
<https://python.langchain.com/docs/templates/sql-pgvector>  
<https://python.langchain.com/docs/templates/sql-research-assistant>  
<https://python.langchain.com/docs/templates/sql-ollama>  
<https://python.langchain.com/docs/templates/sql-llamacpp>  
<https://python.langchain.com/docs/templates/solo-performance-prompting-agent>  
<https://python.langchain.com/docs/templates/sql-llama2>  
<https://python.langchain.com/docs/templates/robocorp-action-server>  
<https://python.langchain.com/docs/templates/skeleton-of-thought>  
<https://python.langchain.com/docs/templates/self-query-qdrant>  
<https://python.langchain.com/docs/templates/self-query-supabase>  
<https://python.langchain.com/docs/templates/retrieval-agent>  
<https://python.langchain.com/docs/templates/rewrite-retrieve-read>  
<https://python.langchain.com/docs/templates/research-assistant>  
<https://python.langchain.com/docs/templates/rag-vectara>  
<https://python.langchain.com/docs/templates/rag-vectara-multiquery>  
<https://python.langchain.com/docs/templates/rag-weaviate>  
<https://python.langchain.com/docs/templates/rag-timescale-hybrid-search-time>  
<https://python.langchain.com/docs/templates/rag-supabase>  
<https://python.langchain.com/docs/templates/rag-timescale-conversation>  
<https://python.langchain.com/docs/templates/rag-self-query>  
<https://python.langchain.com/docs/templates/rag-singlestoredb>  
<https://python.langchain.com/docs/templates/rag-pinecone>  
<https://python.langchain.com/docs/templates/rag-semi-structured>  
<https://python.langchain.com/docs/templates/rag-pinecone-rerank>  
<https://python.langchain.com/docs/templates/rag-redis>  
<https://python.langchain.com/docs/templates/rag-pinecone-multi-query>

<https://python.langchain.com/docs/templates/rag-ollama-multi-query>  
<https://python.langchain.com/docs/templates/rag-opensearch>  
<https://python.langchain.com/docs/templates/rag-multi-modal-mv-local>  
<https://python.langchain.com/docs/templates/rag-multi-modal-local>  
<https://python.langchain.com/docs/templates/rag-multi-index-router>  
<https://python.langchain.com/docs/templates/rag-multi-index-fusion>  
<https://python.langchain.com/docs/templates/rag-mongo>  
<https://python.langchain.com/docs/templates/rag-matching-engine>  
<https://python.langchain.com/docs/templates/rag-momento-vector-index>  
<https://python.langchain.com/docs/templates/rag-gpt-crawler>  
<https://python.langchain.com/docs/templates/rag-gemini-multi-modal>  
<https://python.langchain.com/docs/templates/rag-google-cloud-sensitive-data-protection>  
<https://python.langchain.com/docs/templates/rag-elasticsearch>  
<https://python.langchain.com/docs/templates/rag-google-cloud-vertexai-search>  
<https://python.langchain.com/docs/templates/rag-fusion>  
<https://python.langchain.com/docs/templates/rag-conversation>  
<https://python.langchain.com/docs/templates/rag-conversation-zep>  
<https://python.langchain.com/docs/templates/rag-codellama-fireworks>  
<https://python.langchain.com/docs/templates/rag-chroma>  
<https://python.langchain.com/docs/templates/rag-chroma-private>  
<https://python.langchain.com/docs/templates/rag-chroma-multi-modal>  
<https://python.langchain.com/docs/templates/rag-chroma-multi-modal-multi-vector>  
<https://python.langchain.com/docs/templates/rag-aws-kendra>  
<https://python.langchain.com/docs/templates/rag-aws-bedrock>  
<https://python.langchain.com/docs/templates/python-lint>  
<https://python.langchain.com/docs/templates/rag-astradb>  
<https://python.langchain.com/docs/templates/openai-functions-tool-retrieval-agent>  
<https://python.langchain.com/docs/templates/pirate-speak>  
<https://python.langchain.com/docs/templates/pirate-speak-configurable>  
<https://python.langchain.com/docs/templates/plate-chain>  
<https://python.langchain.com/docs/templates/propositional-retrieval>  
<https://python.langchain.com/docs/templates/pii-protected-chatbot>  
<https://python.langchain.com/docs/templates/openai-functions-agent>  
<https://python.langchain.com/docs/templates/openai-functions-agent-gmail>  
<https://python.langchain.com/docs/templates/nvidia-rag-canonical>  
<https://python.langchain.com/docs/templates/neo4j-generation>  
<https://python.langchain.com/docs/templates/neo4j-parent>  
<https://python.langchain.com/docs/templates/neo4j-semantic-layer>  
<https://python.langchain.com/docs/templates/neo4j-vector-memory>  
<https://python.langchain.com/docs/templates/neo4j-cypher>  
<https://python.langchain.com/docs/templates/neo4j-cypher-memory>  
<https://python.langchain.com/docs/templates/neo4j-advanced-rag>  
<https://python.langchain.com/docs/templates/neo4j-cypher-ft>  
<https://python.langchain.com/docs/templates/mongo-parent-document-retrieval>  
<https://python.langchain.com/docs/templates/llama2-functions>  
<https://python.langchain.com/docs/templates/hyde>  
<https://python.langchain.com/docs/templates/hybrid-search-weaviate>  
<https://python.langchain.com/docs/templates/guardrails-output-parser>  
<https://python.langchain.com/docs/templates/extraction-openai-functions>  
<https://python.langchain.com/docs/templates/elastic-query-generator>  
<https://python.langchain.com/docs/templates/extraction-anthropic-functions>  
<https://python.langchain.com/docs/templates/csv-agent>

<https://python.langchain.com/docs/templates/cohere-librarian>  
<https://python.langchain.com/docs/templates/chat-bot-feedback>  
<https://python.langchain.com/docs/templates/cassandra-synonym-caching>  
<https://python.langchain.com/docs/templates/chain-of-note-wiki>  
<https://python.langchain.com/docs/templates/cassandra-entomology-rag>  
<https://python.langchain.com/docs/templates/bedrock-jcvd>  
<https://python.langchain.com/docs/templates/basic-critique-revise>  
<https://python.langchain.com/docs/templates/anthropic-iterative-search>  
<https://python.langchain.com/docs/contributing/code>  
<https://python.langchain.com/docs/contributing/integrations>  
<https://python.langchain.com/docs/contributing/faq>  
<https://python.langchain.com/docs/contributing/testing>  
<https://python.langchain.com/docs/contributing/documentation>  
<https://python.langchain.com/docs/changelog/langchain>  
<https://python.langchain.com/docs/modules/callbacks/>  
<https://python.langchain.com/docs/modules/callbacks/filecallbackhandler>  
<https://python.langchain.com/docs/changelog/core>  
[https://python.langchain.com/docs/integrations/providers/wandb\\_tracing](https://python.langchain.com/docs/integrations/providers/wandb_tracing)  
<https://python.langchain.com/docs/guides/safety/>  
[https://python.langchain.com/docs/guides/model\\_laboratory](https://python.langchain.com/docs/guides/model_laboratory)  
[https://python.langchain.com/docs/guides/pydantic\\_compatibility](https://python.langchain.com/docs/guides/pydantic_compatibility)  
[https://python.langchain.com/docs/guides/local\\_llms](https://python.langchain.com/docs/guides/local_llms)  
<https://python.langchain.com/docs/guides/fallbacks>  
[https://python.langchain.com/docs/integrations/providers/activeloop\\_deeplake](https://python.langchain.com/docs/integrations/providers/activeloop_deeplake)  
<https://python.langchain.com/docs/guides/evaluation/>  
<https://python.langchain.com/docs/guides/deployments/>  
<https://python.langchain.com/docs/integrations/stores/>  
<https://python.langchain.com/docs/integrations/adapters>  
[https://python.langchain.com/docs/integrations/chat\\_loaders](https://python.langchain.com/docs/integrations/chat_loaders)  
<https://python.langchain.com/docs/integrations/callbacks>  
<https://python.langchain.com/docs/integrations/memory>  
<https://python.langchain.com/docs/integrations/tools>  
<https://python.langchain.com/docs/integrations/toolkits>  
<https://python.langchain.com/docs/integrations/retrievers>  
<https://python.langchain.com/docs/integrations/vectorstores>  
[https://python.langchain.com/docs/integrations/text\\_embedding](https://python.langchain.com/docs/integrations/text_embedding)  
[https://python.langchain.com/docs/integrations/document\\_loaders](https://python.langchain.com/docs/integrations/document_loaders)  
[https://python.langchain.com/docs/integrations/document\\_transformers](https://python.langchain.com/docs/integrations/document_transformers)  
<https://python.langchain.com/docs/integrations/chat/>  
<https://python.langchain.com/docs/integrations/llms/>  
<https://python.langchain.com/docs/integrations/platforms/microsoft>  
<https://python.langchain.com/docs/integrations/components>  
<https://python.langchain.com/docs/integrations/platforms/openai>  
<https://python.langchain.com/docs/integrations/platforms/huggingface>  
<https://python.langchain.com/docs/integrations/platforms/google>  
<https://python.langchain.com/docs/integrations/platforms/aws>  
<https://python.langchain.com/docs/integrations/platforms/anthropic>  
[https://python.langchain.com/docs/modules/model\\_io/llms/](https://python.langchain.com/docs/modules/model_io/llms/)  
[https://python.langchain.com/docs/modules/model\\_io/chat](https://python.langchain.com/docs/modules/model_io/chat)  
[https://python.langchain.com/docs/modules/data\\_connection/retrievers/](https://python.langchain.com/docs/modules/data_connection/retrievers/)  
[https://python.langchain.com/docs/modules/data\\_connection/vectorstores/](https://python.langchain.com/docs/modules/data_connection/vectorstores/)  
[https://python.langchain.com/docs/modules/data\\_connection/text\\_embedding/](https://python.langchain.com/docs/modules/data_connection/text_embedding/)

[https://python.langchain.com/docs/modules/data\\_connection/document\\_transformers/](https://python.langchain.com/docs/modules/data_connection/document_transformers/)  
[https://python.langchain.com/docs/modules/data\\_connection/document\\_loaders/](https://python.langchain.com/docs/modules/data_connection/document_loaders/)  
[https://python.langchain.com/docs/use\\_cases/data\\_generation](https://python.langchain.com/docs/use_cases/data_generation)  
[https://python.langchain.com/docs/use\\_cases/graph/](https://python.langchain.com/docs/use_cases/graph/)  
[https://python.langchain.com/docs/use\\_cases/code\\_understanding](https://python.langchain.com/docs/use_cases/code_understanding)  
[https://python.langchain.com/docs/use\\_cases/web\\_scraping](https://python.langchain.com/docs/use_cases/web_scraping)  
[https://python.langchain.com/docs/use\\_cases/tagging](https://python.langchain.com/docs/use_cases/tagging)  
[https://python.langchain.com/docs/use\\_cases/summarization](https://python.langchain.com/docs/use_cases/summarization)  
[https://python.langchain.com/docs/use\\_cases/extraction](https://python.langchain.com/docs/use_cases/extraction)  
[https://python.langchain.com/docs/use\\_cases/apis](https://python.langchain.com/docs/use_cases/apis)  
[https://python.langchain.com/docs/use\\_cases/question\\_answering/local\\_retrieval\\_qa](https://python.langchain.com/docs/use_cases/question_answering/local_retrieval_qa)  
[https://python.langchain.com/docs/use\\_cases/question\\_answering/conversational\\_retrieval\\_agents](https://python.langchain.com/docs/use_cases/question_answering/conversational_retrieval_agents)  
[https://python.langchain.com/docs/use\\_cases/question\\_answering/streaming](https://python.langchain.com/docs/use_cases/question_answering/streaming)  
[https://python.langchain.com/docs/use\\_cases/question\\_answering/per\\_user](https://python.langchain.com/docs/use_cases/question_answering/per_user)  
[https://python.langchain.com/docs/use\\_cases/question\\_answering/chat\\_history](https://python.langchain.com/docs/use_cases/question_answering/chat_history)  
[https://python.langchain.com/docs/use\\_cases/question\\_answering/quickstart](https://python.langchain.com/docs/use_cases/question_answering/quickstart)  
[https://python.langchain.com/docs/use\\_cases/question\\_answering/sources](https://python.langchain.com/docs/use_cases/question_answering/sources)  
<https://python.langchain.com/docs/contributing>  
[https://python.langchain.com/docs/use\\_cases/qa\\_structured/sql/](https://python.langchain.com/docs/use_cases/qa_structured/sql/)  
[https://python.langchain.com/docs/use\\_cases/chatbots/](https://python.langchain.com/docs/use_cases/chatbots/)  
<https://python.langchain.com/docs/langgraph>  
[https://python.langchain.com/docs/use\\_cases/question\\_answering/](https://python.langchain.com/docs/use_cases/question_answering/)  
<https://python.langchain.com/docs/langserve>  
<https://python.langchain.com/docs/langsmith/>  
<https://python.langchain.com/docs/modules/chains>  
<https://python.langchain.com/docs/modules/agents/>  
[https://python.langchain.com/docs/modules/data\\_connection/](https://python.langchain.com/docs/modules/data_connection/)  
[https://python.langchain.com/docs/modules/model\\_io/](https://python.langchain.com/docs/modules/model_io/)  
<https://python.langchain.com/docs/modules/>  
[https://python.langchain.com/docs/expression\\_language/cookbook](https://python.langchain.com/docs/expression_language/cookbook)  
[https://python.langchain.com/docs/expression\\_language/why](https://python.langchain.com/docs/expression_language/why)  
[https://python.langchain.com/docs/expression\\_language/how\\_to](https://python.langchain.com/docs/expression_language/how_to)  
[https://python.langchain.com/docs/expression\\_language/interface](https://python.langchain.com/docs/expression_language/interface)  
[https://python.langchain.com/docs/expression\\_language/get\\_started](https://python.langchain.com/docs/expression_language/get_started)  
[https://python.langchain.com/docs/expression\\_language/](https://python.langchain.com/docs/expression_language/)  
<https://python.langchain.com/docs/security>  
[https://python.langchain.com/docs/get\\_started/quickstart](https://python.langchain.com/docs/get_started/quickstart)  
[https://python.langchain.com/docs/get\\_started/installation](https://python.langchain.com/docs/get_started/installation)  
[https://python.langchain.com/docs/get\\_started](https://python.langchain.com/docs/get_started)  
[https://python.langchain.com/docs/additional\\_resources/youtube](https://python.langchain.com/docs/additional_resources/youtube)  
[https://python.langchain.com/docs/additional\\_resources/tutorials](https://python.langchain.com/docs/additional_resources/tutorials)  
<https://python.langchain.com/docs/templates/>  
<https://python.langchain.com/docs/contributing>  
<https://python.langchain.com/docs/changelog>  
<https://python.langchain.com/docs/packages>  
<https://python.langchain.com/docs/guides/debugging>  
<https://python.langchain.com/docs/integrations/providers>  
[https://python.langchain.com/docs/use\\_cases/question\\_answering/](https://python.langchain.com/docs/use_cases/question_answering/)  
[https://python.langchain.com/docs/get\\_started/introduction#\\_\\_docusaurus\\_skipToContent\\_fallback](https://python.langchain.com/docs/get_started/introduction#__docusaurus_skipToContent_fallback)  
[https://python.langchain.com/docs/get\\_started/introduction](https://python.langchain.com/docs/get_started/introduction)

html























oadersacremoAirbyte CDKAirbyte GongAirbyte HubspotAirbyte JSONAirbyte SalesforceAirbyte Shopify  
afetyAmazon Comprehend Moderation ChainConstitutional chainHugging Face prompt injection identific  
oadersacremoAirbyte CDKAirbyte GongAirbyte HubspotAirbyte JSONAirbyte SalesforceAirbyte Shopify  
oadersDocument transformersText embedding modelsAleph AlphaAwaDBAzure OpenAI Baidu QianfanE  
oadersDocument transformersText embedding modelsAleph AlphaAwaDBAzure OpenAI Baidu QianfanE  
ud PAI EASAnthropicAnthropic FunctionsAnyscaleAzure OpenAI Azure ML EndpointBaichuan ChatBaid  
n API GatewayAnyscaleAphrodite EngineArceeAzure MLAzure OpenAI Baidu QianfanBananaBasetenBe  
ELInterfaceHow toCookbookModulesModel I/OConceptsPromptsQuickstartLLMsQuick StartCustom LLM  
ELInterfaceHow toCookbookModulesModel I/OConceptsPromptsQuickstartLLMsQuick StartCustom LLM  
n API GatewayAnyscaleAphrodite EngineArceeAzure MLAzure OpenAI Baidu QianfanBananaBasetenBe  
n API GatewayAnyscaleAphrodite EngineArceeAzure MLAzure OpenAI Baidu QianfanBananaBasetenBe  
ELInterfaceHow toCookbookModulesModel I/OConceptsPromptsQuickstartLLMsQuick StartCustom LLM  
ELInterfaceHow toCookbookModulesModel I/OConceptsPromptsQuickstartLLMsQuick StartCustom LLM  
ELInterfaceHow toCookbookModulesModel I/OConceptsPromptsQuickstartLLMsQuick StartCustom LLM  
ELInterfaceHow toCookbookModulesModel I/OConceptsPromptsQuickstartLLMsChat ModelsQuick Start  
ELInterfaceHow toCookbookModulesModel I/ORetrievalDocument loadersText SplittersText embedding  
ELInterfaceHow toCookbookModulesModel I/ORetrievalDocument loadersText SplittersText embedding  
ELInterfaceHow toCookbookModulesModel I/ORetrievalDocument loadersText SplittersText embedding  
ELInterfaceHow toCookbookModulesModel I/ORetrievalDocument loadersText SplittersHTMLHeaderTex  
ELInterfaceHow toCookbookModulesModel I/ORetrievalDocument loadersText SplittersHTMLHeaderTex  
ELInterfaceHow toCookbookModulesModel I/ORetrievalDocument loadersText SplittersHTMLHeaderTex  
ELInterfaceHow toCookbookModulesModel I/ORetrievalDocument loadersCSVFile DirectoryHTMLJSON  
gingWeb scrapingCode understandingSynthetic data generationGraph queryingDiffbot Graph Transform  
ou to the original URL and let them know their link is broken.CommunityDiscordTwitterGitHubPythonJS/1  
ou to the original URL and let them know their link is broken.CommunityDiscordTwitterGitHubPythonJS/1  
ou to the original URL and let them know their link is broken.CommunityDiscordTwitterGitHubPythonJS/1  
ou to the original URL and let them know their link is broken.CommunityDiscordTwitterGitHubPythonJS/1  
oadersDocument transformersBeautiful SoupGoogle Cloud Document AI Doctran: extract propertiesDoc  
ud PAI EASAnthropicAnthropic FunctionsAnyscaleAzure OpenAI Azure ML EndpointBaichuan ChatBaid

ELInterfaceHow toCookbookModulesModel I/ORetrievalAgentsChainsMoreLangServeLangSmithLangGr  
ELInterfaceHow toCookbookModulesModel I/ORetrievalAgentsChainsMoreLangServeLangSmithLangGr  
ud PAI EASAnthropicAnthropic FunctionsAnyscaleAzure OpenAIAzure ML EndpointBaichuan ChatBaid  
ELInterfaceHow toCookbookModulesModel I/ORetrievalAgentsChainsMoreLangServeLangSmithLangGr  
n API GatewayAnyscaleAphrodite EngineArceeAzure MLAzure OpenAIBaidu QianfanBananaBasetenBe  
n API GatewayAnyscaleAphrodite EngineArceeAzure MLAzure OpenAIBaidu QianfanBananaBasetenBe  
n API GatewayAnyscaleAphrodite EngineArceeAzure MLAzure OpenAIBaidu QianfanBananaBasetenBe  
ELInterfaceHow toCookbookModulesModel I/ORetrievalAgentsChainsMoreMemoryChat MessagesMem  
ELInterfaceHow toCookbookModulesModel I/ORetrievalDocument loadersText SplittersText embedding  
ELInterfaceHow toCookbookModulesModel I/ORetrievalDocument loadersText SplittersHTMLHeaderTex  
ELInterfaceHow toCookbookModulesModel I/ORetrievalDocument loadersText SplittersText embedding  
oadersDocument transformersText embedding modelsVector storesActiveloop Deep LakeAlibaba Cloud  
ELInterfaceHow toCookbookModulesModel I/ORetrievalDocument loadersText SplittersText embedding  
oadersDocument transformersText embedding modelsAleph AlphaAwaDBAzure OpenAIBaidu QianfanE  
oadersacreamAirbyte CDKAirbyte GongAirbyte HubspotAirbyte JSONAirbyte SalesforceAirbyte Shopify  
oadersacreamAirbyte CDKAirbyte GongAirbyte HubspotAirbyte JSONAirbyte SalesforceAirbyte Shopify  
ELInterfaceHow toCookbookModulesModel I/ORetrievalDocument loadersText SplittersHTMLHeaderTex  
ELInterfaceHow toCookbookModulesModel I/ORetrievalDocument loadersText SplittersHTMLHeaderTex  
oadersacreamAirbyte CDKAirbyte GongAirbyte HubspotAirbyte JSONAirbyte SalesforceAirbyte Shopify  
oadersDocument transformersText embedding modelsVector storesRetrieversToolsAgents and toolkits/  
oadersDocument transformersText embedding modelsVector storesRetrieversToolsAgents and toolkits/  
oadersDocument transformersText embedding modelsVector storesRetrieversToolsAgents and toolkits/  
oadersacreamAirbyte CDKAirbyte GongAirbyte HubspotAirbyte JSONAirbyte SalesforceAirbyte Shopify  
oadersacreamAirbyte CDKAirbyte GongAirbyte HubspotAirbyte JSONAirbyte SalesforceAirbyte Shopify  
ELInterfaceHow toCookbookModulesModel I/ORetrievalAgentsChainsMoreMemoryCallbacksAsync callb  
ELInterfaceHow toCookbookModulesModel I/ORetrievalAgentsQuickstartConceptsAgent TypesHow-toTr  
ELInterfaceHow toCookbookModulesModel I/ORetrievalAgentsChainsMoreLangServeLangSmithLangSm  
ELInterfaceHow toCookbookModulesModel I/ORetrievalAgentsQuickstartConceptsAgent TypesOpenAI f  
ELInterfaceHow toCookbookModulesModel I/ORetrievalAgentsQuickstartConceptsAgent TypesHow-toCo  
ELInterfaceHow toCookbookModulesModel I/OConceptsPromptsQuickstartLLMsChat ModelsOutput Par  
ELInterfaceHow toCookbookModulesModel I/ORetrievalDocument loadersText SplittersText embedding  
oadersDocument transformersText embedding modelsVector storesRetrieversActiveloop Deep Memory  
ELInterfaceHow toCookbookModulesModel I/ORetrievalDocument loadersText SplittersText embedding  
ELInterfaceHow toCookbookModulesModel I/OConceptsPromptsQuickstartLLMsChat ModelsOutput Par  
ELInterfaceHow toCookbookModulesModel I/OConceptsPromptsQuickstartLLMsChat ModelsOutput Par  
ELInterfaceHow toCookbookModulesModel I/OConceptsPromptsQuick StartCompositionExample Select  
ELInterfaceHow toCookbookModulesModel I/OConceptsPromptsQuickstartLLMsChat ModelsOutput Par  
ELInterfaceHow toCookbookModulesModel I/ORetrievalAgentsChainsMoreMemoryChat MessagesMem  
ELInterfaceHow toCookbookPrompt + LLMRAGMultiple chainsQuerying a SQL DBAgentsCode writingRe







ELInterfaceHow toCookbookModulesModel I/OREtrievalDocument loadersText SplittersHTMLHeaderTeX  
ELInterfaceHow toCookbookModulesModel I/OREtrievalDocument loadersCSVFile DirectoryHTMLJSON  
gingWeb scrapingCode understandingSynthetic data generationGraph queryingUse casesSynthetic data  
gingWeb scrapingCode understandingSynthetic data generationGraph queryingDiffbot Graph Transform  
gingWeb scrapingCode understandingSynthetic data generationGraph queryingUse casesCode underst  
gingWeb scrapingCode understandingSynthetic data generationGraph queryingUse casesWeb scraping  
gingWeb scrapingCode understandingSynthetic data generationGraph queryingUse casesTaggingTaggi  
gingWeb scrapingCode understandingSynthetic data generationGraph queryingUse casesSummarizatio  
gingWeb scrapingCode understandingSynthetic data generationGraph queryingUse casesExtractionExti  
gingWeb scrapingCode understandingSynthetic data generationGraph queryingUse casesInteracting wil  
itsUsing local modelsQ&A over structured dataInteracting with APIsChatbotsExtractionSummarizationT  
ContributorsWelcome ContributorsHi there! Thank you for even being interested in contributing to LangC  
ITaggingWeb scrapingCode understandingSynthetic data generationGraph queryingUse casesQ&A ove  
gingWeb scrapingCode understandingSynthetic data generationGraph queryingUse casesChatbotsChat  
ELInterfaceHow toCookbookModulesModel I/OREtrievalAgentsChainsMoreLangServeLangSmithLangGr  
itsUsing local modelsQ&A over structured dataInteracting with APIsChatbotsExtractionSummarizationT  
ELInterfaceHow toCookbookModulesModel I/OREtrievalAgentsChainsMoreLangServeLangSmithLangGr  
ELInterfaceHow toCookbookModulesModel I/OREtrievalAgentsChainsMoreLangServeLangSmithLangGr  
ELInterfaceHow toCookbookModulesModel I/OREtrievalAgentsChainsMoreLangServeLangSmithLangGr  
ELInterfaceHow toCookbookModulesModel I/OREtrievalAgentsQuickstartConceptsAgent TypesHow-toT  
ELInterfaceHow toCookbookModulesModel I/OREtrievalDocument loadersText SplittersText embedding  
ELInterfaceHow toCookbookModulesModel I/OConceptsPromptsQuickstartLLMsChat ModelsOutput Par  
ELInterfaceHow toCookbookModulesModel I/OREtrievalAgentsChainsMoreLangServeLangSmithLangGr  
ELInterfaceHow toCookbookPrompt + LLMRAGMultiple chainsQuerying a SQL DBAgentsCode writingRe  
ELInterfaceHow toCookbookModulesModel I/OREtrievalAgentsChainsMoreLangServeLangSmithLangGr  
ELInterfaceHow toRunnableParallel: Manipulating dataRunnablePassthrough: Passing data throughRunn  
ELInterfaceHow toCookbookModulesModel I/OREtrievalAgentsChainsMoreLangServeLangSmithLangGr  
roduction to LangChain with Harrison Chase, creator of LangChainBuilding the Future with LLMs, LangC  
langChain, check out the use cases guides.  icon marks a new addition [last update 2023-09-21]LangC  
ndra-synonym-cachingChain-of-Note (Wikipedia)Chat Bot Feedback Templatecohere-librarianscsv-agent  
ContributorsWelcome ContributorsHi there! Thank you for even being interested in contributing to LangC  
main0.1.0 (Jan 5, 2024)Nextlangchain-coreCommunityDiscordTwitterGitHubPythonJS/TSMoreHomepage  
ancy by a maintainer and published to PyPI. The different packages are versioned slightly differently.lang  
afetyDebuggingDebuggingIf you're building with LLMs, at some point something will break, and you'll ne  
oadersDocument transformersText embedding modelsVector storesRetrieversToolsAgents and toolkitsM  
itsUsing local modelsQ&A over structured dataInteracting with APIsChatbotsExtractionSummarizationT  
ELInterfaceHow toCookbookModulesModel I/OREtrievalAgentsChainsMoreLangServeLangSmithLangGr  
ELInterfaceHow toCookbookModulesModel I/OREtrievalAgentsChainsMoreLangServeLangSmithLangGr

Airbyte StripeAirbyte TypeformAirbyte Zendesk SupportAirtable Alibaba Cloud MaxComputeAmazon Te:  
ationLogical Fallacy chainModeration chainSafetyAmazon Comprehend Moderation ChainAmazon Cor:  
Airbyte StripeAirbyte TypeformAirbyte Zendesk SupportAirtable Alibaba Cloud MaxComputeAmazon Te:  
BedrockBGE on Hugging FaceBookend AIClarifaiCloudflare Workers AICohereDashScopeDeepInfraED  
BedrockBGE on Hugging FaceBookend AIClarifaiCloudflare Workers AICohereDashScopeDeepInfraED  
lu QianfanBedrock ChatCohereErnie Bot ChatEverlyAIFireworksGigaChatGoogle AIGoogle Cloud Verte  
eamBedrockBittensorCerebriumAIChatGLMClarifaiCloudflare Workers AICohereC TransformersCTrans  
ICachingStreamingTracking token usageChat ModelsOutput ParsersRetrievalAgentsChainsMoreLangS  
ICachingStreamingTracking token usageChat ModelsOutput ParsersRetrievalAgentsChainsMoreLangS  
eamBedrockBittensorCerebriumAIChatGLMClarifaiCloudflare Workers AICohereC TransformersCTrans  
eamBedrockBittensorCerebriumAIChatGLMClarifaiCloudflare Workers AICohereC TransformersCTrans  
ICachingStreamingTracking token usageChat ModelsOutput ParsersRetrievalAgentsChainsMoreLangS  
ICachingStreamingTracking token usageChat ModelsOutput ParsersRetrievalAgentsChainsMoreLangS  
ICachingStreamingTracking token usageChat ModelsOutput ParsersRetrievalAgentsChainsMoreLangS  
tCachingStreamingTracking token usageOutput ParsersRetrievalAgentsChainsMoreLangServeLangSm  
tCachingStreamingTracking token usageOutput ParsersRetrievalAgentsChainsMoreLangServeLangSm  
tCachingStreamingTracking token usageOutput ParsersRetrievalAgentsChainsMoreLangServeLangSm  
tCachingStreamingTracking token usageOutput ParsersRetrievalAgentsChainsMoreLangServeLangSm  
modelsVector storesRetrieversVector store-backed retrieverMultiQueryRetrieverContextual compressio  
modelsVector storesRetrieversVector store-backed retrieverMultiQueryRetrieverContextual compressio  
modelsVector storesRetrieversVector store-backed retrieverMultiQueryRetrieverContextual compressio  
modelsCacheBackedEmbeddingsVector storesRetrieversIndexingAgentsChainsMoreLangServeLangSm  
xtSplitterSplit by characterSplit codeMarkdownHeaderTextSplitterRecursively split by characterSemantic  
xtSplitterSplit by characterSplit codeMarkdownHeaderTextSplitterRecursively split by characterSemantic  
xtSplitterSplit by characterSplit codeMarkdownHeaderTextSplitterRecursively split by characterSemantic  
IMarkdownPDFText SplittersText embedding modelsVector storesRetrieversIndexingAgentsChainsMore  
ierArangoDB QA chainNeo4j DB QA chainFalkorDBQACChainHugeGraph QA ChainKuzuQAChainMemg  
TSMoreHomepageBlogCopyright © 2024 LangChain, Inc.  
Doctran: interrogate documentsDoctran: language translationGoogle TranslateHTML to textNucliaOpenAI n  
lu QianfanBedrock ChatCohereErnie Bot ChatEverlyAIFireworksGigaChatGoogle AIGoogle Cloud Verte

raphModulesChainsChainsChains refer to sequences of calls - whether to an LLM, a tool, or a data prep  
raphModulesChainsChainsChains refer to sequences of calls - whether to an LLM, a tool, or a data prep  
lu QianfanBedrock ChatCohereErnie Bot ChatEverlyAIFireworksGigaChatGoogle AI Google Cloud Verte  
raphModulesChainsChainsChains refer to sequences of calls - whether to an LLM, a tool, or a data prep  
eamBedrockBittensorCerebriumAIChatGLMClarifaiCloudflare Workers AI Cohere C Transformers C Trans  
eamBedrockBittensorCerebriumAIChatGLMClarifaiCloudflare Workers AI Cohere C Transformers C Trans  
eamBedrockBittensorCerebriumAIChatGLMClarifaiCloudflare Workers AI Cohere C Transformers C Trans  
ory typesMemory in LLMChainMemory in the Multi-Input ChainMemory in AgentMessage Memory in Ag  
modelsVector storesRetrieversVector store-backed retrieverMultiQueryRetrieverContextual compressio  
xtSplitterSplit by characterSplit codeMarkdownHeaderTextSplitterRecursively split by characterSemantic  
modelsVector storesRetrieversVector store-backed retrieverMultiQueryRetrieverContextual compressio  
d OpenSearchAnalyticDBAnnoyAstra DBAtlasAwaDBAzure Cosmos DB Azure AI SearchBagelDBBaidu  
modelsVector storesRetrieversVector store-backed retrieverMultiQueryRetrieverContextual compressio  
BedrockBGE on Hugging FaceBookend AI ClarifaiCloudflare Workers AI Cohere DashScopeDeepInfraED  
'Airbyte StripeAirbyte TypeformAirbyte Zendesk SupportAirtableAlibaba Cloud MaxComputeAmazon Te  
'Airbyte StripeAirbyte TypeformAirbyte Zendesk SupportAirtableAlibaba Cloud MaxComputeAmazon Te  
xtSplitterSplit by characterSplit codeMarkdownHeaderTextSplitterRecursively split by characterSemantic  
xtSplitterSplit by characterSplit codeMarkdownHeaderTextSplitterRecursively split by characterSemantic  
'Airbyte StripeAirbyte TypeformAirbyte Zendesk SupportAirtableAlibaba Cloud MaxComputeAmazon Te  
AINetworkAirbyte Question AnsweringAmadeusAzure Cognitive ServicesClickUpCSVDocument Compa  
AINetworkAirbyte Question AnsweringAmadeusAzure Cognitive ServicesClickUpCSVDocument Compa  
AINetworkAirbyte Question AnsweringAmadeusAzure Cognitive ServicesClickUpCSVDocument Compa  
'Airbyte StripeAirbyte TypeformAirbyte Zendesk SupportAirtableAlibaba Cloud MaxComputeAmazon Te  
'Airbyte StripeAirbyte TypeformAirbyte Zendesk SupportAirtableAlibaba Cloud MaxComputeAmazon Te  
backsCustom callback handlersCallbacks for custom chainsLogging to fileMultiple callback handlersTags  
oolsToolkitsDefining Custom ToolsTools as OpenAI FunctionsChainsMoreLangServeLangSmithLangGr  
nith WalkthroughLangGraphLangSmithLangSmith WalkthroughLangSmith WalkthroughOpen In ColabL  
functionsOpenAI toolsXML AgentJSON Chat AgentStructured chatReActSelf-ask with searchHow-toToo  
ustom agentStreamingRunning Agent as an IteratorReturning Structured OutputHandle parsing errorsA  
ustom agentStreamingRunning Agent as an IteratorReturning Structured OutputHandle parsing errorsA  
ustom agentStreamingRunning Agent as an IteratorReturning Structured OutputHandle parsing errorsA  
ustom agentStreamingRunning Agent as an IteratorReturning Structured OutputHandle parsing errorsA  
ustom agentStreamingRunning Agent as an IteratorReturning Structured OutputHandle parsing errorsA  
oolsToolkitsDefining Custom ToolsTools as OpenAI FunctionsChainsMoreLangServeLangSmithLangGr  
functionsOpenAI toolsXML AgentJSON Chat AgentStructured chatReActSelf-ask with searchHow-toToo  
oolsChainsMoreLangServeLangSmithLangGraphModulesAgentsQuickstartQuickstartTo best understand  
ersQuickstartTypesCSV parserDatetime parserEnum parserJSON parserOpenAI FunctionsOutput-fixir  
modelsVector storesRetrieversVector store-backed retrieverMultiQueryRetrieverContextual compressio  
'Amazon KendraArcee RetrieverArxivAzure Cognitive SearchBedrock (Knowledge Bases)BM25Chainde  
modelsVector storesRetrieversVector store-backed retrieverMultiQueryRetrieverContextual compressio  
modelsVector storesRetrieversIndexingAgentsChainsMoreLangServeLangSmithLangGraphModulesRe  
'ersQuickstartTypesRetrievalAgentsChainsMoreLangServeLangSmithLangGraphModulesModel I/O Ou  
'ersRetrievalAgentsChainsMoreLangServeLangSmithLangGraphModulesModel I/O QuickstartQuickstar  
tor TypesExample selectorsFew-shot prompt templatesFew-shot examples for chat modelsTypes of 'Me  
'ersRetrievalAgentsChainsMoreLangServeLangSmithLangGraphModulesModel I/O ConceptsConcepts  
ory typesMemory in LLMChainMemory in the Multi-Input ChainMemory in AgentMessage Memory in Ag  
outing by semantic similarityAdding memoryAdding moderationManaging prompt sizeUsing toolsModule  
outing by semantic similarityAdding memoryAdding moderationManaging prompt sizeUsing toolsModule





.elastic-query-generatorextraction-anthropic-functionsextraction-openai-functionsguardrails-output-parse  
.elastic-query-generatorextraction-anthropic-functionsextraction-openai-functionsguardrails-output-parse  
.elastic-query-generatorextraction-anthropic-functionsextraction-openai-functionsguardrails-output-parse  
.elastic-query-generatorextraction-anthropic-functionsextraction-openai-functionsguardrails-output-parse  
.elastic-query-generatorextraction-anthropic-functionsextraction-openai-functionsguardrails-output-parse  
.elastic-query-generatorextraction-anthropic-functionsextraction-openai-functionsguardrails-output-parse  
.elastic-query-generatorextraction-anthropic-functionsextraction-openai-functionsguardrails-output-parse  
.elastic-query-generatorextraction-anthropic-functionsextraction-openai-functionsguardrails-output-parse  
.elastic-query-generatorextraction-anthropic-functionsextraction-openai-functionsguardrails-output-parse  
lease do not try to push directly to this repo unless you are a maintainer. Please follow the checked-in pu  
le on Contributing Code. There are a few different places you can contribute integrations for LangChain:  
: a pull request, there may be additional changes necessary before merging it. Oftentimes, it is more effi  
sts. Unit tests run on every pull request, so they should be fast and reliable. Integration tests run once a d  
ce. Documentation is built using Quarto and Docusaurus 2. API Reference are largely autogenerated by :  
ctorDBChainConversationalRetrievalChainMore general to all retrieverscreate\_ernie\_fn\_chaincreate\_eri  
>acksCustom callback handlersCallbacks for custom chainsLogging to fileMultiple callback handlersTags  
>acksCustom callback handlersCallbacks for custom chainsLogging to fileMultiple callback handlersTags  
moved in 0.2.0. Use BaseChatModel.invoke instead. BaseChatModel methods apredict, apredict\_messa  
ivAstra DBAtlasAwaDBAZLyricsBagelDBBananaBasetenBeamBeautiful SoupBiliBiliNIBittensorBlackbo  
cationLogical Fallacy chainModeration chainSafetySafetyOne of the key concerns with using LLMs is tha  
ed choosing between many different options of prompts, models, and even chains to use. When doing s  
/dantic.dev/2.0/blog/pydantic-v2-final/)v2 contains has a number of breaking changes (<https://docs.pydai>  
nd GPT4All underscore the demand to run LLMs locally (on your own device). This has at least two impc  
underlying APIs, whether these be rate limiting or downtime. Therefore, as you go to move your LLM ap  
ivAstra DBAtlasAwaDBAZLyricsBagelDBBananaBasetenBeamBeautiful SoupBiliBiliNIBittensorBlackbo  
ng applications with language models involves many moving parts. One of the most critical components  
e of Large Language Models (LLMs) is rapidly expanding. As a result, it is crucial for developers to unde  
MemoryCallbacksChat loadersAdaptersStoresIn MemoryLocal FilesystemAstra DBRedisUpstash Redis  
MemoryCallbacksChat loadersAdaptersOpenAI Adapter(Old)OpenAI AdapterStoresComponentsAdapte  
MemoryCallbacksChat loadersDiscordFacebook MessengerGMailiMessageLangSmith Chat DatasetsLa  
MemoryCallbacksArgillaConfidentContextInfiniLabel StudioLLMonitorPromptLayerSageMaker Tracking  
MemoryAstra DBAWS DynamoDBCassandraElasticsearchMomento CacheMongoDBMotörheadNeo4jP  
/ArXivAWS LambdaShell (bash)Bearly Code InterpreterBing SearchBrave SearchChatGPT PluginsDall  
-AI NetworkAirbyte Question AnsweringAmadeusAzure Cognitive ServicesClickUpCSVDocument Compa  
/Amazon KendraArcee RetrieverArxivAzure Cognitive SearchBedrock (Knowledge Bases)BM25Chainde  
d OpenSearchAnalyticDBAnnoyAstra DBAtlasAwaDBAzure Cosmos DBAzure AI SearchBagelDBBaidu  
BedrockBGE on Hugging FaceBookend AIClarifaiCloudflare Workers AICohereDashScopeDeepInfraED  
/Airbyte StripeAirbyte TypeformAirbyte Zendesk SupportAirtableAlibaba Cloud MaxComputeAmazon Te  
:tran: interrogate documentsDoctran: language translationGoogle TranslateHTML to textNucliaOpenAI n  
lu QianfanBedrock ChatCohereErnie Bot ChatEverlyAIFireworksGigaChatGoogle AIGoogle Cloud Verte  
eamBedrockBittensorCerebriumAIChatGLMClarifaiCloudflare Workers AICohereC TransformersCTrans  
MemoryCallbacksChat loadersAdaptersStoresProvidersMicrosoftMicrosoftAll functionality related to Mic  
MemoryCallbacksChat loadersAdaptersStoresComponentsComponents LLMs78 items Cha  
MemoryCallbacksChat loadersAdaptersStoresProvidersOpenAIOpenAIAll functionality related to OpenA  
MemoryCallbacksChat loadersAdaptersStoresProvidersHugging FaceHugging FaceAll functionality relat  
MemoryCallbacksChat loadersAdaptersStoresProvidersGoogleGoogleAll functionality related to Google  
MemoryCallbacksChat loadersAdaptersStoresProvidersAWSAWSThe LangChain integrations related to  
MemoryCallbacksChat loadersAdaptersStoresProvidersAnthropicAnthropicAll functionality related to An  
!CachingStreamingTracking token usageChat ModelsOutput ParsersRetrievalAgentsChainsMoreLangS  
tCachingStreamingTracking token usageOutput ParsersRetrievalAgentsChainsMoreLangServeLangSm  
modelsVector storesRetrieversVector store-backed retrieverMultiQueryRetrieverContextual compressio  
modelsVector storesRetrieversIndexingAgentsChainsMoreLangServeLangSmithLangGraphModulesRe  
modelsCacheBackedEmbeddingsVector storesRetrieversIndexingAgentsChainsMoreLangServeLangSr

`xtSplitterSplit` by characterSplit codeMarkdownHeaderTextSplitterRecursively split by characterSemanticIMarkdownPDFText SplittersText embedding modelsVector storesRetrieversIndexingAgentsChainsMore a generationSynthetic data generationOpen In ColabUse caseSynthetic data is artificially generated dat:merArangoDB QA chainNeo4j DB QA chainFalkorDBQAChainHugeGraph QA ChainKuzuQACChainMemg tandingCode understandingOpen In CollabUse caseSource code analysis is one of the most popular LLI gWeb scrapingOpen In CollabUse caseWeb research is one of the killer LLM applications:Users have hi:ngOpen In ColabUse caseTagging means labeling a document with classes such as:sentimentlanguage: onSummarizationOpen In ColabUse caseSuppose you have a set of documents (PDFs, Notion pages, cractionOpen In CollabUse caseGetting structured output from raw LLM generations is hard.For example th APIsInteracting with APIsOpen In ColabUse caseSuppose you want an LLM to interact with external , aggingWeb scrapingCode understandingSynthetic data generationGraph queryingUse casesQ&A with F aggingWeb scrapingCode understandingSynthetic data generationGraph queryingUse casesQ&A with F aggingWeb scrapingCode understandingSynthetic data generationGraph queryingUse casesQ&A with F aggingWeb scrapingCode understandingSynthetic data generationGraph queryingUse casesQ&A with F aggingWeb scrapingCode understandingSynthetic data generationGraph queryingUse casesQ&A with F aggingWeb scrapingCode understandingSynthetic data generationGraph queryingUse casesQ&A with F aggingWeb scrapingCode understandingSynthetic data generationGraph queryingUse casesQ&A with F Chain. As an open-source project in a rapidly developing field, we are extremely open to contributions, v: r structured dataSQLSQLOpen In ColabUse caseEnterprise data is often stored in SQL databases.LLM tbotsOpen In ColabUse caseChatbots are one of the central LLM use-cases. The core features of chatb raphLangGraph LangGraph Building language agents as graphs ↗OverviewLangGraph is a aggingWeb scrapingCode understandingSynthetic data generationGraph queryingUse casesQ&A with F raphLangServe LangServe We will be releasing a hosted version of LangServe for one-clic nith WalkthroughLangGraphLangSmithLangSmithLangSmith helps you trace and evaluate your languag raphModulesChainsChainsChains refer to sequences of calls - whether to an LLM, a tool, or a data prep oolsChainsMoreLangServeLangSmithLangGraphModulesAgentsAgentsThe core idea of agents is to us modelsVector storesRetrieversIndexingAgentsChainsMoreLangServeLangSmithLangGraphModulesRe:rsRetrievalAgentsChainsMoreLangServeLangSmithLangGraphModulesModel I/OModel I/OThe core raphModulesModulesLangChain provides standard, extendable interfaces and external integrations for t outing by semantic similarityAdding memoryAdding moderationManaging prompt sizeUsing toolsModule raphLangChain Expression LanguageWhy use LCELWhy use LCELWe recommend reading the LCEL C nableLambda: Run Custom FunctionsRunnableBranch: Dynamically route logic based on inputBind runt raphLangChain Expression LanguageInterfaceInterfaceTo make it as easy as possible to create custom raphLangChain Expression LanguageGet startedGet startedLCEL makes it easy to build complex chains raphLangChain Expression LanguageLangChain Expression Language (LCEL)LangChain Expression L raphGet startedSecuritySecurityLangChain has a large ecosystem of integrations with various external r: raphGet startedQuickstartQuickstartIn this quickstart we'll show you how to:Get setup with LangChain, L raphGet startedInstallationInstallationOfficial releaseTo install LangChain run:PipCondapip install langch raphGet startedGet startedGet started with LangChain IntroductionLangChain is a framework for d Chain, & Pinecone by PineconeLangChain and Weaviate with Harrison Chase and Bob van Luijt - Weav Chain on WikipediaBooks Generative AI with LangChain by Ben Auffrath, © 2023 Packt PublishingI elastic-query-generatorextraction-anthropic-functionsextraction-openai-functionsguardrails-output-parse Chain. As an open-source project in a rapidly developing field, we are extremely open to contributions, v: eBlogCopyright © 2024 LangChain, Inc.

gchain-corelangchain-core is currently on version 0.1.x.As langchain-core contains the base abstraction ed to debug. A model call will fail, or the model output will be misformatted, or there will be some nested MemoryCallbacksChat loadersAdaptersStoresProvidersProviders AnthropicAll functionality related aggingWeb scrapingCode understandingSynthetic data generationGraph queryingUse casesQ&A with F raphGet startedIntroductionIntroductionLangChain is a framework for developing applications powered b raphGet startedIntroductionIntroductionLangChain is a framework for developing applications powered b

xtractApify DatasetArcGISArxivAssemblyAI Audio TranscriptsAstraDBAsync ChromiumAsyncHtmlAWS  
nprehend Moderation ChainAmazon Comprehend is a natural-language processing (NLP) service that u  
xtractApify DatasetArcGISArxivAssemblyAI Audio TranscriptsAstraDBAsync ChromiumAsyncHtmlAWS  
)EN AIElasticsearchEmbaasERNIEFake EmbeddingsFastEmbed by QdrantGoogle Generative AI Embe  
)EN AIElasticsearchEmbaasERNIEFake EmbeddingsFastEmbed by QdrantGoogle Generative AI Embe  
»x AIGPTRouterHugging Face Chat WrapperJinaChatKonkoLiteLLMllama 2 ChatLlama APIllamaEdgel  
»late2DatabricksDeepInfraDeepSparseEden AlFireworksForefrontAIGigaChatGoogle AIGoogle Cloud V  
»erveLangSmithLangGraphModulesModel I/OLLMsStreamingStreamingAll LLMs implement the Runnabl  
»erveLangSmithLangGraphModulesModel I/OLLMsQuick StartQuick StartLarge Language Models (LLM:  
»late2DatabricksDeepInfraDeepSparseEden AlFireworksForefrontAIGigaChatGoogle AIGoogle Cloud V  
»late2DatabricksDeepInfraDeepSparseEden AlFireworksForefrontAIGigaChatGoogle AIGoogle Cloud V  
»erveLangSmithLangGraphModulesModel I/OLLMsCustom LLMCustom LLMThis notebook goes over hc  
»erveLangSmithLangGraphModulesModel I/OLLMsCachingCachingLangChain provides an optional cach  
»erveLangSmithLangGraphModulesModel I/OLLMsTracking token usageTracking token usageThis notel  
»ithLangGraphModulesModel I/OChat ModelsTracking token usageTracking token usageThis notebook (»  
»ithLangGraphModulesModel I/OChat ModelsCachingCachingLangChain provides an optional caching la  
»ithLangGraphModulesModel I/OChat ModelsStreamingStreamingAll ChatModels implement the Runnat  
»ithLangGraphModulesModel I/OChat ModelsQuick StartQuick StartChat models are a variation on lang  
»nEnsemble RetrieverLong-Context ReorderMultiVector RetrieverParent Document RetrieverSelf-queryir  
»nEnsemble RetrieverLong-Context ReorderMultiVector RetrieverParent Document RetrieverSelf-queryir  
»nEnsemble RetrieverLong-Context ReorderMultiVector RetrieverParent Document RetrieverSelf-queryir  
»ithLangGraphModulesRetrievalText embedding modelsCacheBackedEmbeddingsCacheBackedEmbe  
» ChunkingSplit by tokensText embedding modelsVector storesRetrieversIndexingAgentsChainsMoreLa  
» LangServeLangSmithLangGraphModulesRetrievalDocument loadersPDFPDFPortable Document Form  
» LangServeLangSmithLangGraphModulesRetrievalDocument loadersJSONJSON (JavaScript Ob  
» LangServeLangSmithLangGraphModulesRetrievalDocument loadersFile DirectoryFile DirectoryThis co  
» LangServeLangSmithLangGraphModulesRetrievalDocument loadersMarkdownMarkdownMarkdown is  
» LangServeLangSmithLangGraphModulesRetrievalDocument loadersCSVCSVA comma-separated val  
» LangServeLangSmithLangGraphModulesRetrievalDocument loadersHTMLHTMLThe HyperText Marku  
» graph QA chainNebulaGraphQAChainNetworkXGraphSparqlQAChainNeptune Open Cypher QA ChainL  
» graph QA chainNebulaGraphQAChainNetworkXGraphSparqlQAChainNeptune Open Cypher QA ChainL

netadata taggerText embedding modelsVector storesRetrieversToolsAgents and toolkitsMemoryCallbac  
»x AIGPTRouterHugging Face Chat WrapperJinaChatKonkoLiteLLMllama 2 ChatLlama APIllamaEdgel

processing step. The primary supported way to do this is with LCEL. LCEL is great for constructing your processing step. The primary supported way to do this is with LCEL. LCEL is great for constructing your c  
x AIGPTRouterHugging Face Chat WrapperJinaChatKonkoLiteLLMllama 2 ChatLlama APILlamaEdge processing step. The primary supported way to do this is with LCEL. LCEL is great for constructing your c  
late2DatabricksDeepInfraDeepSparseEden AIFireworksForefrontAIGigaChatGoogle AI Google Cloud V  
late2DatabricksDeepInfraDeepSparseEden AIFireworksForefrontAIGigaChatGoogle AI Google Cloud V  
late2DatabricksDeepInfraDeepSparseEden AIFireworksForefrontAIGigaChatGoogle AI Google Cloud V  
gent backed by a databaseCustomizing Conversational MemoryCustom MemoryMultiple Memory classe  
nEnsemble RetrieverLong-Context ReorderMultiVector RetrieverParent Document RetrieverSelf-queryir  
c ChunkingSplit by tokensText embedding modelsVector storesRetrieversIndexingAgentsChainsMoreLa  
nEnsemble RetrieverLong-Context ReorderMultiVector RetrieverParent Document RetrieverSelf-queryir  
Cloud ElasticSearch VectorSearchBigQuery Vector SearchChromaClarifaiClickHouseDashVectorData  
nEnsemble RetrieverLong-Context ReorderMultiVector RetrieverParent Document RetrieverSelf-queryir  
EN AIElasticsearchEmbaasERNIEFake EmbeddingsFastEmbed by QdrantGoogle Generative AI Embe  
xtractApify DatasetArcGISArxivAssemblyAI Audio TranscriptsAstraDBAsync ChromiumAsyncHtmlAWS  
xtractApify DatasetArcGISArxivAssemblyAI Audio TranscriptsAstraDBAsync ChromiumAsyncHtmlAWS  
c ChunkingSplit by tokensText embedding modelsVector storesRetrieversIndexingAgentsChainsMoreLa  
c ChunkingSplit by tokensText embedding modelsVector storesRetrieversIndexingAgentsChainsMoreLa  
xtractApify DatasetArcGISArxivAssemblyAI Audio TranscriptsAstraDBAsync ChromiumAsyncHtmlAWS  
xtractApify DatasetArcGISArxivAssemblyAI Audio TranscriptsAstraDBAsync ChromiumAsyncHtmlAWS  
risonGithubGitlabGmailJiraJSONMultiOnNASAOffice365OpenAPINatural Language APIsPandas Dataf  
risonGithubGitlabGmailJiraJSONMultiOnNASAOffice365OpenAPINatural Language APIsPandas Dataf  
risonGithubGitlabGmailJiraJSONMultiOnNASAOffice365OpenAPINatural Language APIsPandas Dataf  
xtractApify DatasetArcGISArxivAssemblyAI Audio TranscriptsAstraDBAsync ChromiumAsyncHtmlAWS  
xtractApify DatasetArcGISArxivAssemblyAI Audio TranscriptsAstraDBAsync ChromiumAsyncHtmlAWS  
sToken countingLangServeLangSmithLangGraphModulesMoreCallbacksToken countingToken counting  
aphModulesAgentsToolsTools as OpenAI FunctionsTools as OpenAI FunctionsThis notebook goes ove  
angChain makes it easy to prototype LLM applications and Agents. However, delivering LLM applicati  
olsChainsMoreLangServeLangSmithLangGraphModulesAgentsAgent TypesOpenAI functionsOpenAI fu  
.access intermediate stepsCap the max number of iterationsTimeouts for agentsToolsChainsMoreLangSe  
.access intermediate stepsCap the max number of iterationsTimeouts for agentsToolsChainsMoreLangSe  
.access intermediate stepsCap the max number of iterationsTimeouts for agentsToolsChainsMoreLangSe  
.access intermediate stepsCap the max number of iterationsTimeouts for agentsToolsChainsMoreLangSe  
.access intermediate stepsCap the max number of iterationsTimeouts for agentsToolsChainsMoreLangSe  
.access intermediate stepsCap the max number of iterationsTimeouts for agentsToolsChainsMoreLangSe  
.access intermediate stepsCap the max number of iterationsTimeouts for agentsToolsChainsMoreLangSe  
.access intermediate stepsCap the max number of iterationsTimeouts for agentsToolsChainsMoreLangSe  
aphModulesAgentsToolsTools are interfaces that an agent can use to interact with the world. They  
olsChainsMoreLangServeLangSmithLangGraphModulesAgentsAgent TypesAgent TypesThis categorize  
id the agent framework, let's build an agent that has two tools: one to look things up online, and one to lo  
jents is to use a language model to choose a sequence of actions to take. In chains, a sequence of acti  
ng parserPandas DataFrame ParserPydantic parserRetry parserStructured output parserXML parserYA  
nEnsemble RetrieverLong-Context ReorderMultiVector RetrieverParent Document RetrieverSelf-queryir  
eskChatGPT PluginCohere RerankerCohere RAG retrieverDocArrayElasticSearch BM25EmbedchainFle  
nEnsemble RetrieverLong-Context ReorderMultiVector RetrieverParent Document RetrieverSelf-queryir  
trievalIndexingIndexingHere, we will look at a basic indexing workflow using the LangChain indexing AP  
tput ParsersOutput ParsersOutput parsers are responsible for taking the output of an LLM and transform  
rtThe quick start will cover the basics of working with language models. It will introduce the two different  
essagePromptTemplate`Partial prompt templatesPipelineQuickstartLLMsChat ModelsOutput ParsersRe  
The core element of any language model application is...the model. LangChain gives you the building bl  
gent backed by a databaseCustomizing Conversational MemoryCustom MemoryMultiple Memory classe  
esModel I/ORetrievalAgentsChainsMoreLangServeLangSmithLangGraphLangChain Expression Langua  
esModel I/ORetrievalAgentsChainsMoreLangServeLangSmithLangGraphLangChain Expression Langua







ChunkingSplit by tokensText embedding modelsVector storesRetrieversIndexingAgentsChainsMoreLangServeLangSmithLangGraphModulesRetrievalDocument loadersDocument loadersINFOHead to Ira, rather than data collected from real-world events. It's used to simulate real data without compromising graph QA chainNebulaGraphQACChainNetworkXGraphSparqlQACChainNeptune Open Cypher QA ChainLM applications (e.g., GitHub Co-Pilot, Code Interpreter, Codium, and Codeium) for use-cases such as: Qhighlighted it as one of his top desired AI tools.OSS repos like gpt-researcher are growing in popularity. One style (formal, informal etc.) covered topics political tendencyOverviewTagging has a few components: full customer questions, etc.) and you want to summarize the content.LLMs are a great tool for this given the fact, suppose you need the model output formatted with a specific schema for:Extracting a structured row to APIs.This can be very useful for retrieving context for the LLM to utilize.And, more generally, it allows us to RAGUsing local modelsUsing local modelsThe popularity of projects like PrivateGPT, llama.cpp, and GPT4RAGUsing agentsUsing agentsThis is an agent specifically optimized for doing retrieval when necessary.RAGStreamingStreamingOften in Q&A applications it's important to show users the sources that were used.RAGPer-User RetrievalPer-User RetrievalWhen building a retrieval app, you often have to build it with many RAGAdd chat historyAdd chat historyIn many Q&A applications we want to allow the user to have a backlog.RAGQuickstartQuickstartLangChain has a number of components designed to help build question-answer systems.RAGReturning sourcesReturning sourcesOften in Q&A applications it's important to show users the sources whether they involve new features, improved infrastructure, better documentation, or bug fixes. Guis make it possible to interact with SQL databases using natural language.LangChain offers SQL Chains. The benefits are that they can have long-running conversations and have access to information that users want to query a library for building stateful, multi-actor applications with LLMs, built on top of (and intended to be used with) RAGQ&A with RAGOvewOne of the most powerful applications enabled by LLMs is sophisticated quick deployments of LangChain applications. Sign up here to get on the waitlist.OverviewLangServe helps you to build model applications and intelligent agents to help you move from prototype to production.Check out the processing step. The primary supported way to do this is with LCEL.LCEL is great for constructing your own language model to choose a sequence of actions to take. In chains, a sequence of actions is hard-coded into retrievalRetrievalMany LLM applications require user-specific data that is not part of the model's training set. This element of any language model application is...the model. LangChain gives you the building blocks to implement the following main modules:Model I/OInterface with language modelsRetrievalInterface with application-specificsModel I/ORetrievalAgentsChainsMoreLangServeLangSmithLangGraphLangChain Expression LanguageGet started section first.LCEL makes it easy to build complex chains from basic components. It does this by argsConfigure chain internals at runtimeCreate a runnable with the `@chain` decoratorAdd fallback chains, we've implemented a "Runnable" protocol. The Runnable protocol is implemented for most components from basic components, and supports out of the box functionality such as streaming, parallelism, and language, or LCEL, is a declarative way to easily compose chains together. LCEL was designed from the ground up to work with resources like local and remote file systems, APIs and databases. These integrations allow developers to use LangSmith and LangServeUse the most basic and common components of LangChain: prompt template chainThis will install the bare minimum requirements of LangChain. A lot of the value of LangChain comes from developing applications powered by language models. It enables applications that: InstallationOfficial Weaviate Podcast #36 by Weaviate • Vector DatabaseLangChain Demo + Q&A with Harrison Chase by Full Stack DeepLearning.AI coursesby Harrison Chase and Andrew NgLangChain for LLM Application DevelopmentHybrid Search in Weaviatehydellama2-functionsmongo-parent-document-retrievalneo4j-advanced-ragWhether they involve new features, improved infrastructure, better documentation, or bug fixes. Guis

is and runtime for the whole LangChain ecosystem, we will communicate any breaking changes with advanced model calls and it won't be clear where along the way an incorrect output was created. Here are a few changes to Anthropic models. AWSThe LangChain integrations related to Amazon AWS platform. GPT4RAGQ&A with RAGOvewOne of the most powerful applications enabled by LLMs is sophisticated query language models. It enables applications that:Are context-aware: connect a language model to source query language models. It enables applications that:Are context-aware: connect a language model to source

cksChat loadersAdaptersStoresComponentsDocument transformersOpenAI metadata taggerOpenAI me  
MiniMaxMistralAINVIDIA AI Foundation EndpointsOllamaOllama FunctionsOpenAIPromptLayer ChatOr







neo4j-cypher-ftneo4j-cypher-memoryneo4j\_cypherneo4j-generationneo4j-parentneo4j-semantic-layerne  
equests cannot land without passing the formatting, linting, and testing checks first. See Testing and For  
Source Community.Partner Packages: For independent packages that are co-maintained by LangChain ar  
do so in code review.By default, most pull requests will have a  Maintainers are allowed to edit this pu  
al services.Unit TestsUnit tests cover modular logic that does not require calls to outside APIs. If you add  
documentation to all classes and methods.Similar to linting, we recognize documentation can be annoying.  
unableUse LCEL under the hoodNatBotChainNot usedcreate\_openai\_fn\_chaincreate\_openai\_fn\_runna  
tions for documentation on built-in callbacks integrations with 3rd-party tools.LangChain provides a calli  
his example shows how to print logs to file. It shows how to use the FileCallbackHandler, which does the  
ct, predict\_messages. Will be removed in 0.2.0. Use BaseLLM.invoke` instead.BasellLM methods apred  
College ConfidentialCometConfident AIConfluenceContextC TransformersDashVectorDatabricksDatadog  
esent some built-in chains inspired by this research, which are intended to make the outputs of LLMs sa  
vay.LangChain provides the concept of a ModelLaboratory to test out and try different models.%pip insta  
be installed at the same timeLangChain Pydantic migration planAs of langchain>=0.0.267, LangChain v  
of a commercial serviceCost: There is no inference fee, which is important for token-intensive applicatio  
we've introduced the concept of fallbacks.A fallback is an alternative plan that may be used in an emerge  
College ConfidentialCometConfident AIConfluenceContextC TransformersDashVectorDatabricksDatadog  
inputs, and that they work well with your application's other software components. Ensuring reliability us  
into two categories:Case 1: Utilizing External LLM Providers (OpenAI, Anthropic, etc.) In this scenario, r  
ul. In this section, we will look at a few different ways to store key-value pairs using implementations of th  
to OpenAI AdapterPlease ensure OpenAI library is version 1.0.0 or higher; otherwise, PreviousWh  
it loadersChat loaders DiscordThis notebook shows how to create your own chat loader that works  
open-source data curation ConfidentDeepEval package for unit testing LLMs. ContextConte  
Astra DBDataStax [Astra] ZepCallbacksChat loadersAdaptersStoresComponentsMemoryMemory Astra DBDataStax [Astra]  
echFile SystemGolden QueryGoogle Cloud Text-to-SpeechGoogle DriveGoogle FinanceGoogle JobsGr  
framePlayWright BrowserPowerBI DatasetPythonRobocorpSlackSpark DataframeSpark SQLSQL Datab  
eet AI Libraries ContextGoogle DriveGoogle Vertex AI SearchJaguar Vector DatabaseKay.aikNNLOTR (bricks Vector SearchDingoDBDocArray HnswSearchDocArray InMemorySearchElasticsearchEpsillaFais  
embeddingsGoogle Vertex AI PaLMGPT4AllGradientHugging FaceInfinityInstruct Embeddings on Hugging F  
S3 DirectoryAWS S3 FileAZLyricsAzure AI DataAzure Blob Storage ContainerAzure Blob Storage FileA  
cksChat loadersAdaptersStoresComponentsDocument transformersDocument transformers Beaut  
MiniMaxMistralAINVIDIA AI Foundation EndpointsOllamaOllama FunctionsOpenAIPromptLayer ChatOr  
ertext AIGooseAIGPT4AllGradientHugging Face HubHugging Face Local PipelinesHuggingface TextGe  
as Azure is a cloud computing platform run by Microsoft, which offers access, management, and develop  
embedding models49 items Vector stores71 items Retrievers39 items Tools56 items  
incorporated and its for-profit subsidiary corporation OpenAI Limited Partnership. OpenAI conducts AI rese  
er 350k models, 75k datasets, and 150k demo apps (Spaces), all open source and publicly available, in a  
ch as gemini-pro and gemini-pro-vision through the ChatGoogleGenerativeAI class.pip install -U langcha  
of high-performing foundation models (FMs) from leading AI companies like AI21 Labs, Anthropic, Cohes  
covers all integrations between Anthropic models and LangChain.Prompting OverviewClaude is chat-b  
ponent of LangChain. LangChain does not serve its own LLMs, but rather provides a standard interface  
ain. LangChain does not serve its own ChatModels, but rather provides a standard interface for interacti  
ngTime-weighted vector store retrieverIndexingAgentsChainsMoreLangServeLangSmithLangGraphMod  
th 3rd-party vector stores.One of the most common ways to store and search over unstructured data is t  
ons for documentation on built-in integrations with text embedding model providers.The Embeddings cla

LangServeLangSmithLangGraphModulesRetrievalText SplittersText SplittersOnce you've loaded documentations for documentation on built-in document loader integrations with 3rd-party tools. Use documentation privacy or encountering real-world limitations. Benefits of Synthetic Data: Privacy and Security: No real-life casesGraph queryingGraph databases give us a powerful way to represent and query over the code base to understand how it worksUsing LLMs for suggesting refactors or improvement overviewGathering content from the web has a few components: Search: Query to url (e.g., using Google's function: Like extraction, tagging uses functions to specify how the model should tag a document's schema: higher proficiency in understanding and synthesizing text. In this walkthrough we'll go over how to perform document insert into a databaseExtracting API parametersExtracting different parts of a user query (e.g., for servers to interact with APIs using natural language! OverviewThere are two primary ways to interface LLMs with LangChain: Local and Cloud. Both underscore the importance of running LLMs locally. LangChain has integrations with many open-source frameworks and also holding a conversation. To start, we will set up the retriever we want to use, and then turn it into a chain that can be used to generate the answer. The simplest way to do this is for the chain to return the documents that we have in mind. This means that you may be storing data not just for one user, but for many different users. This is useful for k-and-forth conversations, meaning the application needs some sort of "memory" of past questions and answers. There are many ways to contribute to LangChain. Here are some common ways to contribute:

- Agents to build and run SQL queries based on natural language prompts. These are compatible with LangChain.
- APIs to interact with LangChain. It extends the LangChain Expression Language with the ability to coordinate multiple chains.
- Chatbots that can answer questions about specific domains. Developers can deploy LangChain runnables and chains as a REST API. This library is integrated with FastAPI.
- Interactive walkthroughs to get started. For more information, please refer to the LangSmith documentation.
- Off-the-shelf chains, but it's also nice to have chains that you can use off-the-shelf. There are two types of off-the-shelf chains: agents and models.
- Reasoning engines that can be used as a reasoning engine to determine which actions to take. The primary way of accomplishing this is through Retrieval Augmented Generation (RAG). In this process, the LangChain Expression Language is used to interface with any language model.
- Conceptual GuideA conceptual explanation of messages, prompts, and specific data.
- AgentsLet chains choose which tools to use given high-level directivesAdditionalChainsCorporationCookbookCookbookExample code for accomplishing common tasks with the LangChain Expression Language by providing: 1. A unified interface: Every LCEL object implements the Runnable interface, which defines Stream custom generator functionsInspect your runnablesAdd message history (memory)CookbookComponents. This is a standard interface, which makes it easy to define custom chains as well as invoke them.
- Logging. Basic example: prompt + model + output parserThe most basic and common use case is chainiray 1 to support putting prototypes in production, with no code changes, from the simplest "prompt + LLM" to create versatile applications that combine the power of LLMs with the ability to access, interact with APIs, models, and output parsersUse LangChain Expression Language, the protocol that LangChain is built upon when integrating it with various model providers, datastores, etc. By default, the dependencies needed for LangChain are released.
- QuickstartIn this quickstart we'll show you how to: SecurityLangChain has a large Stack Deep LearningLangChain Agents: Build Personal Assistants For Your Data (Q&A with Harrison ClintonLangChain Chat with Your DataLangChain Functions, Tools and Agents with LangChainHandbookLangChain, neo4j-cypher-ftneo4j-cypher-memoryneo4j\_cypherneo4j-generationneo4j-parentneo4j-semantic-layerne
- Ways to contributeThere are many ways to contribute to LangChain. Here are some common ways to contribute:

vance notice and version bumps. The exception for this is anything marked with the beta decorator (you can use different tools and functionalities to aid in debugging). TracingPlatforms with tracing capabilities like LangChain, GoogleAll functionality related to Google Cloud Platform and other Google products. Hugging FaceReasoning engines (Q&A) chatbots. These are applications that can answer questions about specific sources of context (prompt instructions, few shot examples, content to ground its response in, etc.) Reasoning engines (Q&A) chatbots. These are applications that can answer questions about specific sources of context (prompt instructions, few shot examples, content to ground its response in, etc.)

Azure AI Document Intelligence BibTeX BiliBili Blackboard Blockchain Brave Search Browserless ChatGPT [use Amazon Comprehend to detect and handle Personally Identifiable Information (PII) and toxicity. Set Azure AI Document Intelligence BibTeX BiliBili Blackboard Blockchain Brave Search Browserless ChatGPT [face Jina John Snow Labs Llama-cpp LLM Rails Local AI MiniMax Model Scope Mosaic ML NLP Cloud NVIDIA AI face Jina John Snow Labs Llama-cpp LLM Rails Local AI MiniMax Model Scope Mosaic ML NLP Cloud NVIDIA AI openAI Tencent Hunyuan Tongyi QwenvLLM ChatVolc Enging Maas Yandex GPTZHIPU AI Document loade n Inference Javelin AI Gateway Tutorial JSONFormer KoboldAI API Llama.cpp LLM Caching integrationsLI n, astream. This gives all LLMs basic support for streaming. Streaming support defaults to returning an It standard interface for interacting with many different LLMs. There are lots of LLM providers (OpenAI, Col n Inference Javelin AI Gateway Tutorial JSONFormer KoboldAI API Llama.cpp LLM Caching integrationsLI n Inference Javelin AI Gateway Tutorial JSONFormer KoboldAI API Llama.cpp LLM Caching integrationsLI one that is supported in LangChain. There are only two required things that a custom LLM needs to impl 'I calls you make to the LLM provider, if you're often requesting the same completion multiple times. It ca OpenAI API. Let's first look at an extremely simple example of tracking token usage for a single LLM call. AI API. Let's first look at an extremely simple example of tracking token usage for a single Chat model ca PI calls you make to the LLM provider, if you're often requesting the same completion multiple times. It c im, astream. This gives all ChatModels basic support for streaming. Streaming support defaults to returni fferent. Rather than using a "text in, text out" API, they use an interface where "chat messages" are the i Jules Retrieval Retrievers Time-weighted vector store retriever Time-weighted vector store retriever This ref Jules Retrieval Retrievers Contextual compression Contextual compression One challenge with retrieval is t Jules Retrieval Retrievers Long-Context Reorder Long-Context Reorder No matter the architecture of your r ocompute them. Caching embeddings can be done using a CacheBackedEmbeddings. The cache backed the simplest method. This splits based on characters (by default "") and measure chunk length by numb models have a token limit. You should not exceed the token limit. When you split your text into chunks it i: olits the text based on semantic similarity. Taken from Greg Kamradt's wonderful notebook: <https://github> vs you to split your code with multiple languages supported. Import enum Language and specify the lang xtSplitterDescription and motivationSimilar in concept to the `MarkdownHeaderTextSplitter`, the `HTMLH ts, including text formatting and images, in a manner independent of application software, hardware, and xt to store and transmit data objects consisting of attribute-value pairs and arrays (or other serializable v der. from langchain\_community.document\_loaders import DirectoryLoader We can use the glob parameter load Markdown documents into a document format that we can use downstream. # !pip install unstructure a record. Each record consists of one or more fields, separated by commas. Load CSV data with a single e b browser. This covers how to load HTML documents into a document format that we can use downstream queries Neptune graph database using openCypher and returns human readable response from lang t choice for applications based on network-like models. To standardize the syntax and semantics of such Ms to provide a natural language interface to a Memgraph database. To complete this tutorial, you will n : LLMs to provide a natural language interface to NebulaGraph database. You will need to have a running on, and study of the structure, dynamics, and functions of complex networks. This notebook goes over hi : to provide a natural language interface to a graph database you can query with the Cypher query langu o provide a natural language interface to FalkorDB database. FalkorDB is a low latency property graph d : a natural language interface to Kùzu database. Kùzu is an in-process property graph database manage LLMs to provide a natural language interface to HugeGraph database. You will need to have a running H now to use LLMs to provide a natural language interface to an ArangoDB database. You can get a local / Text data often contain rich relationships and insights that can be useful for various analytics, recommen

etadata tagger It can often be useful to tag ingested documents with structured metadata, such as the title openAI Tencent Hunyuan Tongyi QwenvLLM ChatVolc Enging Maas Yandex GPTZHIPU AI Document loade









ents, you'll often want to transform them to better suit your application. The simplest example is you may want loaders to load data from a source as Document's. A Document is a piece of text and associated meta personal data at risk of breaches.

**Data Augmentation:** Expands datasets for machine learning.

**Flexibility:** Many real-world relationships. There are a number of chains that make it easy to use LLMs to interact with them.

**Using LLMs for documenting the code**

**Overview:** The pipeline for QA over code follows the steps we do:

- SearchAPIWrapper**: Loading: Url to HTML (e.g., using AsyncHtmlLoader, AsyncChromiumLoader, etc).
- defines how we want to tag the document

**Quickstart:** Let's see a very straightforward example of how we can build a document summarization using LLMs.

**Overview:** A central question for building a summarizer is how to pass context (semantic vs keyword search)

**Functions:** There are two primary approaches for this:

- Functions:** Some LLMs can call external APIs.
- LLM-generated source LLMs:** These are LLMs that can be run locally. See here for setup instructions for these LLMs. For example, here we will use a retriever tool. Next, we will use the high level constructor for this type of agent. Finally, we will walk through how documents are retrieved in each generation. We'll work off of the Q&A app with sources we built over the LLM Power user users, and they should not be able to see each other's data. This means that you need to be able to cache answers, and some logic for incorporating those into its current thinking.

In this guide we focus on adding a simple Q&A application over a text data source. Along the way we'll go over a typical Q&A architecture, documents that were retrieved in each generation. We'll work off of the Q&A app we built over the LLM Power user users, and they should not be able to see each other's data. This means that you need to be able to cache answers, and some logic for incorporating those into its current thinking.

**Common ways people contribute:**

- Documentation:** Help improve our docs, including this one!
- Code:** Help us with any SQL dialect supported by SQLAlchemy (e.g., MySQL, PostgreSQL, Oracle SQL, Databricks, SQL) or a chatbot. Memory allows a chatbot to remember past interactions, and retrieval provides a chatbot with chains (or actors) across multiple steps of computation in a cyclic manner. It is inspired by Pregel and ARGE.
- Information:** These applications use a technique known as Retrieval-Augmented Generation, or RAG.
- tAPI:** tAPI and uses pydantic for data validation.
- In addition:** In addition, it provides a client that can be used to call into running LLMs.
- For tutorials and other end-to-end examples:** demonstrating ways to integrate LangSmith in your workflow.
- e-shelf chains that LangChain supports:** Chains that are built with LCEL. In this case, LangChain offers a choice of which order to take and in which order.
- Quickstart:** For a quick start to working with agents, please check out this getting started guide, where external data is retrieved and then passed to the LLM when doing the generation step.
- LangChain:** Covers the common building block compositions.
- Memory:** Persist application state between runs of a chain.
- Persist:** Callbacks for Language (LCEL). These examples show how to compose different Runnable (the core LCEL interface) components into a single Runnable.
- LCEL:** These examples show how to compose different Runnable (the core LCEL interface) components into a single Runnable.
- Language (LCEL):** These examples show how to compose different Runnable (the core LCEL interface) components into a single Runnable.
- Interfaces:** These examples show how to compose different Runnable (the core LCEL interface) components into a single Runnable.
- Streams:** stream: stream back chunks of the response in chunks, using a prompt template and a model together.
- Invocation:** To see how this works, let's create a chain that takes a topic and manipulates external resources.
- Best Practices:** When building such applications developers should remember which components are installed and which facilitate component chaining.
- Build:** Build a simple application with LangChainTrace your application to do that are NOT installed. You will need to install the dependencies for specific integrations separately.
- Ecosystem:** The ecosystem of integrations with various external resources like local and remote file systems, APIs and databases (such as PostgreSQL and MySQL).
- Chat:** Chat with data (sorted by views)
- Videos:** Using ChatGPT with YOUR OWN Data
- AI Handbook:** By James Briggs and Francisco Ingham
- Short Tutorials:** LangChain Explained in 13 Minutes
- OpenAI Functions Agent - Gmail:** openai-functions-agent
- openai-fi**
- mon ways people contribute:**
- Documentation:** Help improve our docs, including this one!
- Code:** Help us with any SQL dialect supported by SQLAlchemy (e.g., MySQL, PostgreSQL, Oracle SQL, Databricks, SQL) or a chatbot.

I can see this in the API reference and will see warnings when using such functionality). The reason for this is that Smith and WandB are the most comprehensive solutions for debugging. These platforms make it easy to track all functionality related to the Hugging Face Platform.

 Microsoft All functionality related to Microsoft

re information. These applications use a technique known as Retrieval-Augmented Generation, or RAG.

ely on a language model to reason (about how to answer based on provided context, what actions to take).

ely on a language model to reason (about how to answer based on provided context, what actions to take).

DataCollege Confidential Concurrent LoaderConfluenceCoNLL-UCopy PasteCouchbaseCSVCube Semantic  
tting up%pip install --upgrade --quiet boto3 nltk%pip install --upgrade --quiet langchain\_experimental%  
DataCollege Confidential Concurrent LoaderConfluenceCoNLL-UCopy PasteCouchbaseCSVCube Semantic  
AI Foundation EndpointsOllamaOpenClipOpenAISageMakerSelf HostedSentence Transformers on Hug  
AI Foundation EndpointsOllamaOpenClipOpenAISageMakerSelf HostedSentence Transformers on Hug  
ersDocument transformersText embedding modelsVector storesRetrieversToolsAgents and toolkitsMem  
M Format EnforcerManifestMinimaxModalMosaicMLNLP CloudOCI Data Science Model Deployment Er  
terator (or AsyncIterator in the case of async streaming) of a single value, the final result returned by the  
here, Hugging Face, etc) - the LLM class is designed to provide a standard interface for all of them.In thi  
M Format EnforcerManifestMinimaxModalMosaicMLNLP CloudOCI Data Science Model Deployment Er  
M Format EnforcerManifestMinimaxModalMosaicMLNLP CloudOCI Data Science Model Deployment Er  
ement:A \_call method that takes in a string, some optional stop words, and returns a string.A \_llm\_type  
an speed up your application by reducing the number of API calls you make to the LLM provider.from lar  
.from langchain.callbacks import get\_openai\_callbackfrom langchain\_openai import OpenAIllm = OpenAI  
all.from langchain.callbacks import get\_openai\_callbackfrom langchain\_openai import ChatOpenAIllm =  
can speed up your application by reducing the number of API calls you make to the LLM provider.from Is  
ing an Iterator (or AsyncIterator in the case of async streaming) of a single value, the final result returned  
inputs and outputs.SetupFor this example we'll need to install the OpenAI partner package:pip install lan  
triever uses a combination of semantic similarity and a time decay.The algorithm for scoring them is:sen  
that usually you don't know the specific queries your document storage system will face when you inges  
model, there is a substantial performance degradation when you include 10+ retrieved documents. In bri  
embedder is a wrapper around an embedder that caches embeddings in a key-value store. The text is t  
per of characters.How the text is split: by single character.How the chunk size is measured: by number o  
s therefore a good idea to count the number of tokens. There are many tokenizers. When you count tok  
.com/FullStackRetrieval-com/RetrievalTutorials/blob/main/5\_Levels\_Of\_Text\_Splitting.ipynbAll credit to  
guage.from langchain.text\_splitter import ( Language, RecursiveCharacterTextSplitter,# Full list of s  
leaderTextSplitter` is a “structure-aware” chunker that splits text at the element level and adds metadata  
d operating systems.This covers how to load PDF documents into the Document format that we use dov  
values).JSON Lines is a file format where each line is a valid JSON value.The JSONLoader uses a spec  
er to control which files to load. Note that here it doesn't load the .rst file or the .html files.loader = Direct  
id > /dev/nullfrom langchain\_community.document\_loaders import UnstructuredMarkdownLoadermarkdc  
> row per document.from langchain\_community.document\_loaders.csv\_loader import CSVLoaderloader  
sam.from langchain\_community.document\_loaders import UnstructuredHTMLLoaderloader = Unstructur  
jchain\_community.graphs import NeptuneGraphhost = "<neptune-host>"port = 8182use\_https = Truegra  
n graphs, the W3C recommends Semantic Web Technologies, cp. Semantic Web. SPARQL serves as a  
eed Docker and Python 3.x installed.To follow along with this tutorial, ensure you have a running Memg  
g NebulaGraph cluster, for which you can run a containerized cluster by running the following script:c  
ow to do question answering over a graph data structure.Setting upWe have to install a Python package  
uage.You will need to have a running Neo4j instance. One option is to create a free Neo4j database inst  
latabase management system. You can simply run its docker locally:docker run -p 6379:6379 -it --rm fall  
ement system. You can simply install it with pip:pip install kuzuOnce installed, you can simply import it ar  
hugeGraph instance. You can run a local docker container by running the executing the following script:c  
ArangoDB instance running via the ArangoDB Docker image:docker run -p 8529:8529 -e ARANGO\_RO  
idation engines, or knowledge management applications.Diffbot's NLP API allows for the extraction of er

e, tone, or length of a document, to allow for a more targeted similarity search later. However, for large r  
ersDocument transformersText embedding modelsVector storesRetrieversToolsAgents and toolkitsMem







GMailThis loader goes over how to load data from GMail. There are many ways iMessage  
Label Studio is an LLMonitorLLMonitor is an open-source observability platform that provides cost searchElasticsearch is a Momento CacheMomento Cache is the world's first MongoDBMongo  
loggingFace Hub ToolsHuman as a toolIFTTT WebHooksLemon AgentMemorizeMetaphor SearchNuclie  
esComponentsAgents and toolkitsAgents and toolkits AI NetworkAI Network is a layer 1 Airbr  
ailleRePhraseQuerySEC filingSelf-querying retrieverSingleStoreDBSVMTavily Search APITF-IDFVespa  
LLMRailsMarqoMeilisearchMilvusMomento Vector Index (MVI)MongoDB AtlasMyScaleNeo4j Vector In  
AI Foundation EndpointsOllamaOpenClipOpenAISageMakerSelf HostedSentence Transformers on Hug  
DataCollege ConfidentialConcurrent LoaderConfluenceCoNLL-UCopy PasteCouchbaseCSVCube Sema  
g platform from Google Cloud to Doctran: extract propertiesWe can extract useful features of docu  
ersDocument transformersText embedding modelsVector storesRetrieversToolsAgents and toolkitsMem  
M Format EnforcerManifestMinimaxModalMosaicMLNLP CloudOCI Data Science Model Deployment Er  
ding software as a service (SaaS), platform as a service (PaaS), and infrastructure as a service (IaaS). M  
1 items Adapters2 items Stores6 itemsPreviousZillizNextLLMsCommunityDiscordTwitterGitl  
Azure-based supercomputing platform from Microsoft. The OpenAI API is powered by a diverse set of n  
ral place where anyone can explore, experiment, collaborate, and build technology with Machine Learnin  
\_genai import ChatGoogleGenerativeAlllm = ChatGoogleGenerativeAI(model="gemini-pro")llm.invoke('  
uild generative AI applications with security, privacy, and responsible AI. Using Amazon Bedrock, you ca  
s in single string. It expects this string to be in a particular format. This means that it is up the user to er  
g and returns a string. There are lots of LLM providers (OpenAI, Cohere, Hugging Face, etc) - the LLM cl  
s and returns a message. There are lots of model providers (OpenAI, Cohere, Hugging Face, etc) - the C  
ed query. It is more general than a vector store. A retriever does not need to be able to store documents  
d query and retrieve the embedding vectors that are 'most similar' to the embedded query. A vector store  
providers (OpenAI, Cohere, Hugging Face, etc) - this class is designed to provide a standard interface fo

want to split a long document into smaller chunks that can fit into your model's context window. LangChain has several document loaders for loading simple .txt files, for loading the text content of various graph DBs, or for document question answering, with some differences: In particular, we can employ a splitting strategy for transforming HTML to formatted text (e.g., using HTML2Text or BeautifulSoup). Quickstart: pip install -q > can use OpenAI functions for tagging in LangChain. %pip install --upgrade --quiet langchain langchain-qa-chain is your documents into the LLM's context window. Two common approaches for this are:

- Stuff:** Simply "stuff" your documents into the LLM's context window.
- Span:** Call functions to extract arbitrary entities from LLM responses.

**Parsing:** Output parsers are classes that implement an iterated interface: Use an LLM with access to API documentation to create an interface. Quickstart: Many examples show how to run GPT4All or LLaMA2 locally (e.g., on your laptop) using local embeddings and a local API. Through how to construct a conversational retrieval agent from components. %pip install --upgrade --quiet langchain langchain-retrieval-autonomous-agents

We'll show how to run GPT4All or LLaMA2 locally (e.g., on your laptop) using local embeddings and a local API. Through how to construct a conversational retrieval agent from components. %pip install --upgrade --quiet langchain langchain-retrieval-autonomous-agents

SetupDependencies: We'll use an LLM to generate queries that will be run based on natural language questions. We'll also show how to retrieve up-to-date, domain-specific information. Overview: The chat model interface is based around messages and a message history. The current interface exposed is one inspired by NetworkX. The main use is for adding cycles. What is RAG? RAG is a technique for augmenting LLM knowledge with additional data. LLMs can reason about data deployed on a server. A javascript client is available in LangChainJS. Features: Input and Output, Workflow, check out the LangSmith Cookbook. Some of the guides therein include:

- Leveraging user feedback:** Higher-level constructor method. However, all that is being done under the hood is constructing a chain of LCEL objects.
- Getting started guide:** This covers basics like initializing an agent, creating tools, and adding memory.
- Conceptual guide:** Provides all the building blocks for RAG applications - from simple to complex.

This section of the documentation provides the basics of getting started working with different types of models. You should walk through [this section] and stream intermediate steps of any chain. PreviousLangChain Expression Language (LCEL) NextModel components to achieve various tasks. If you're just getting acquainted with LCEL, the Prompt + LLM paradigm allows chains of LCEL objects to also automatically support these invocations. That is, every chain of LCEL objects can be manipulated in parallel. How to:

- RunnableParallel:** Manipulating datamanipulating-inputs-output
- RunnableJoke:** call the chain on an inputbatch: call the chain on a list of inputs
- RunnableAsynchronous:** These also have corresponding asynchronous methods and generates a joke: %pip install --upgrade --quiet langchain-core langchain-community langchain-openai

To highlight a few of the reasons you might want to use LCEL:

- Streaming support:** When you're building a system that needs to follow good security practices.
- Limit Permissions:** Scope permissions specifically to the application with LangSmithServe.
- LangServe:** Your application with LangServeThat's a fair amount to cover! Let's dive in.

From sourceIf you want to install from source, you can do so by cloning the repo and be sure that the databases. These integrations allow developers to create versatile applications that combine the power of multiple APIs. This is magical. (LangChain OpenAI API) by TechLeadFirst look - ChatGPT + WolframAlpha (GPT-3.5) | QuickStart Tutorial for Beginners by RabbitMetricsLangChain Crash Course: Build an AutoGPT app in minutes with LangChain's built-in functions: tool-retrieval-agent, pii-protected-chatbot, pirate-speak, configurable-pirate-speak, plate-chain, and write code, fix bugs, or improve our infrastructure. Integrations: Help us integrate with your favorite vendor!

beta features is that given the rate of change of the field, being able to move quickly is still a priority. Mind not only log and visualize LLM apps, but also to actively debug, test and refine them. For anyone building Azure and other Microsoft products.  OpenAI All functionality related to OpenAI  More 201 iterations

3. What is RAG? RAG is a technique for augmenting LLM knowledge with additional data. LLMs can reason, etc.) This framework consists of several parts. LangChain Libraries: The Python and JavaScript libraries, etc.) This framework consists of several parts. LangChain Libraries: The Python and JavaScript libraries

antic LayerDatadog LogsDiffbotDiscordDocugamiDocusaurusDropboxDuckDBEmailEPubEtherscanEvent  
pip install --upgrade --quiet langchain pydanticimport osimport boto3comprehend\_client = boto3.client("comprehend")  
antic LayerDatadog LogsDiffbotDiscordDocugamiDocusaurusDropboxDuckDBEmailEPubEtherscanEvent  
ging FaceSpaCyTensorFlow HubText Embeddings InferenceTogether AIVolc EngineVoyage AIXorbits i  
ging FaceSpaCyTensorFlow HubText Embeddings InferenceTogether AIVolc EngineVoyage AIXorbits i  
noryCallbacksChat loadersAdaptersStoresComponentsChat modelsBedrock ChatBedrockChatAmazon I  
ndpointOctoAIllamaOpaquePromptsOpenAIOpenLLMOpenLMAliCloud PAI EASPetalsPipelineAIPredilection  
e underlying LLM provider. This obviously doesn't give you token-by-token streaming, which requires nat  
is walkthrough we'll work with an OpenAI LLM wrapper, although the functionalities highlighted are gene  
ndpointOctoAIllamaOpaquePromptsOpenAIOpenLLMOpenLMAliCloud PAI EASPetalsPipelineAIPredilection  
ndpointOctoAIllamaOpaquePromptsOpenAIOpenLLMOpenLMAliCloud PAI EASPetalsPipelineAIPredilection  
property that returns a string. Used for logging purposes only. There is a second optional thing it can imp  
ngchain.globals import set\_llm\_cachefrom langchain\_openai import OpenAI# To make the caching reall  
AI(model\_name="gpt-3.5-turbo-instruct", n=2, best\_of=2)with get\_openai\_callback() as cb: result = llm  
ChatOpenAI(model\_name="gpt-4")with get\_openai\_callback() as cb: result = llm.invoke("Tell me a joke")  
angchain.globals import set\_llm\_cachefrom langchain\_openai import ChatOpenAIllm = ChatOpenAI()In  
d by the underlying ChatModel provider. This obviously doesn't give you token-by-token streaming, whic  
ngchain-openaiAccessing the API requires an API key, which you can get by creating an account and he  
nantic\_similarity + (1.0 - decay\_rate) ^ hours\_passedNotably, hours\_passed refers to the hours passed  
st data into the system. This means that the information most relevant to a query may be buried in a doc  
ief: When models must access relevant information in the middle of long contexts, they tend to ignore th  
ashed and the hash is used as the key in the cache. The main supported way to initialized a CacheBack  
of characters.# This is a long document we can split up.with open("../state\_of\_the\_union.txt") as f:  
stens in your text you should use the same tokenizer as used in the language model.tiktokentiktoken is a  
him. At a high level, this splits into sentences, then groups into groups of 3 sentences, and then merges  
supported languages[e.value for e in Language]['cpp', 'go', 'java', 'kotlin', 'js', 'ts', 'php', 'proto', 'python', 'rs'  
a for each header "relevant" to any given chunk. It can return chunks element by element or combine ele  
wnstream.Using PyPDFLoad PDF using pypdf into array of documents, where each document contains  
cified jq schema to parse the JSON files. It uses the jq python package. Check this manual for a detailed  
toryLoader('../', glob="\*\*/\*.md")docs = loader.load()len(docs) 1Show a progress barBy default a progre  
own\_path = "../.../README.md"loader = UnstructuredMarkdownLoader(markdown\_path)data = loader.  
= CSVLoader(file\_path='./example\_data/mlb\_teams\_2012.csv')data = loader.load()print(data) [Docum  
redHTMLLoader("example\_data/fake-content.html")data = loader.load()data [Document(page\_content:  
aph = NeptuneGraph(host=host, port=port, use\_https=use\_https)from langchain.chains import NeptuneC  
a query language analogously to SQL or Cypher for these graphs. This notebook demonstrates the appli  
graph instance. You can download and run it in a local Docker container by executing the following script  
-fsSL nebula-up.siwei.io/install.sh | bashOther options are: - Install as a Docker Desktop Extension. See  
e.%pip install --upgrade --quiet networkxCreate the graphIn this section, we construct an example graph  
tance in their Aura cloud service. You can also run the database locally using the Neo4j Desktop applica  
kordb/falkordb:edgeOnce launched, you can simply start creating a database on the local machine and  
nd start creating a database on the local machine and connect to it:import kuzudb = kuzu.Database("tes  
docker run \ --name=graph \ -itd \ -p 8080:8080 \ hugegraph/hugegraphIf we want to connect Hu  
OT\_PASSWORD= arangodb/arangodbAn alternative is to use the ArangoDB Cloud Connector package  
ntities, relationships, and semantic meaning from unstructured text data.By coupling Diffbot's NLP API w

numbers of documents, performing this labelling process manually can be tedious. The OpenAIMetadata  
noryCallbacksChat loadersAdaptersStoresComponentsChat modelsAnthropicChatAnthropicThis notebook

with LCEL.[Legacy] Chains constructed by subclassing from a legacy Chain class. These chains do no  
with LCEL.[Legacy] Chains constructed by subclassing from a legacy Chain class. These chains do no  
itoryCallbacksChat loadersAdaptersStoresComponentsChat modelsAnthropic FunctionsAnthropic Funct  
with LCEL.[Legacy] Chains constructed by subclassing from a legacy Chain class. These chains do no  
ndpointOctoAIllamaOpaquePromptsOpenAIOpenLLMOpenLMAliCloud PAI EASPetalsPipelineAIPredil  
ndpointOctoAIllamaOpaquePromptsOpenAIOpenLLMOpenLMAliCloud PAI EASPetalsPipelineAIPredil  
ndpointOctoAIllamaOpaquePromptsOpenAIOpenLLMOpenLMAliCloud PAI EASPetalsPipelineAIPredil  
story class. This is a super lightweight wrapper that provides convenience methods for saving HumanMe  
e MultiVectorRetriever which makes querying this type of setup easy. A lot of the complexity lies in how i  
ain has a number of built-in document transformers that make it easy to split, combine, filter, and other  
ed on “distance”. But, retrieval may produce different results with subtle changes in query wording or if tr  
dexNucliaDBOpenSearchPostgres EmbeddingPGVecto.rsPGVectorPineconeQdrantRedisRocksetSca  
ctor store class to make it conform to the retriever interface. It uses the search methods implemented by  
ging FaceSpaCyTensorFlow HubText Embeddings InferenceTogether AIVolc EngineVoyage AIOrbits i  
antic LayerDatadog LogsDiffbotDiscordDocugamiDocusaurusDropboxDuckDBEmailEPubEtherscanEve  
antic LayerDatadog LogsDiffbotDiscordDocugamiDocusaurusDropboxDuckDBEmailEPubEtherscanEve  
h or document is embedded, the embedding process considers both the overall context and the relation  
", "\n", " ", ""). This has the effect of trying to keep all paragraphs (and then sentences, and then words) t  
antic LayerDatadog LogsDiffbotDiscordDocugamiDocusaurusDropboxDuckDBEmailEPubEtherscanEve  
a in CSV format. It is mostly optimized for question answering.NOTE: this agent calls the Pandas DataF  
signed to interact with a SQL databases. The agent builds off of SQLDatabaseChain and is designed to  
se agents to interact with a Pandas DataFrame. It is mostly optimized for question answering.NOTE: this  
antic LayerDatadog LogsDiffbotDiscordDocugamiDocusaurusDropboxDuckDBEmailEPubEtherscanEve  
antic LayerDatadog LogsDiffbotDiscordDocugamiDocusaurusDropboxDuckDBEmailEPubEtherscanEve  
ert total\_tokens > 0 with get\_openai\_callback() as cb: llm("What is the square root of 4?") llm("What is  
\_tool\_to\_openai\_function(t) for t in tools] message = model.predict\_messages([HumanMessage(conte  
Quickly debug a new chain, agent, or set of toolsCreate and manage datasets for fine-tuning, few-shot p  
containing arguments to call those functions. The goal of the OpenAI Function APIs is to more reliably r  
hubfrom langchain.agents import AgentExecutor, create\_react\_agentfrom langchain\_community.tools in  
all wikipediafrom langchain import hubfrom langchain.agents import AgentExecutor, create\_openai\_func  
from langchain.agents import AgentExecutor, create\_react\_agentfrom langchain\_community.tools impc  
onality with handle\_parsing\_errors! Let's explore how SetupWe will be using a wikipedia tool, so need tc  
e three prime numbers from a ToolMultiply these together.In this simple problem we can demonstrate ac  
iod example of this is an agent tasked with doing question-answering over some sources. Let's say we v  
he intermediate steps an agent takes.Let's take a look at how to do this.Set up the agentLet's set up a s  
tory, but we will then show how to add memory in. Memory is needed to enable conversation.Load the L  
ems! The name, description, and JSON schema can be used to prompt the LLM so it knows how to spe  
than it is intended for, but it likely won't produce results of the same quality.Supports Chat HistoryWheth  
output. This makes debugging these systems particularly tricky, and observability particularly important.  
ne action an agent should take. It has a tool property (which is the name of the tool that should be invoke  
that large language models are leaky abstractions! You'll have to use an LLM with sufficient capacity to c  
results based on the Reciprocal Rank Fusion algorithm.By leveraging the strengths of different algorithms  
MongoDB AtlasMyScaleOpenSearchPineconeQdrantRedisSupabaseTimescale Vector (Postgres) self-  
embeddings can most accurately reflect their meaning. If too long, then the embeddings can lose meanin  
as well as improve your vector search results.Crucially, the indexing API will work even with documents  
start guide for an introduction to output parsers and how to work with them.Output Parser TypesLangCha  
ovide two options: using OpenAI (a popular model available via API) or using a local open source mode  
answering questions, completing sentences, or engaging in a conversation.QuickstartThis quick start prc  
models that LangChain integrates with: LLMs and Chat Models. These are defined by their input and ou  
e to access some window of past messages directly. A more complex system will need to have a world i  
ChatPromptTemplatefrom langchain\_openai import ChatOpenAISearch = DuckDuckGoSearchRun(itemp  
e model context window.With LCEL, it's easy to add custom functionality for managing the size of promp





isitional-retrievalpython-lintrag-astradbrag-aws-bedrockrag-aws-kendrarag-chroma-multi-modal-multi-ve  
'unit\_tests and tests/integration\_tests. Make an improvementUpdate any affected example notebooks an  
before creating a new partner package. In the following sections, we'll walk through how to contribute to  
doesn't allow this setting to be enabled for forks in organizations (issue). If you are working in an organiz  
s/unit\_tests/test\_imports.py make testIntegration TestsIntegration tests cover logic that requires making i  
enciesQuarto - package that converts Jupyter notebooks (.ipynb files) into mdx files for serving in Docuse  
e hoodVectorDBQARetrievalQAMore general to all retrieversSequential ChainLCELObviated by LCELS  
s argument available throughout the API. This argument is list of handler objects, which are expected to  
<Handlerfrom langchain.chains import LLMLChainfrom langchain.prompts import PromptTemplatefrom lar  
load.loads as beta.Marked langchain\_core.beta.runnables.context.ContextGet and langchain\_core.beta.  
>frontAIIGitBookGoldenSerper - Google Search APIGooseAIGPT4AllGradientGraphsignalGrobidGu  
which should guide the model behavior.Hugging Face prompt injection identification: Detect and handle  
jingFaceHubfrom langchain\_openai import OpenAIllms = [ OpenAI(temperature=0), Cohere(model="  
using pydantic v2 throughout their code, but avoiding mixing v1 and v2 code for LangChain (see below).  
n this LLM on your device w/ acceptable latencyOpen-source LLMsUsers can now gain access to a rapi  
ls, you don't just want to send the same prompt to Anthropic - you probably want to use a different prom  
>frontAIIGitBookGoldenSerper - Google Search APIGooseAIGPT4AllGradientGraphsignalGrobidGu  
ite your applications. Evaluation and testing are both critical when thinking about deploying LLM applicat  
nplating, chat message generation, caching, vector embedding database creation, preprocessing, etc.C  
s]]: get the contents of multiple keys, returning None if the key does not existmset(key\_value\_pairs: Seq

This notebook shows how to use the iMessage chat loader. This classLangSmith Chat Datasets  
and usage analytics, user tracking, tracing and evaluation tools.PromptLayerPromptLayer is a pla  
oDB is a source-available cross-platform document-orientedMotörheadMotörhead is a memory se  
a UnderstandingOpenWeatherMapPolygon Stock Market APIPubMedPython REPLReddit SearchReque  
yte Question AnsweringThis notebook shows how to do question answering over structured data,  
a Weaviate Hybrid SearchWikipediaUsing the You.com RetrieverZepToolsAgents and toolkitsMemoryCa  
dexNucliaDBOpenSearchPostgres EmbeddingPGVecto.rsPGVectorPineconeQdrantRedisRocksetScal  
ging FaceSpaCyTensorFlow HubText Embeddings InferenceTogether AIVolc EngineVoyage AIXorbits i  
antic LayerDatadog LogsDiffbotDiscordDocugamiDocusaurusDropboxDuckDBEmailEPubEtherscanEven  
ments using theDoctran: interrogate documentsDocuments used in a vector store knowledge bas  
toryCallbacksChat loadersAdaptersStoresComponentsChat modelsChat modelsFeatures (natively supp  
ndpointOctoAIllamaOpaquePromptsOpenAIOpenLLMOpenLMAliCloud PAI EASPetalsPipelineAIPredil  
Microsoft Azure supports many programming languages, tools, and frameworks, including Microsoft-spe  
HubPythonJS/TSMoreHomepageBlogCopyright © 2024 LangChain, Inc.  
nodes with different capabilities and price points.ChatGPT is the Artificial Intelligence (AI) chatbot develo  
ng.To use, we should have the huggingface\_hub python package installed.pip install huggingface\_hubS  
"Sing a ballad of LangChain.")Gemini vision model supports image inputs when providing a single chat r  
in easily experiment with and evaluate top FMs for your use case, privately customize them with your da  
nsure that is the case. LangChain provides several utilities and helper functions to make sure prompts th  
ass is designed to provide a standard interface for all of them.Quick StartCheck out this quick start to ge  
ChatModel class is designed to provide a standard interface for all of them.Quick StartCheck out this qui  
s, only to return (or retrieve) them. Vector stores can be used as the backbone of a retriever, but there ar  
e takes care of storing embedded data and performing vector search for you.Get startedThis walkthrough  
or all of them.Embeddings create a vector representation of a piece of text. This is useful because it mea

chain has a number of built-in document transformers that make it easy to split, combine, filter, and others of any web page, or even for loading a transcript of a YouTube video. Document loaders provide a "loa

ory Compliance: Helps navigate strict data protection laws. Model Robustness: Can lead to better genera

Colab

Neo4j DB QA chain

This notebook shows how to use LLMs to provide a natural language

iy that does a few things:

- Keeps each top-level function and class in the code is loaded into separate doc
- | langchain-openai langchain playwright beautifulsoup4playwright install# Set env var OPENAI\_API\_KEY
- openai# Set env var OPENAI\_API\_KEY or load from a .env file:# import dotenv# dotenv.load\_dotenv()

tuff" all your documents into a single prompt. This is the simplest approach (see here for more on the St

at structure LLM responses. Only some LLMs support functions (e.g., OpenAI), and they are more genera

APIs are already compatible with OpenAI function calling. For example, Klarna has a YAML file that desc

LLM. Document LoadingFirst, install packages needed for local embeddings and vector storage.%pip in:

et langchain langchain-community langchainhub langchain-openai faiss-cpu

The RetrieverTo start, we ne

/e'll use an OpenAI chat model and embeddings and a Chroma vector store in this walkthrough, but eve

1: Make sure the retriever you are using supports multiple usersAt the moment, there is no unified flag c

ement is covered here. We'll work off of the Q&A app we built over the LLM Powered Autonomous Agen

Q&A techniques. We'll also see how LangSmith can help us trace and understand our application. Lang

OpenAI chat model and embeddings and a Chroma vector store in this walkthrough, but everything sho

ors and tools.

► GitHub IssuesOur issues page is kept up to date with bugs, improvements, and feature

estionsCreating chatbots that can answer questions based on database dataBuilding custom dashboard:

rather than raw text. Several components are important to consider for chat:chat model: See here for a l

cles to your LLM application. Crucially, this is NOT a DAG framework. If you want to build a DAG, you sh

on about wide-ranging topics, but their knowledge is limited to the public data up to a specific point in tim

schemas automatically inferred from your LangChain object, and enforced on every API call, with rich ei

ack in your JS application (link).Building an automated feedback pipeline (link).How to evaluate and audi

1 with LCEL.[Legacy] Chains constructed by subclassing from a legacy Chain class. These chains do no

itsThere are several key concepts to understand when building agents: Agents, AgentExecutor, Tools, T

umentation covers everything related to the retrieval step - e.g. the fetching of the data. Although this so

] if you want to get an overview of the functionality.PromptsThis section deep dives into the different type

idel I/OAdditionalCommunityDiscordTwitterGitHubPythonJS/TSMoreHomepageBlogCopyright © 2024 L

age is a good place to start.

Prompt + LLMThe most common and valuable composition is taking:

objects is itself an LCEL object. 2. Composition primitives: LCEL provides a number of primitives that mak

nablePassthrough: Passing data throughpassing-data-through} RunnableLambda: Run Custom Fu

nc methods:astream: stream back chunks of the response asyncainvoke: call the chain on an input asyn

enaifrom langchain\_core.output\_parsers import StrOutputParserfrom langchain\_core.prompts import Cr

/you build your chains with LCEL you get the best possible time-to-first-token (time elapsed until the first c

ation's need. Granting broad or excessive permissions can introduce significant security vulnerabilities.

.SetupJupyter NotebookThis guide (and most of the other guides in the documentation) use Jupyter note

e directory is PATH/TO/REPO/langchain/libs/langchain running:pip install -e .LangChain communityThe

of LLMs with the ability to access, interact with and manipulate external resources.NextIntroductionCom

.5 and Wolfram|Alpha via LangChain by James Weaver) by Dr Alan D. ThompsonLangChain explained

n 25 minutes by Nicholas RenotteLangChain Crash Course - Build apps with language models by Patric

isitional-retrievalpython-lintrag-astradbrag-aws-bedrockrag-aws-kendrarag-chroma-multi-modal-multi-ver

ors and tools.

► GitHub IssuesOur issues page is kept up to date with bugs, improvements, and feature

ior version increases will occur for:Breaking changes for any public interfaces marked as beta. Patch ver

ing production-grade LLM applications, we highly recommend using a platform like this.set\_debug and s

msNextAnthropicCommunityDiscordTwitterGitHubPythonJS/TSMoreHomepageBlogCopyright © 2024 L

on about wide-ranging topics, but their knowledge is limited to the public data up to a specific point in tim

es. Contains interfaces and integrations for a myriad of components, a basic run time for combining thes

es. Contains interfaces and integrations for a myriad of components, a basic run time for combining thes

```
rNoteFacebook ChatFaunaFigmaGeopandasGitGitBookGitHub Google BigQueryGoogle Cloud Storage  
'comprehend', region_name="us-east-1")from langchain_experimental.comprehend_moderation import  
rNoteFacebook ChatFaunaFigmaGeopandasGitGitBookGitHub Google BigQueryGoogle Cloud Storage  
inference (Xinference)YandexGPTVector storesRetrieversToolsAgents and toolkitsMemoryCallbacksCh  
inference (Xinference)YandexGPTVector storesRetrieversToolsAgents and toolkitsMemoryCallbacksCh  
Bedrock is a fully managed service that offers a choice of high-performing foundation models (FMs) from  
basePrediction GuardPromptLayer OpenAIRELLMReplicateRunhouseSageMakerEndpointStochasticAI  
tive support from the LLM provider, but ensures your code that expects an iterator of tokens ca  
eric for all LLM types. SetupFor this example we'll need to install the OpenAI Python package:  
basePrediction GuardPromptLayer OpenAIRELLMReplicateRunhouseSageMakerEndpointStochasticAI  
basePrediction GuardPromptLayer OpenAIRELLMReplicateRunhouseSageMakerEndpointStochasticAI  
lement: An _identifying_params property that is used to help with printing of this class. Should r  
ly obvious, lets use a slower model.llm = OpenAI(model_name="gpt-3.5-turbo-instruct", n=2, best_of=2)  
.invoke("Tell me a joke") print(cb)Tokens Used: 37 Prompt Tokens: 4 Completion  
<e> print(cb)Tokens Used: 24 Prompt Tokens: 11 Completion Tokens: 13Successfu  
Memory Cache%%timefrom langchain.cache import InMemoryCacheset_llm_cache(InMemoryCache())  
ch requires native support from the ChatModel provider, but ensures your code that expects an i  
ead here. Once we have a key we'll want to set it as an environment variable by runnin  
since the object in the retriever was last accessed, not since it was created. This mea  
ument with a lot of irrelevant text. Passing that full document through your application can lead  
ne provided documents. See: https://arxiv.org/abs/2307.03172 To avoid this issue you can re-order  
kedEmbeddings is from_bytes_store. This takes in the following parameters: underlying_embedder: The  
ate_of_the_union = f.read()from langchain.text_splitter import CharacterTextS  
fast BPE tokenizer created by OpenAI. We can use it to estimate tokens used. It will p  
one that are similar in the embedding space. Install Dependencies!pip install --quiet  
st', 'ruby', 'rust', 'scala', 'swift', 'markdown',  
ments with the same metadata, with the objectives of (a) keeping related text grouped (more or  
the page content and metadata with page number.pip install pypdffrom langchain_community.document  
l documentation of the jq syntax.#!pip install jqfrom langchain_community.document_loaders imp  
ss bar will not be shown. To show a progress bar, install the tqdm libra  
ider.load()data [Document(page_content="ñ\x9f\x9c\x8f\x9f\x97 LangChain\n\nñ\x9a; Building app  
nent(page_content='Team: Nationals\n"Payroll (millions)": 81.34\n"Wins": 98', lookup_str="", metadata=  
='My First Heading\n\nMy first paragraph.', lookup_str="", metadata={'source': 'example_data/fake-conte  
OpenCypherQACChainfrom langchain_openai import ChatOpenAlllm = ChatOpenAI(temperature=0, mod  
ication of LLMs as a natural language interface to a graph database by generating SPARQL.Disclaimer:  
t:docker run \ -it \ -p 7687:7687 \ -p 7444:7444 \ -p 3000:3000 \ -e MEMGRAPH="--bolt-server-  
e here - NebulaGraph Cloud Service. See here - Deploy from package, source code, or via Kubernetes.  
n. At the moment, this works best for small pieces of text.from langchain.indexes import GraphIndexCre  
ation, or running a docker container. You can run a local docker container by running the executing the fc  
connect to it.from langchain.chains import FalkorDBQACChainfrom langchain_community.graphs import I  
it_db")conn = kuzu.Connection(db)First, we create the schema for a simple movie database:conn.execu  
geGraph in the application, we need to install python sdk:pip3 install hugegraph-pythonIf you are using t  
o get a temporary cloud instance running:%%capture%pip install --upgrade --quiet python-arango # T  
with Neo4j, a graph database, you can create powerful, dynamic graph structures based on the informati
```

Tagger document transformer automates this process by extracting metadata from each provided docu  
ok covers how to get started with Anthropic chat models.from langchain.schema import HumanMessage

it use LCEL under the hood but are rather standalone classes. We are working creating  
it use LCEL under the hood but are rather standalone classes. We are working creating  
tionsThis notebook shows how to use an experimental wrapper around Anthropic that gives it the same ,  
it use LCEL under the hood but are rather standalone classes. We are working creating  
basePrediction GuardPromptLayer OpenAIRELLMReplicateRunhouseSageMakerEndpointStochasticAI  
basePrediction GuardPromptLayer OpenAIRELLMReplicateRunhouseSageMakerEndpointStochasticAI  
basePrediction GuardPromptLayer OpenAIRELLMReplicateRunhouseSageMakerEndpointStochasticAI  
ssages, AIMessages, and then fetching them all. You may want to use this class directly if you are mana  
to create the multiple vectors per document. This notebook covers some of the common ways  
wise manipulate documents. When you want to deal with long pieces of text, it is necess  
ne embeddings do not capture the semantics of the data well. Prompt engineering / tuning is someti  
INNSemaDBSingleStoreDBscikit-learnSQLite-VSSStarRocksSupabase (Postgres)SurrealDBTairTencent  
' a vector store, like similarity search and MMR, to query the texts in the vector stor  
inference (Xinference)YandexGPTVector storesRetrieversToolsAgents and toolkitsMemoryCallbacksCh  
rNoteFacebook ChatFaunaFigmaGeopandasGitGitBookGitHubGoogle BigQueryGoogle Cloud Storage  
rNoteFacebook ChatFaunaFigmaGeopandasGitGitBookGitHubGoogle BigQueryGoogle Cloud Storage  
ships between the sentences and phrases within the text. This can result in a more comprehensive vect  
together as long as possible, as those would generically seem to be the  
rNoteFacebook ChatFaunaFigmaGeopandasGitGitBookGitHubGoogle BigQueryGoogle Cloud Storage  
rame agent under the hood, which in turn calls the Python agent, which executes LLM generated Python  
answer more general questions about a database, as well as recover from errors. Note that, as this ager  
s agent calls the Python agent under the hood, which executes LLM generated Python code - this can be  
rNoteFacebook ChatFaunaFigmaGeopandasGitGitBookGitHubGoogle BigQueryGoogle Cloud Storage  
rNoteFacebook ChatFaunaFigmaGeopandasGitGitBookGitHubGoogle BigQueryGoogle Cloud Storage  
s the square root of 4?"})assert cb.total\_tokens == total\_tokens \* 2# You can ki  
nt="move file foo to bar")], functions=functions)messageAIMessage(content=", additional\_kwargs={'func  
prompting, and evaluationRun regression tests on your application to confidently developCapture p  
return valid and useful function calls than a generic text completion or chat API. The OpenAI Funct  
nport WikipediaQueryRunfrom langchain\_community.utilities import WikipediaAPIWrapperfrom langcha  
ctions\_agentfrom langchain\_community.tools import WikipediaQueryRunfrom langchain\_com  
ort WikipediaQueryRunfrom langchain\_community.utilities import WikipediaAPIWrapperfrom langchain\_()  
o install that%pip install --upgrade --quiet wikipediafrom langchain import hubfro  
dding some logic to verify intermediate steps by checking whether their outputs are prime.fro  
want the agent to respond not only with the answer, but also a list of the sources used. We then  
simple agent for demonstration purposes. For our tool, we will use Tavily. Make sure that y  
.LMFirst, let's load the language model we're going to use to control the agent.from langchain\_o  
ecify what action to take, and then the function to call is equivalent to taking that  
mer or not these agent types support chat history. If it does, that means it can be used  
LangSmith is especially useful for such cases. When building with LangChain, all s  
ed) and a tool\_input property (the input to that tool)AgentFinishThis represents the f  
generate well-formed YAML. In the OpenAI family, DaVinci can do reliably but Curie's ability already dro  
s, the EnsembleRetriever can achieve better performance than any single algorithm. The most com  
queryingVectara self-queryingWeaviateSingleStoreDBSVMTavily Search APITF-IDFVespaWeaviate Hy  
ng. You want to have long enough documents that the context of each chunk is retained. The Paren  
s that have gone through several transformation steps (e.g., via text chunking) with res  
ain has lots of different types of output parsers. This is a list of output parsers L  
el.OpenAILocalFirst we'll need to install their partner package:pip install langcha  
ovides a basic overview of how to work with prompts. How-To GuidesWe have many how-to guides for  
tput types. LLMs in LangChain refer to pure text completion models. The APIs they  
model that it is constantly updating, which allows it to do things like maintain information about entities a  
plate = """turn the following user input into a search query for a search engine:{input}"""prompt = ChatPr  
pts within your chain or agent. Let's look at simple agent example that can search Wikipedia for informa





ctorrag-chroma-multi-modalrag-chroma-privaterag-chromarag-codellama-fireworksrag-conver  
id documentation. These live in docs. Update unit and integration tests w  
each of these packages from a fake company, Parrot Link AI. Community Package The langchain-comm  
zation, we recommend submitting your PR from a personal fork in order to enable  
calls to outside APIs (often integratio  
aurus.poetry install from the monorepo rootBuildingIn the following commands, the pref  
SimpleSequentialChainLCEL Obviated by LCEL TransformChainLCEL/RunnableLambda Obviated by LCE  
o implement one or more of the methods described below in more detail. Callback handl  
ngchain\_openai import OpenAI from loguru import logger logfile = "output.log" logger.ad  
.runnables.context.ContextSet as beta. PreviousChangelog Next langchain 0.1.7 (Jan 5, 2024) Community  
itenbergHacker News Hazy Research Helicone Hologres HTML to text i Fixit! MSDB! Infinity Infino Jaguar J  
e prompt injection attacks. Logical Fallacy chain: Checks the model output against logical fa  
"command-xlarge-20221108", max\_tokens=20, temperature=0), HuggingFaceHub(repo\_id="google/fla  
.User can either pin to pydantic v1, and upgrade their code in one go once LangCh  
idly growing set of open-source LLMs. These LLMs can be assessed across at lea  
ipt template and send a different version there. Fallback for LLM API Errors This is m  
itenbergHacker News Hazy Research Helicone Hologres HTML to text i Fixit! MSDB! Infinity Infino Jaguar J  
tions, since production environments require repeatable and useful outcomes. LangChain  
ase 2: Self-hosted Open-Source Models Alternatively, developers can opt to use smaller, yet c  
quence[Tuple[str, bytes]]) -> None: set the contents of multiple keys m delete(key):

This notebook demonstrates an easy way to load a LangSmith chat dataset LangSmith LLM Runs  
atform SageMaker Tracking Amazon SageMaker is a fully Streamlit Streamlit is a fa  
server Neo4j Neo4j is an open-source graph Postgres PostgreSQL also known as Redis|I  
ests Scene Xplain Search Tools Search API SearxNG Search Semantic Scholar API ToolSerp API StackExch  
Amadeus This notebook walks you through connecting LangChain to the Amadeus Azure Cognitive  
Callbacks Chat loaders Adapters Stores Components Retrievers Retrievers ActiveLoop Deep Memory [Ac  
NSema DB Single Store DB scikit-learn SQLite VSS StarRocks Supabase (Postgres) SurrealDB Tair Tencent  
inference (Xinference) Yandex GPT Vector stores Retrievers Tools Agents and toolkits Memory Callbacks Ch  
rNote Facebook Chat Fauna Figma Geopandas Git GitBook GitHub Google BigQuery Google Cloud Storage  
se are typically stored in Doctran: language translation Comparing documents through embeddings  
ported) All ChatModels implement the Runnable interface, which comes with default implementations of a  
basePrediction Guard Prompt Layer OpenAIRE LLM Replicate Runhouse SageMaker Endpoint Stochastic AI  
cific and third-party software and systems. Azure OpenAI is an Azure service with powerful

oped by OpenAI. Installation and Setup Install the integration package with pip insta  
See a usage example. from langchain\_community.llms import HuggingFaceHub Hugging Face Local Pipe  
message. Example: from langchain\_core.messages import HumanMessage from langchain\_google\_gen  
ata using techniques such as fine-tuning and Retrieval Augmented Generation (RAG), and build agents t  
hat you write - whether formatted as a string or as a list of messages - end up form  
et an overview of working with LLMs, including all the different methods they expose Integ  
ick start to get an overview of working with ChatModels, including all the different methods  
re other types of retrievers as well. Retrievers accept a string query as input and re  
gh showcases basic functionality related to vector stores. A key part of working with vect  
ans we can think about text in the vector space, and do things like semantic search where we look f

wise manipulate documents. When you want to deal with long pieces of text, it is necessary to use the "load\_documents" method for loading data as documents from a configured source. They optionally implement a "load\_data" method for loading data as documents from a configured source. They optionally implement a "load\_data" method for loading data as documents from a configured source.

This notebook shows how to use LLMs to provide a natural language interface. It uses LangChain's `load\_documents` method to load data from a file. It then uses the `create\_tagging\_chain` function to create a chain that can tag documents with their category. Finally, it uses the `map\_reduce` function to summarize each document on its own. This is useful for quickly prototyping AI models without having to worry about parsing the data.

The code here is mostly just example code. Feel free to try anything shown here with any ChatModel or LLM, Embeddings, and VectorStore or Retriever. We'll show how to filter for this in LangChain. Rather, each vectorstore and retriever may have their own blog post by Lilian Weng in the Quickstart. We'll need to update two things about our existing app: Smith will become increasingly helpful as our application grows in complexity. Architecture We'll create a new endpoint for LangChain's expression language. Cycles are important for agent-like behavior because they were trained on. If you want to build AI applications that can reason about private data or error messages, see the API docs page with JSONSchema and Swagger (insert example link). Efficient RAG workflows (link). How to fine-tune an LLM on real usage data (link). How to use the LCEL under the hood but are rather standalone classes. We are working on creating Toolkits. For an in-depth explanation, please check out this conceptual guide.

Agent Toolkit sounds simple, it can be subtly complex. This encompasses several key modules. Examples of prompt templates and how to use them. LLMs This section covers functional LangChain, Inc.

RAGLet's look at adding in a retrieval step to a prompt and LLM, which adds multiple chains. It makes it easy to compose chains, parallelize components, add fallbacks, dynamically configure chain internal functions, run custom functions, etc. RunnableBranch: Dynamically route logic based on input. dynamicall icabatch: call the chain on a list of inputs. asynccastream\_log: stream back intermediate steps. ChatPromptTemplate from langchain\_openai import ChatOpenAI prompt = ChatPromptTemplate.from\_template(chunk of output comes out). For some chains this means eg. we stream tokens straight from an LLM to the client. To avoid such vulnerabilities, consider using read-only credentials, disallowing a Jupyter notebooks and assume the reader is as well. Jupyter notebooks are perfect for learning how to work with LangChain. The LangChain community package contains third-party integrations. It includes Discord, Twitter, GitHub, Python, JS, TS, More, Homepage, Blog, Copyright © 2024 LangChain, Inc.

- The hottest new Python framework by AssemblyAIChatbot with INFINITE MEMORY using OpenAI & Fck LoeberTutorialsLangChain for Gen AI and LLMs by James Briggs#1 Getting Started with GPT-3 vs. ChromaDBrag-chroma-multi-modalrag-chroma-private rag-chromarag-codellama-fireworksrag-conver requests. There is a taxonomy of labels to help with sorting and discovery of issues of integration increases will occur for: Bug fixes, New features, Any changes to private set\_verbose if you're prototyping in Jupyter Notebooks or running Python scripts. LangChain, Inc.

ne that they were trained on. If you want to build AI applications that can reason about private data or use components into chains and agents, and off-the-shelf implementations of chains and agents. Use components into chains and agents, and off-the-shelf implementations of chains and agents.

DirectoryGoogle Clo

DirectoryGoogle Clo

iat loadersAdaptersStoresComponen

iat loadersAdaptersStoresComponen

n leading AI companies like AI

INebula (Symbl.ai)TextGenTitan

INebula (Symbl.ai)TextGenTitan

INebula (Symbl.ai)TextGenTitan

!#

embedder to use for embeddi

t\_loaders imp

olicat

nt.html'}, lookup\_

lel="gpt-4")chain = NeptuneOpenCypherQACChain.from\_I

: To date, SPARQL query generation via LLMs is sti

-name-for-init=Ne

. See hereOnce the cluster

atorfrom lang

ollowing script:docker ru

FalkorDBGraphfrom langchain\_ope

ite("CREATE NODE TAB

the docker cont

The ArangoDB Python Driver%pip install --upgra

ion extracted from

ment acco

efrom langchain\_community.chat

API as OpenAI Functions.fro

INebula (Symbl.ai)TextGenTitan  
INebula (Symbl.ai)TextGenTitan  
INebula (Symbl.ai)TextGenTitan  
aging memory outsi

t Cloud VectorDBTigrisTileDBTi

lat loadersAdaptersStoresComponen  
DirectoryGoogle Clo  
DirectoryGoogle Clo  
t

DirectoryGoogle Clo  
n code - this can be bad  
nt is in active development,  
e bad if the LLM generated P  
DirectoryGoogle Clo  
DirectoryGoogle Clo

ctio

opena

)p

ibrid SearchWikipediaUsi

and their r  
'omptTempl

```
hain.invoke({  
    chatOpenAIModel = Cha  
}  
:  
nple DB for this  
nd i  
s/ChatModels  
_openai imp  
BaseMessag  
||  
sy  
  
trac  
ie  
e arguments  
g helps p  
er inp  
unpacks  
, will si  
t Cloud VectorDBTigrisTileDBTi  
Search API is a sea
```



run

ELcreate\_tagging\_c

DiscordTw

an

This notebook demonstrates h

Redis (Remote Dictionary) Remembrall  
RangeTavily SearchTwilioWikipediaWolfram  
Cloud ServicesThis toolkit is us  
ctiveloop Deep Ama  
t Cloud VectorDBTigrisTileDBTi  
flat loadersAdaptersStoresComponen  
DirectoryGoogle Clo  
s has the benefit  
all methods, ie. ainvoke, b  
INebula (Symbl.ai)TextGenTitan

```
eline
ai import Ch
that
```

Graph QA ChainThis notebook shows how t

I u

:abases:

:Runnables can e

:

ly-route-logic-base

iplate("

a str

Pinecone - GPT-3, Embeddings, ADA, Vector

Open Source











