

More than ChatGPT: The benefits of Large Language Models for agriculture

Project *diabek*

Dr. Sebastian Fritsch,
Heiko Fabritius

20.10.2023

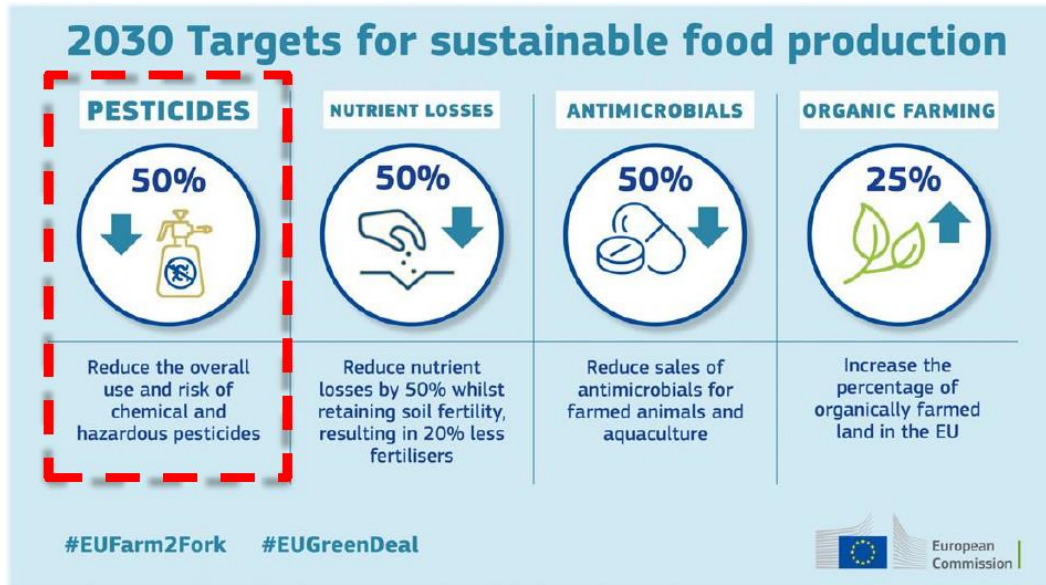


(Created with ChatGPT using the DALL-E 3 model)

Project background

- **Project Diabek:** Application and review of existing digital technologies (e.g. *Precision Farming*)
- **Living Lab:** Problem deep dive and development of new solutions in cooperation with farmers (based on AI)

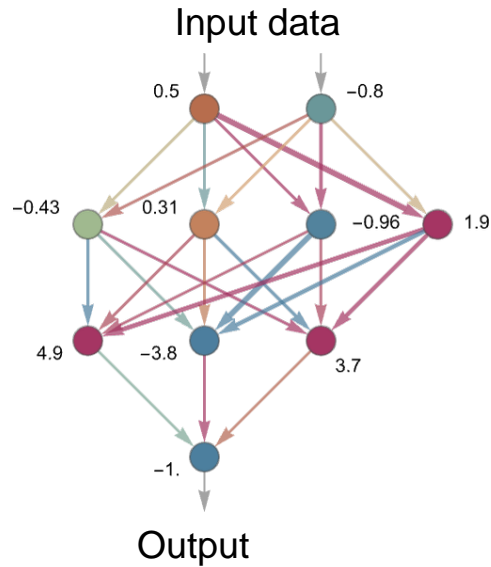
Why are we doing this?



Goals:

- Comfort for farmers
- Reduction of errors and securing compliance
- Reduction of pesticides

How can AI help achieve these goals?



Large Language
Models

{ The best thing about AI is its ability to,

The best thing about AI is its ability to create,

The best thing about AI is its ability to create worlds,

The best thing about AI is its ability to create worlds that,

The best thing about AI is its ability to create worlds that are,

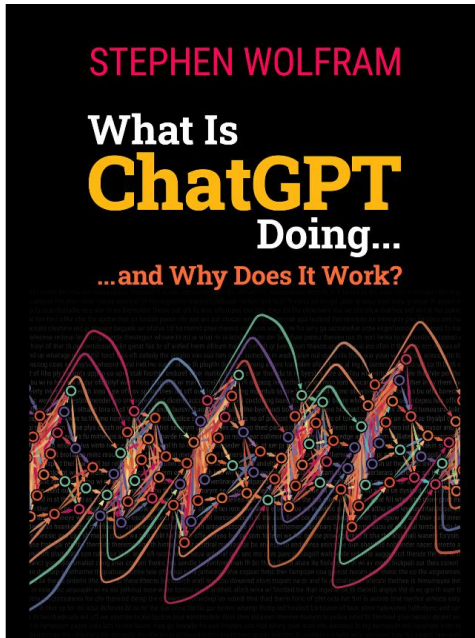
The best thing about AI is its ability to create worlds that are both,

The best thing about AI is its ability to create worlds that are both exciting,

The best thing about AI is its ability to create worlds that are both exciting, }

Image recognition

How do Large Language Models (LLMs) work?



The best thing about AI is its ability to

learn	4.5%
predict	3.5%
make	3.2%
understand	3.1%
do	2.9%

`Out[]=` { The best thing about AI is its ability to,
The best thing about AI is its ability to learn,
The best thing about AI is its ability to learn from,
The best thing about AI is its ability to learn from experience,
The best thing about AI is its ability to learn from experience.,
The best thing about AI is its ability to learn from experience. It,
The best thing about AI is its ability to learn from experience. It's,
The best thing about AI is its ability to learn from experience. It's not }

Why is ChatGPT not suited for our task?

Digitalisierung

Agrar-Professor lässt KI düngen - das Ergebnis überrascht



© IMAGO/imagebroker Kann eine künstliche Intelligenz die Düngung selbstständig planen? Ein Professor hat das ausprobiert.

 Teilen  Twitttern  Pinnen  XING  Mail  Druck



Thomas Göggerle, agrarheute
am Freitag, 14.04.2023 - 06:02 (7 Kommentare)

Ein Professor an einer Agrarfachhochschule stellte der künstlichen Intelligenz ChatGPT eine Aufgabe: Er soll eine Empfehlung zur Düngung von Weizen erstellen und dabei einige Parameter berücksichtigen. Die KI erledigt den Job und das Ergebnis hat viele überrascht.

Düngeempfehlungen zu erstellen ist komplex geworden. Neben dem Bedarf der Pflanzen müssen auch Vorfrucht, maximale Stickstoffmenge und Nmin-Werte berücksichtigt werden. Professor Peter Breunig von der Fachhochschule Weihenstephan-Triesdorf ist eigentlich Agrarökonom. Er sitzt aber auch als Landwirt auf dem Traktor und streut schon mal selbst Dünger. Er hat den Chatbot ChatGPT eine Aufgabe erstellt.

- Incorrect answers
- Hallucinations of the model
- Results are non-deterministic

Chatbot development using Langchain and Retrieval Augmented Generation (RAG)

Vector database



Prompt Engineering

GPT 3.5 or 4



+



+

```
primer = f"""Deine Rolle ist die eines landwirtschaftlichen Beraters.
Die Nutzer stellen dir Fragen zu spezifischen Präparaten und du gibst Ihnen Antworten mit Informationen zur Zulassung,
den Auflagen und weiteren Bemerkungen zu diesen Präparaten.

Merke dir, in der Datenbank steht W für Winterweizen, G für "Gerste", T für "Triticale", WW für "Winterweizen", WG für "Wintergerste",
WR für "Winterraps", WT für Winter-Triticale, SH für Sommerhafer, WDU für Winterdurum und DI steht für Dinkel.

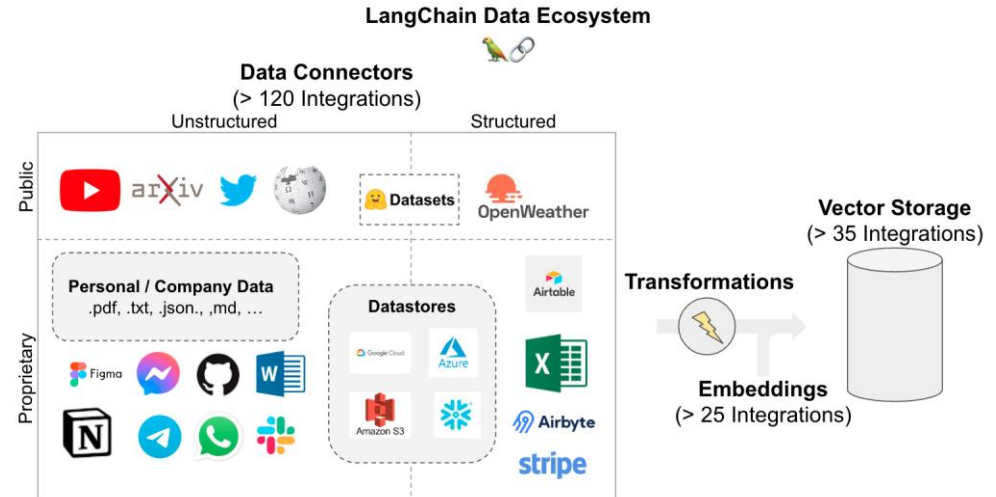
Wenn du die Antwort nicht in der Datenbank finden kannst, antworte nur "Ich kann Ihnen leider nicht weiterhelfen" und stoppe anschließend.
Weigere dich jegliche Fragen zu beantworten bei denen du dir nicht sicher bist.

Antworte in Deutsch und so präzise wie möglich.

Antworte immer in der folgenden Reihenfolge:
1. Zulassung des Präparats
2. Auflage
3. weitere Bemerkungen und Anbauhinweise

Gib zu allen Antworten eine Quelle an, damit der Nutzer bei Bedarf deine Antworten nachprüfen kann.
....
```


What is LangChain?



Source: <https://langchain.com>; https://python.langchain.com/docs/use_cases/question_answering.html

General Framework:

Open source UI visual tool to build
customized LLM flow using Langchain:

Hosting of the chatbot:



LangChain

FlowiseAI

render

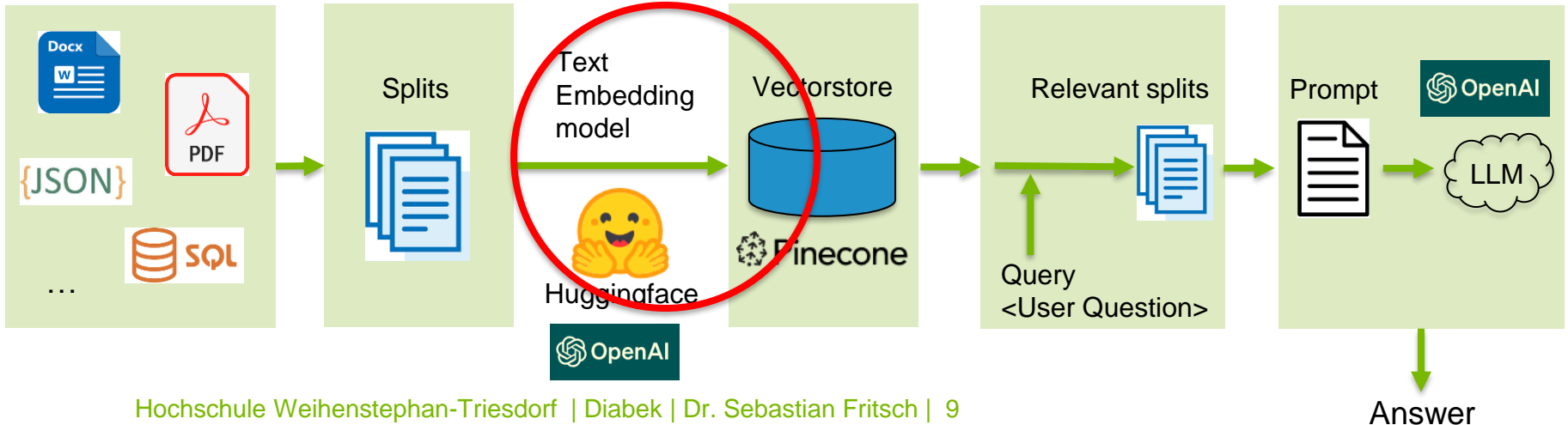
Document and
data loaders

Text Splitting

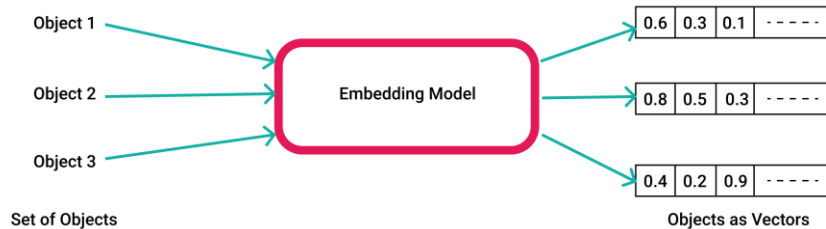
Data storage

Retrieval

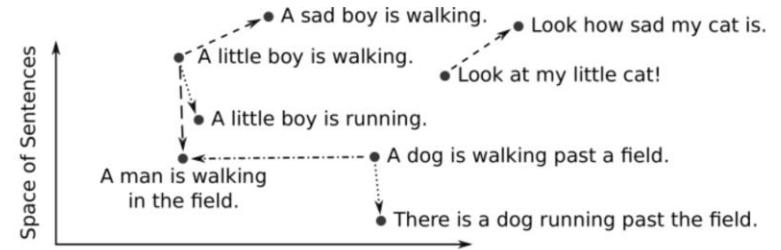
Output



How does text embedding work?



Source: <https://www.pinecone.io/learn/vector-embeddings/>



Source: DeepAI

- Deep neural networks commonly used for embeddings (transformer models)
- Resulting embeddings usually highly dimensional (up to 2000 dimensions)
- Common models: Word2Vec, BERT

Hugging Face is way more fun with friends and colleagues! 🥳 [Join an organization](#)

[Dismiss this message](#)

+ New

HarrierDuBois

- Profile
- Inbox (0)
- Settings
- Get **Pro**

Organizations

+ Create New

Resources

- Hub guide
- Transformers doc
- Forum
- Tasks
- Learn

☀️ Light theme

Sebastian Fritsch's Activity

All Models Datasets Spaces Collections Community Upvotes Likes

Updated a space 1 minute ago

Bert Base Uncased

Liked a space about 1 month ago

Auto Evaluator

Updated a space about 1 month ago

Stabilityai Stable Diffusion XL Base 1.0

Trending last 7 days

All Models Datasets Spaces

HuggingFaceH4/zephyr-7b-alpha

Text Generation • Updated 1 day... • ↓ 26.6k • ♥ 587

mistralai/Mistral-7B-v0.1

Text Generation • Updated 6 day... • ↓ 192k • ♥ 1.29k

open-web-math/open-web-math

Viewer • Updated about 14 hours ago • ↓ 408 • ♥ 123

mistralai/Mistral-7B-Instruct-v0.1

Text Generation • Updated 7 days... • ↓ 154k • ♥ 762

replit/replit-code-v1_5-3b

Text Generation • Updated about 12... • ↓ 4k • ♥ 192

Running on A10G

IllusionDiffusion

♥ 2.21k

stabilityai/stable-diffusion-xl-base-1..

Text-to-Image • Updated 16 da... • ↓ 5.56M • ♥ 3.16k

(huggingface.co)

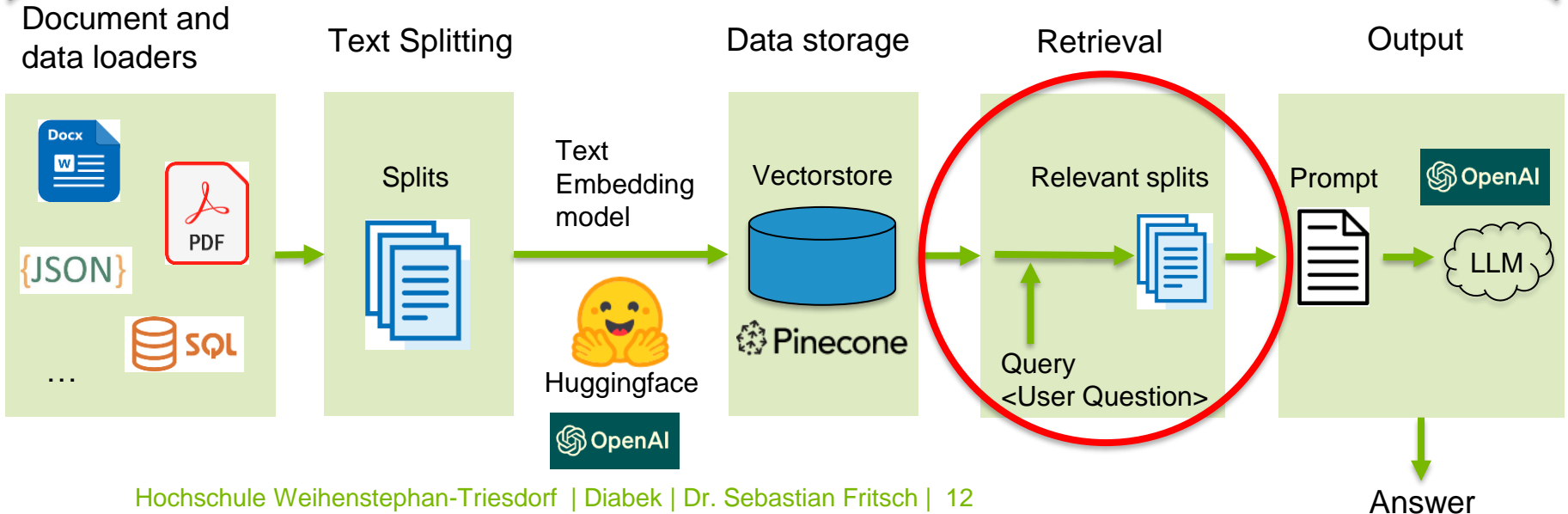
General Framework:



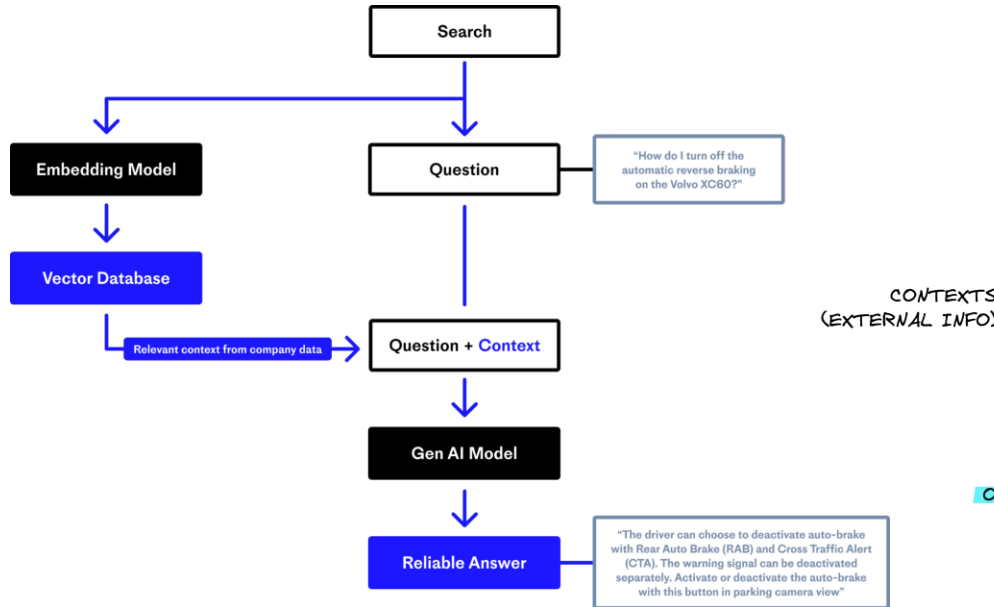
Open source UI visual tool to build
customized LLM flow using Langchain:



Hosting of the chatbot:



What is RAG?



INSTRUCTIONS

"" Answer the question based on the context below. If the question cannot be answered using the information provided answer with "I don't know".

CONTEXTS (EXTERNAL INFO)

Context: Large Language Models (LLMs) are the latest models used in NLP. Their superior performance over smaller models has made them incredibly useful for developers building NLP enabled applications. These models can be accessed via Hugging Face's 'transformers' library, via OpenAI using the 'openai' library, and via Cohere using the 'cohere' library.

Question: Which libraries and model providers offer LLMs?

Answer: ""

PROMPTER INPUT

OUTPUT INDICATOR

Source: <https://www.pinecone.io/learn/retrieval-augmented-generation/>

Source: <https://www.pinecone.io/learn/series/langchain/langchain-prompt-templates/>

General Framework:



Open source UI visual tool to build
customized LLM flow using Langchain:



Hosting of the chatbot:



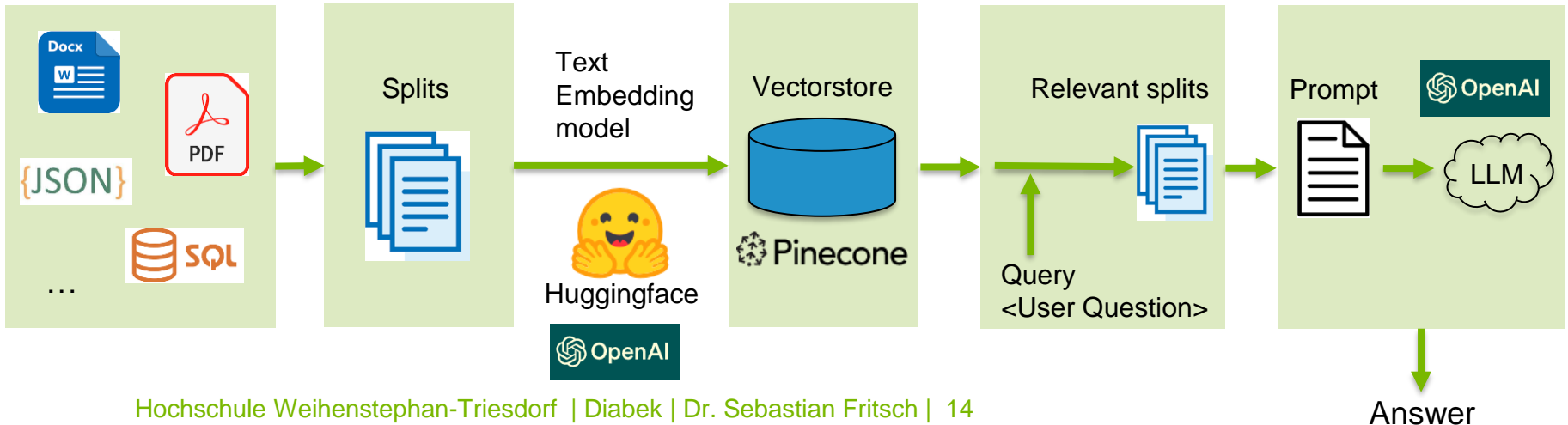
Document and
data loaders

Text Splitting

Data storage

Retrieval

Output



No code chatbot development with Flowise

```
Upload_PDF_to_Pinecone.ipynb ☆
Datei Bearbeiten Anzeige Einfügen Laufzeit Tools Hilfe Alle Änderungen wurden gespeichert

+ Code + Text

[ ] from langchain.embeddings import HuggingFaceInstructEmbeddings
instructor_embeddings = HuggingFaceInstructEmbeddings(model_name="hknlp/instructor-xl", model_kwargs={"device": "cuda",
embeddings = instructor_embeddings

[ ] from langchain.embeddings import HuggingFaceEmbeddings, SentenceTransformerEmbeddings
sentence_embeddings = HuggingFaceEmbeddings(model_name="multi-qa-mpnet-base-dot-v1", model_kwargs={"device": "cuda",
embeddings_sen = sentence_embeddings

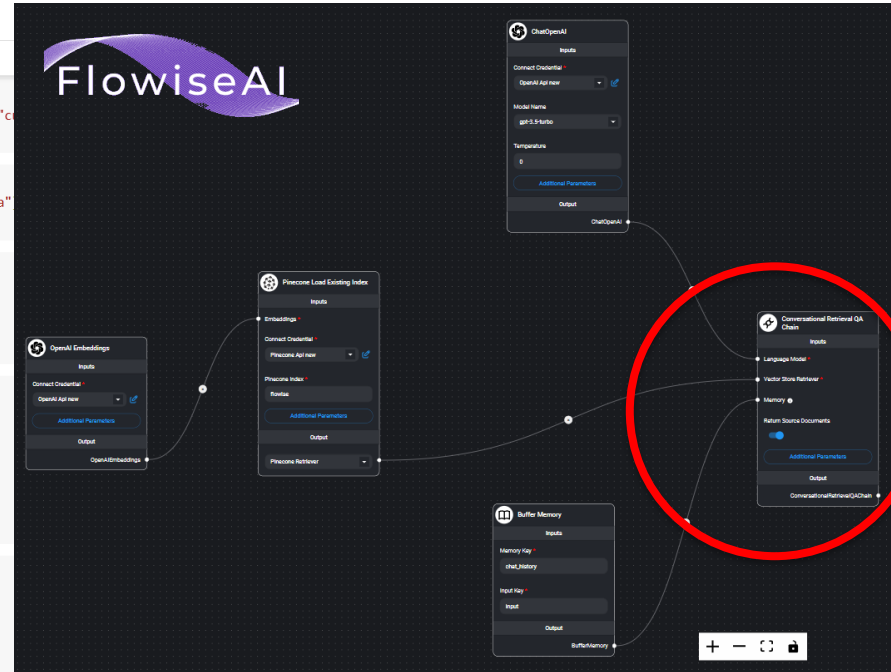
[ ] from sentence_transformers import SentenceTransformer

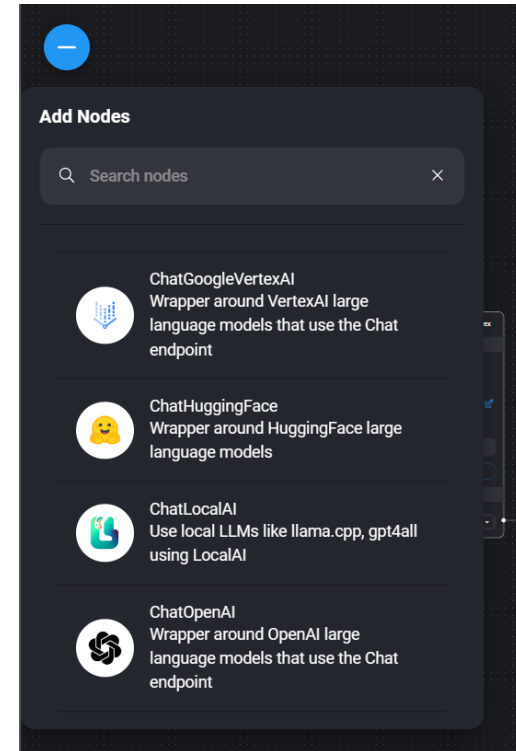
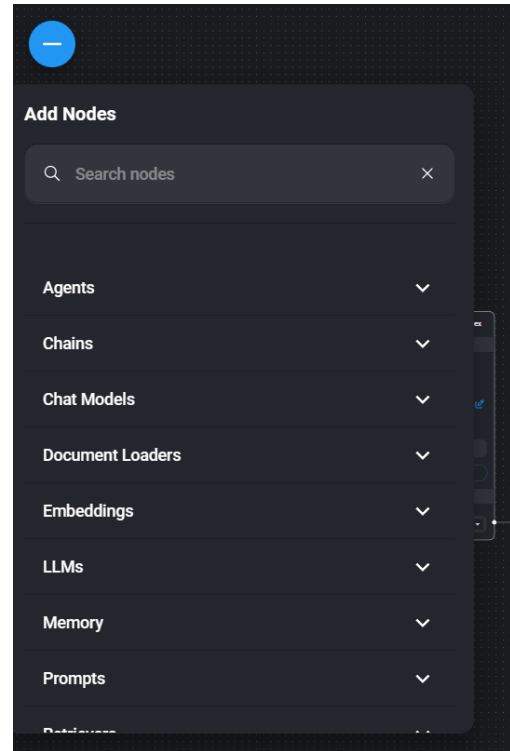
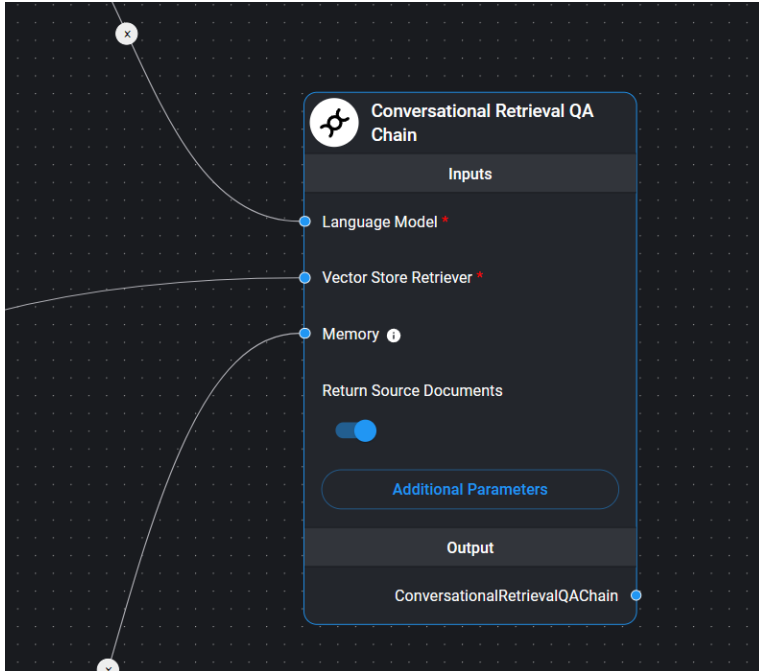
sentence_embeddings = SentenceTransformer("sentence-transformers/multi-qa-mpnet-base-dot-v1")

embeddings_sen = sentence_embeddings

[ ] import pinecone
from langchain.vectorstores import Pinecone
# initialize pinecone
pinecone.init(
    api_key=[REDACTED] # find at app.pinecone.io
    environment="us-central1-gcp" # next to api key in console
)
index_name = "flowise" # put in the name of your pinecone index here

[ ] # First, check if our index already exists. If it doesn't, we create it
if index_name not in pinecone.list_indexes():
    # we create a new index
    pinecone.create_index(
        name=index_name,
        metric="cosine",
```





General Framework:

Open source UI visual tool to build
customized LLM flow using Langchain:

Hosting of the chatbot



LangChain

FlowiseAI

render

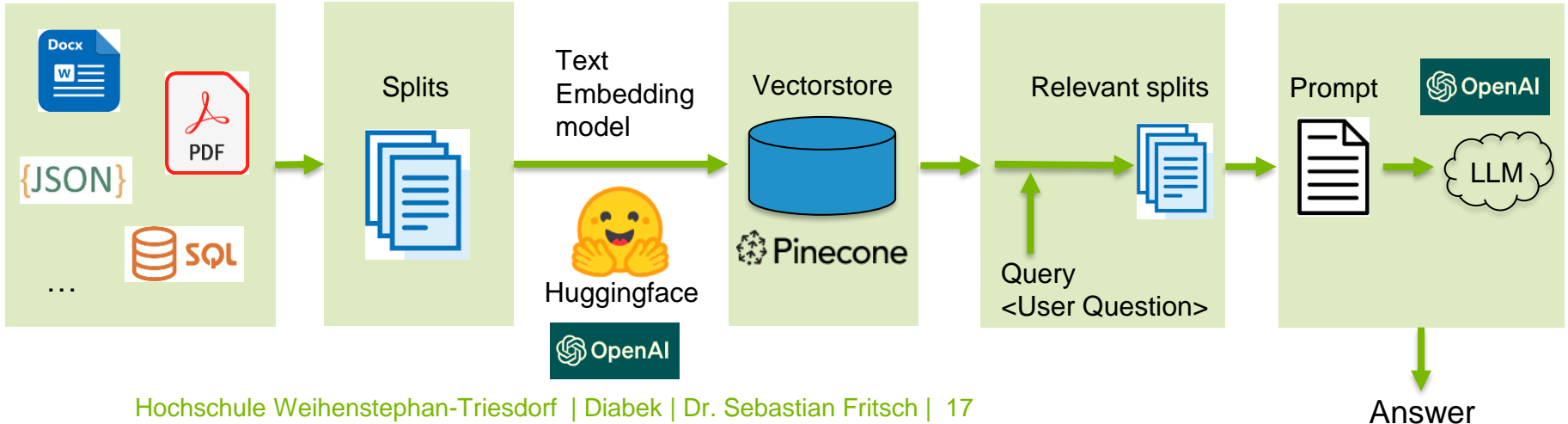
Document and
data loaders

Text Splitting

Data storage

Retrieval

Output



(Very quick) Demo

LLM Evaluation

- Manual evaluation and debugging

Ackerfuchsschwanz ist ein Unkraut, das in Getreidefeldern vorkommt. Es kann zu Ertragsverlust

zu 14 Tag...

197 8 • ...

lauf am 6...

Stadium ...

man von ...

415 11 B...

416 Kapit...

von 15-2...

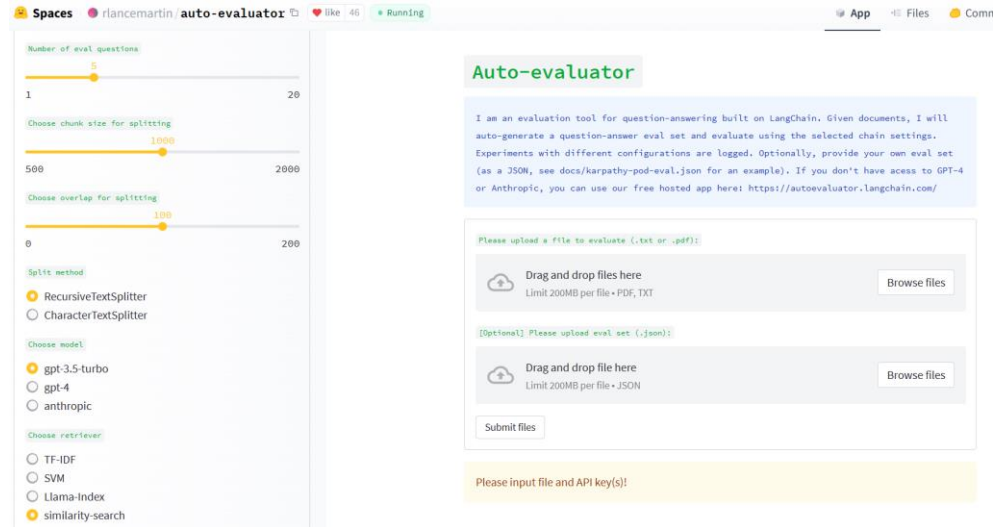
LLM Evaluation

- Manual evaluation and debugging

```
"pageContent": "415\n11\nBehandlung\nPr\u00e4parate/Wirkung in  
\nAckerfuchsschwanz\nSH-Westk\u00fcste\nPeldon (England)\nSensitiver  
Standard\nSH-Ostk\u00fcste\nDeutsches Weidelgr. \n Taube Trespe\n Roggentrespe\nUnna  
(NRW)\n65\n40\n10\n10\n10\n40\n30\n35\n0\n100\n60\n40\n20\n80\n0\n100\n60\n40\n20\n80\n100\n100\n100\n98\n95\n95\n97\n98\n85\n70\n70\n95\n65\n0\n50\n0\n100\n60\n40\n20\n80\nVorsaateneinbarung  
n 01.10.14\nVorauf\nlauf\nn 06.10.14\n EC 25\nn 04.12.14\nDevrinol  
Fl. 2,75 l\n Kerb Flo 1,875 l\n Cadou SC 0,5 l\n  
(Getreideherbizid)\nVersuch zur Ackerfuchsschwanzbek\u00e4mpfung im  
Jahr 2014/15\nSaar von verschiedenen Ackerfuchsschwanzherk\u00fcnften  
am 1. Oktober 2014; Behandlung und \nEinarbeitung kurz vor der  
Saar mit 2,75 l/ha Devrinol Fl\u00fcssig (450 g/l Napropamid). Neben  
der \nWirkung von Devrinol Fl\u00fcssig ist zum Vergleich auch die  
Wirkung von 0,5 l / ha Cadou SC (250 g/l \nFlufenacet) und 1,875  
l/ha Kerb Flo (400 g/l Propyzamid) aufgef\u00fchrt. Cadou SC wurde im  
Vorauf\nlauf am 6. Oktober 2014 eingesetzt. Die Anwendung von  
Kerb Flo erfolgte am 4. Dezember 2014"  
  
"metadata":  
  "blobType": ""  
  "loc.lines.from": 1  
  "loc.lines.to": 71  
  "loc.pageNumber": 417  
  "pdf.info.CreationDate": "D:20221220140309+01'00"  
  "pdf.info.Creator": "Adobe InDesign 18.0 (Macintosh)"  
  "pdf.info.IsAcroFormPresent": false  
  "pdf.info.IsXFAPresent": false  
  "pdf.info.ModDate": "D:20221221110821+01'00"  
  "pdf.info.PDFFormatVersion": "1.6"  
  "pdf.info.Producer": "Adobe PDF Library 17.0"  
  "pdf.info.Trapped.name": "False"  
  "pdf.metadata._metadata.dc:format": "application/pdf"  
  "pdf.metadata._metadata.pdf:producer": "Adobe PDF Library  
17.0"  
  "pdf.metadata._metadata.pdf:trapped": "False"  
  "pdf.metadata._metadata.xmp:createdate": "2022-12-
```

LLM Evaluation

- Manual evaluation and debugging
- LLM assisted evaluation



The screenshot shows the 'Auto-evaluator' application interface. On the left, there are several configuration sliders and radio buttons for settings like 'Number of eval questions', 'Choose chunk size for splitting', 'Choose overlap for splitting', 'Split method', 'Choose model', and 'Choose retriever'. The right side of the interface features a text box with an introduction to the tool, two file upload sections (one for documents and one for an optional JSON eval set), and a 'Submit files' button. A yellow banner at the bottom prompts the user to 'Please input file and API key(s)!'. The application is running in a browser window titled 'Spaces - rlanccmartin / auto-evaluator'.

LLM Evaluation

	question	answer	result
0	What is the main focus of the text?	The text focuses on the recommendations for growing maize and the factors that affect soil health and management in agriculture.	The main focus of the text is on soil health and management in agriculture.
1	What is the name of the herbicide with the active ingredient Florasulam 2.5?	Agni	The name of the herbicide with the active ingredient Florasulam 2.5 is Flu.
2	What are some methods mentioned in the text for controlling weeds and grasses?	Some methods mentioned in the text for controlling weeds and grasses include crop rotation, mechanical weeding, and the use of herbicides.	Some methods mentioned in the text for controlling weeds and grasses include crop rotation, mechanical weeding, and the use of herbicides.
3	What are some of the host plants of Rhizoctonia violacea?	Some of the host plants of Rhizoctonia violacea include beets, rapeseed, clover, potato, and maize.	Some of the host plants of Rhizoctonia violacea include sugar beets, rapeseed, clover, and maize.
4	What are the recommended storage conditions for the product?	Storage temperature should be below 10 °C and CO2 concentration should be less than 1000 ppm.	The recommended storage conditions for the product are as follows:
5	What is the purpose of using multiple active substances against the marked weeds?	The purpose is to control the weeds that appear in waves, requiring multiple applications.	The purpose of using multiple active substances against the marked weeds is to control the weeds that appear in waves, requiring multiple applications.
6	What are the principles of organic farming?	The principles of organic farming include a closed farm system, preservation of soil fertility, and the use of natural pest control.	The principles of organic farming include: a closed farm system, preservation of soil fertility, and the use of natural pest control.
7	What is the expiration date for the pesticide Cyperkill Max?	The expiration date for Cyperkill Max is February 28, 2023.	The expiration date for the pesticide Cyperkill Max is February 28, 2023.
8	What is the main component of Rübenmark?	Zellulose	The main component of Rübenmark is cellulose.
9	What were the yields of untreated control plots in the experiments?	The yields of untreated control plots in the experiments were 62 dt/ha.	The yields of untreated control plots in the experiments were 89.6 dt/ha in 2021 and 89.6 dt/ha in 2022.

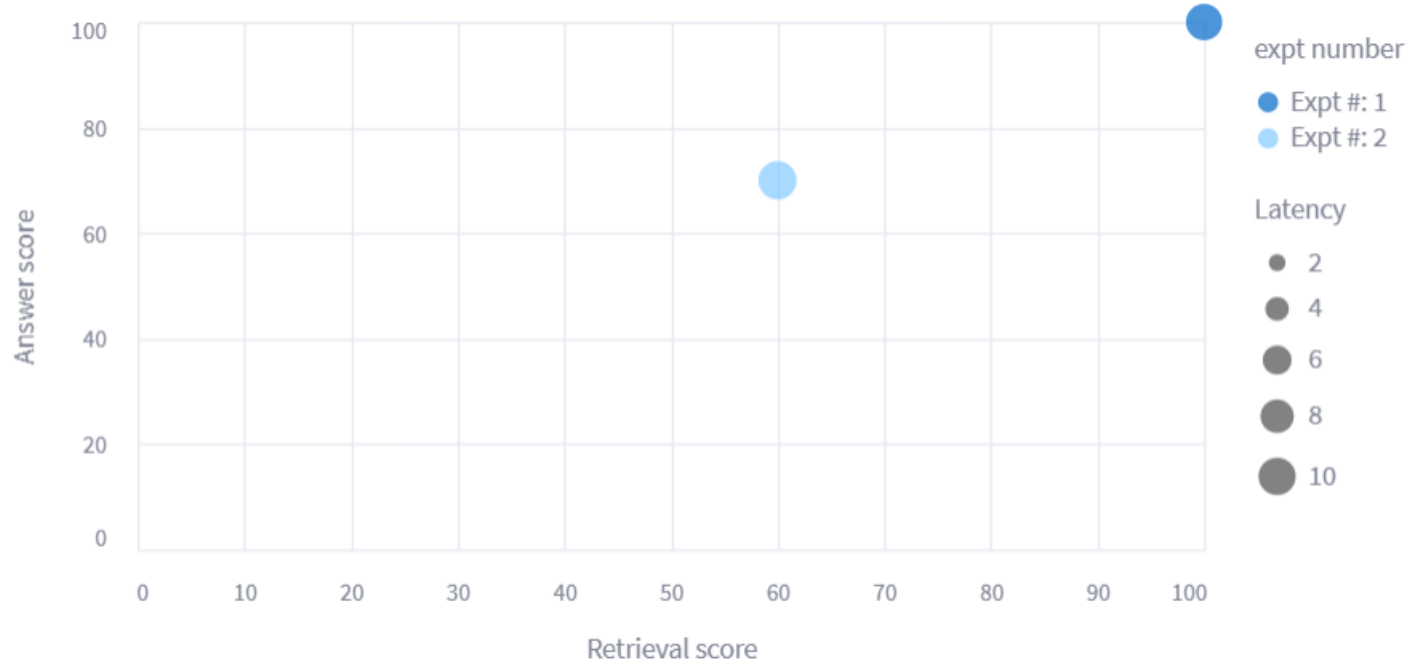
LLM Evaluation

answer score	docs score	latency
relevance: The student answer does not directly quote the text, but it does mention several topics that are discussed in the text, such as soil health, nutrient supply, pH levels, soil structure, and soil organisms. However, it does not mention the specific focus on recommendations for growing maize and the factors that affect its growth and yield. Therefore, the submission does not meet the relevance criterion.	Context is relevant: False.	5.084
conciseness: The student answer is concise and to the point. It clearly states the main focus of the text in a single sentence. Therefore, the submission meets the conciseness criterion.	Context is relevant: True	3.7176
correct: The student answer does not accurately reflect the main focus of the text. It mentions soil health and management in agriculture, which is a broader topic than the specific focus on recommendations for growing maize and the factors that affect its growth and yield. Therefore, the submission does not meet the correct criterion.	Context is relevant: False.	20.9474
CORRECT	Context is relevant: True.	7.0067
relevance: The student answer does not directly quote the true answer, but it does address the purpose of using multiple active substances against the marked weeds, which is mentioned in the text.	Context is relevant: False.	15.2548
relevance: The student answer refers to the principles of organic farming mentioned in the text, including preservation and enhancement of soil fertility, use of natural inputs, and avoidance of synthetic pesticides. Therefore, the submission meets the relevance criterion.	Context is relevant: True.	10.5981
Relevance: The submission is referring to the pesticide Cyperkill Max, which is mentioned in the question. Therefore, it is relevant to the question.	Context is relevant: True.	32.2561
To assess the relevance criterion, we need to determine if the student's answer is referring to a real quote from the text. However, since no text is provided, we cannot determine the relevance of the submission.	Context is relevant: True.	2.8393
Relevance: The student answer does not refer to a real quote from the text. The true answer states that the yields of untreated control plots were 62 dt/ha, while the student's answer states that the yields were 62 dt/ha. Therefore, the submission does not meet the relevance criterion.	Context is relevant: True.	1.8922
	Context is relevant: False.	5.3619

LLM Evaluation

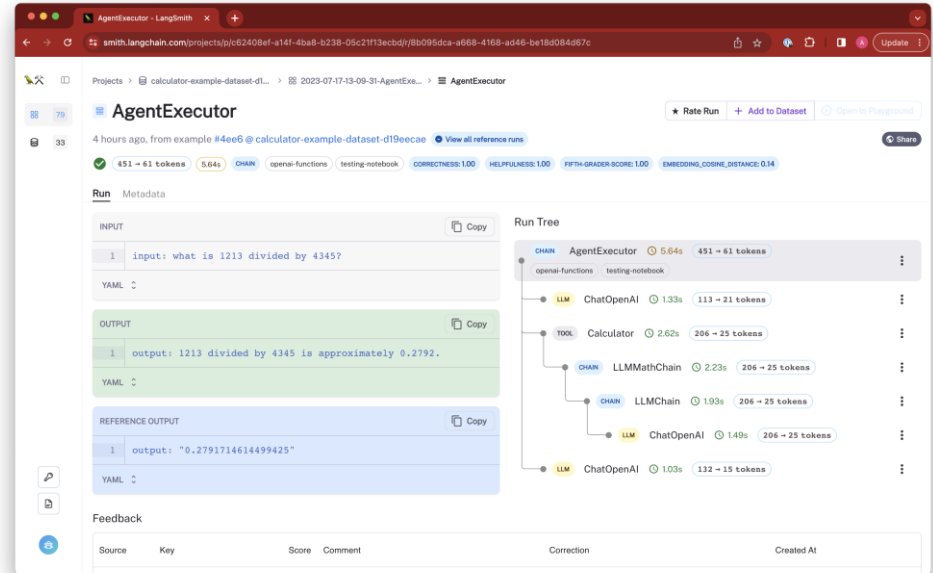
	chunk_chars	overlap	split	model	retriever	embedding	num_neighbors	Latency	Retrieval score	Answer score
0	1,000	100	RecursiveTextSplitter	gpt-3.5-turbo	similarity-search	OpenAI	3	9.4948	100	100
1	1,000	150	RecursiveTextSplitter	gpt-3.5-turbo	similarity-search	OpenAI	6	10.4958	60	70

LLM Evaluation



LLM Evaluation

- Manual evaluation and debugging
- LLM assisted evaluation
- LangChain evaluation platform



The screenshot displays the LangChain AgentExecutor evaluation interface. The main content area shows the following details:

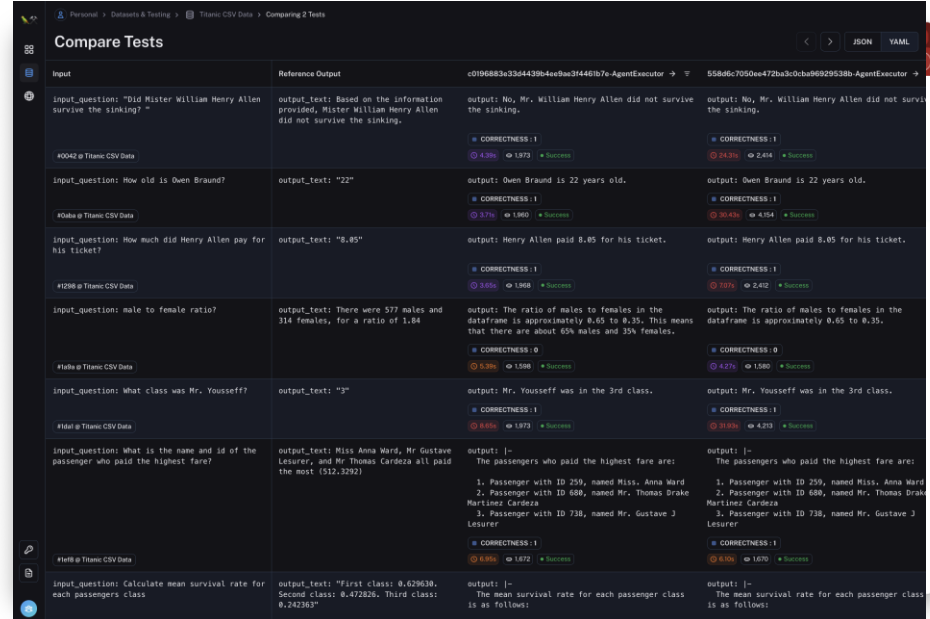
- Run Metadata:** 4 hours ago, from example #4ee6 @ calculator-example-dataset-d19eecae. Metrics include CORRECTNESS: 1.00, HELPFULNESS: 1.00, FIFTH GRADER SCORE: 1.00, and EMBEDDING COSINE DISTANCE: 0.14.
- Run:** Metadata section with input, output, and reference output.
- Run Tree:** A flowchart showing the execution path: CHAIN AgentExecutor (5.64s, 451 tokens) → CHAIN LLMChain (2.23s, 206 tokens) → LLM ChatOpenAI (1.03s, 132 tokens).
- Feedback Table:**

Source	Key	Score	Comment	Correction	Created At
--------	-----	-------	---------	------------	------------

Source: <https://docs.smith.langchain.com/tracing>

LLM Evaluation

- Manual evaluation and debugging
- LLM assisted evaluation
- LangChain evaluation platform



The screenshot displays the 'Compare Tests' interface in the LangChain evaluation platform. It shows a table with columns for 'Input', 'Reference Output', and two columns of LLM outputs. Each output is accompanied by a 'CORRECTNESS' score and a 'Success' indicator. The interface also includes navigation buttons for 'JSON' and 'YAML' at the top right.

Input	Reference Output	LLM 1 Output	LLM 2 Output
Input_question: "Did Mister William Henry Allen survive the sinking?"	output_text: Based on the information provided, Mister William Henry Allen did not survive the sinking.	output: No, Mr. William Henry Allen did not survive the sinking.	output: No, Mr. William Henry Allen did not survive the sinking.
Input_question: How old is Owen Braund?	output_text: "22"	output: Owen Braund is 22 years old.	output: Owen Braund is 22 years old.
Input_question: How much did Henry Allen pay for his ticket?	output_text: "8.85"	output: Henry Allen paid 8.85 for his ticket.	output: Henry Allen paid 8.05 for his ticket.
Input_question: male to female ratio?	output_text: There were 577 males and 314 females, for a ratio of 1.84	output: The ratio of males to females in the dataframe is approximately 0.55 to 0.35. This means that there are about 65% males and 35% females.	output: The ratio of males to females in the dataframe is approximately 0.65 to 0.35.
Input_question: what class was Mr. Yousseff?	output_text: "3"	output: Mr. Yousseff was in the 3rd class.	output: Mr. Yousseff was in the 3rd class.
Input_question: what is the name and id of the passenger who paid the highest fare?	output_text: Miss Anna Ward, Mr Gustave Lesieur, and Mr. Thomas Carcra all paid the most (512.3292)	output: - The passengers who paid the highest fare are: 1. Passenger with ID 259, named Miss, Anna Ward 2. Passenger with ID 689, named Mr, Thomas Drake Martinez Carcra 3. Passenger with ID 738, named Mr, Gustave J Lesieur	output: - The passengers who paid the highest fare are: 1. Passenger with ID 259, named Miss, Anna Ward 2. Passenger with ID 689, named Mr, Thomas Drake Martinez Carcra 3. Passenger with ID 738, named Mr, Gustave J Lesieur
Input_question: Calculate mean survival rate for each passengers class	output_text: "First class: 0.629630. Second class: 0.472826. Third class: 0.242393"	output: - The mean survival rate for each passenger class is as follows:	output: - The mean survival rate for each passenger class is as follows:

Source: <https://blog.langchain.dev/test-run-comparisons/>

Next steps

Short-term:

- Prompt engineering
- Test and evaluation with real-life users
- Connecting the BVL API
- Integrating technical data

A bit further away:

- Use for fertilizer application
- Embed in messenger
- Use of localized weather information



Thank you very much.

sebastian.fritsch@hswt.de

*Applied Sciences
for Life*

Gefördert durch:



Bundesministerium
für Ernährung
und Landwirtschaft

ptble
Projekträger Bundesanstalt
für Landwirtschaft und Ernährung

aufgrund eines Beschlusses
des Deutschen Bundestages