EnBW LernGPT

Millions of employees’ prompts served proudly with MongoDB

Timm Friebe, EnBW

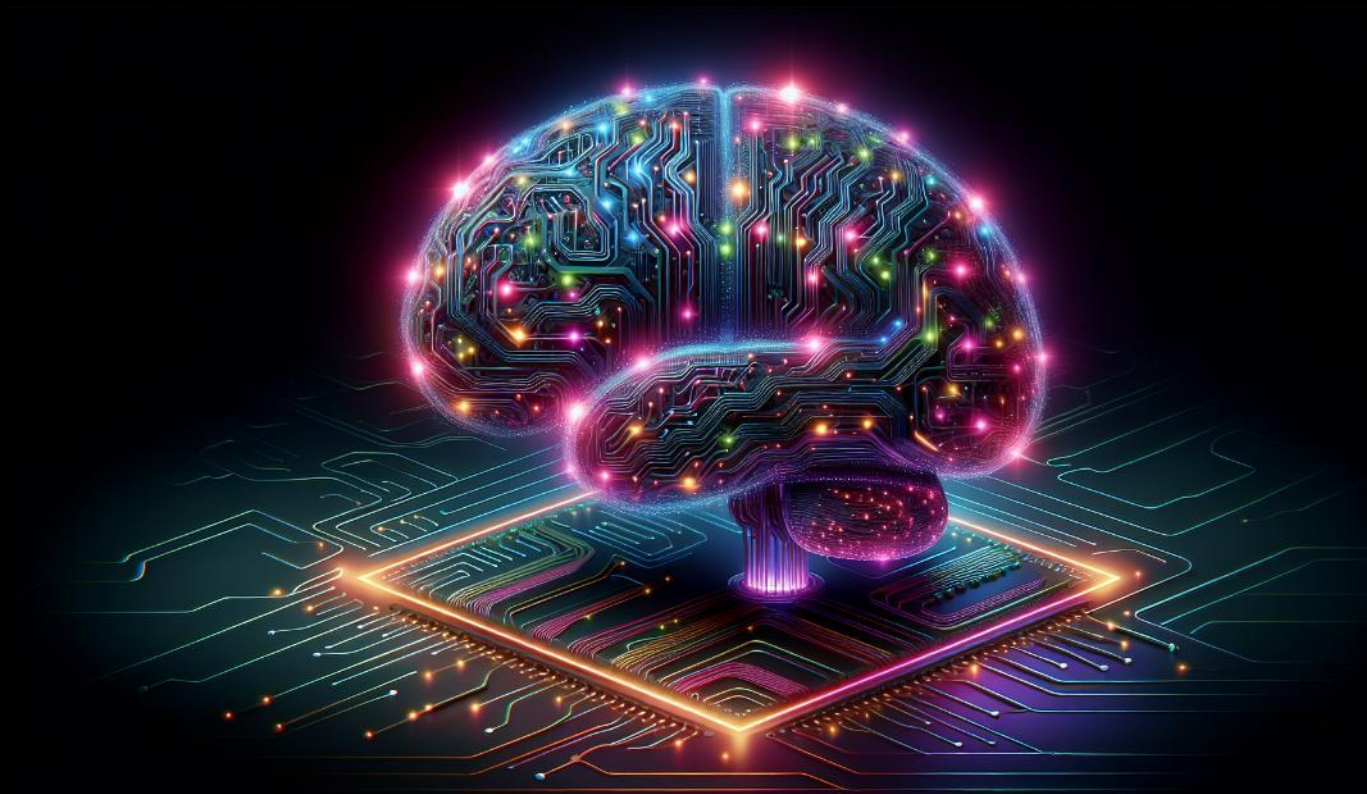
MongoDB Day, Frankfurt am Main, 30.09.2025

Timm Friebe

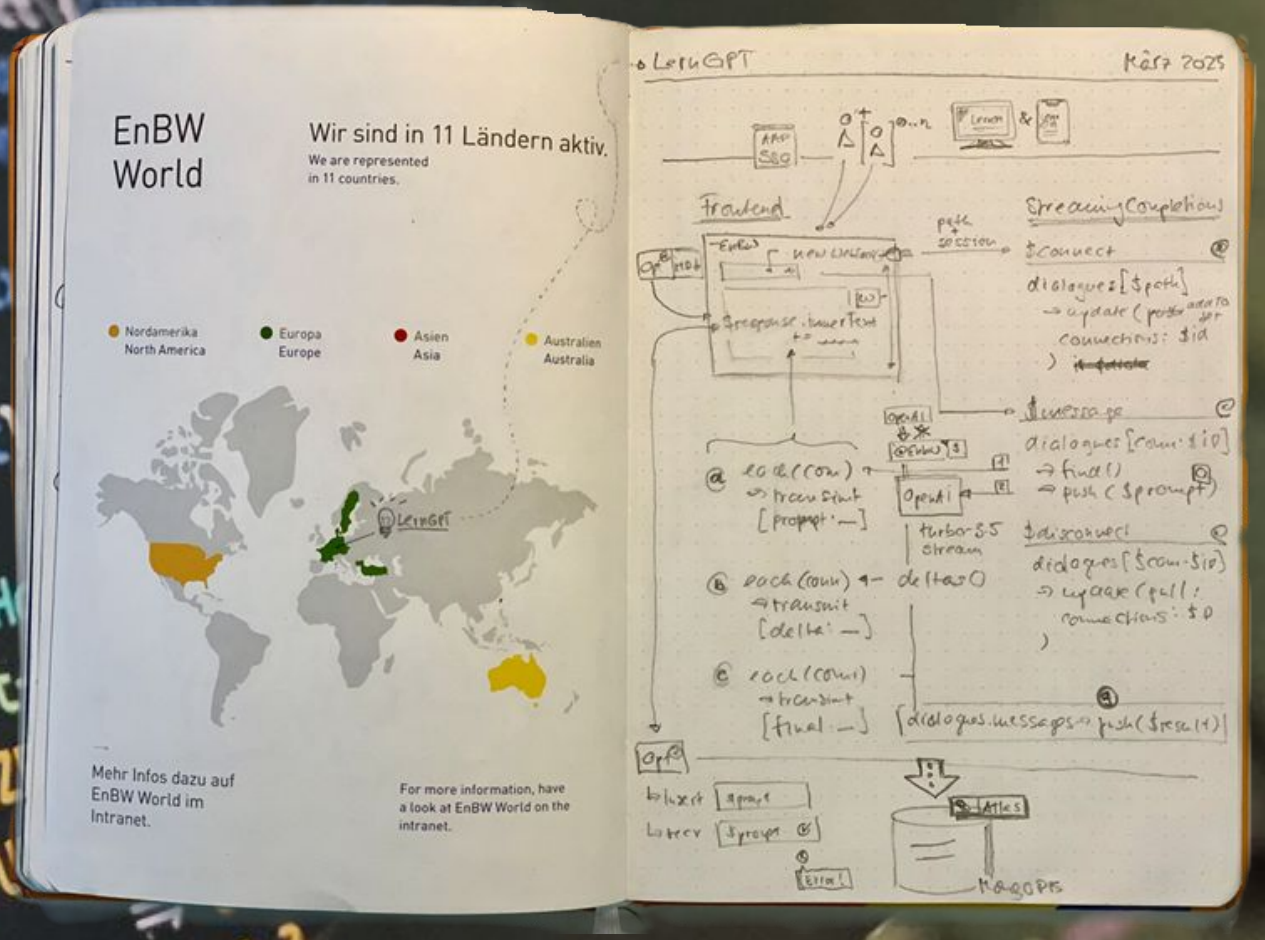
1999 – 2019 @ IT | 2019 - … @ HR

November 2022

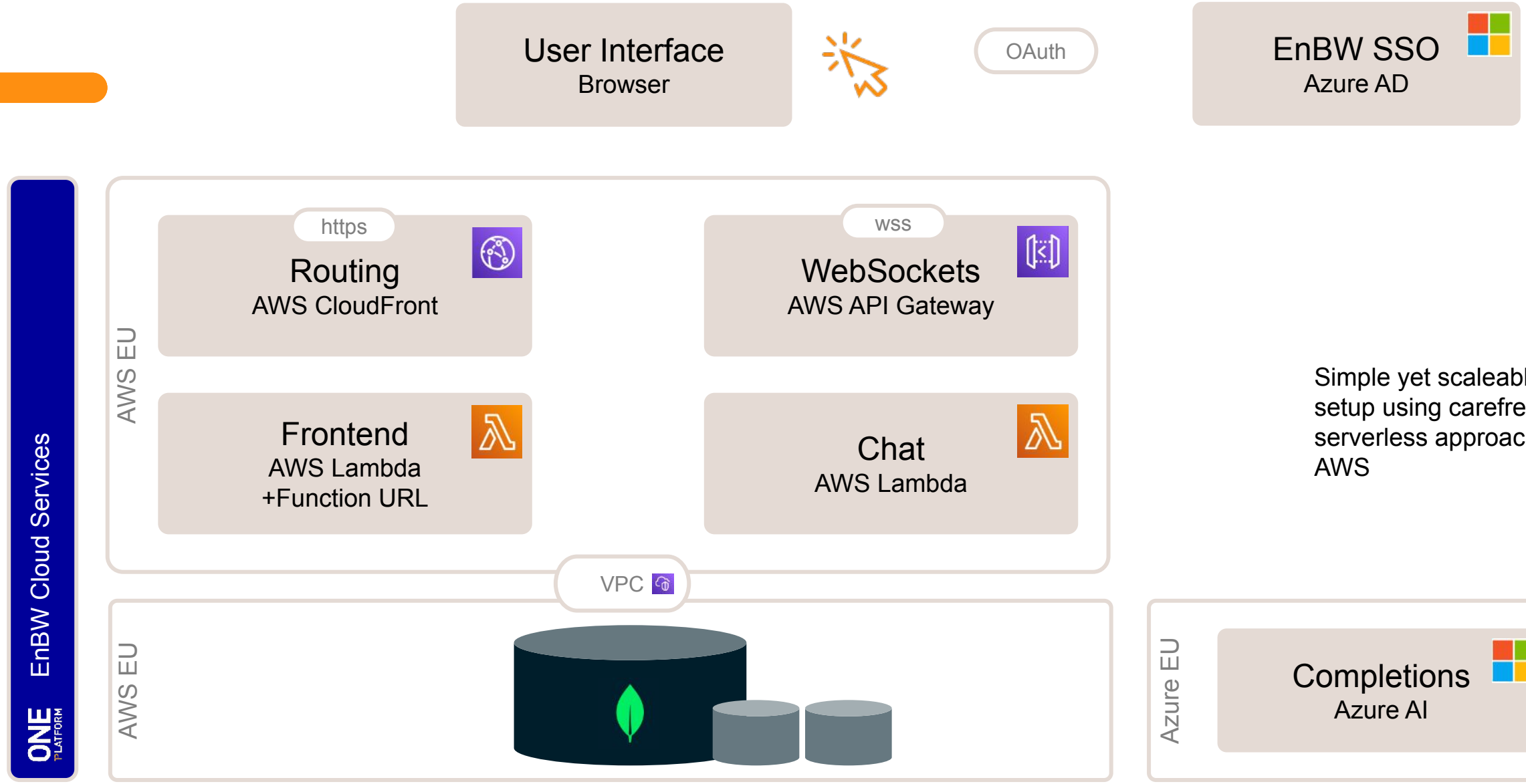
Hello World!



February 2023

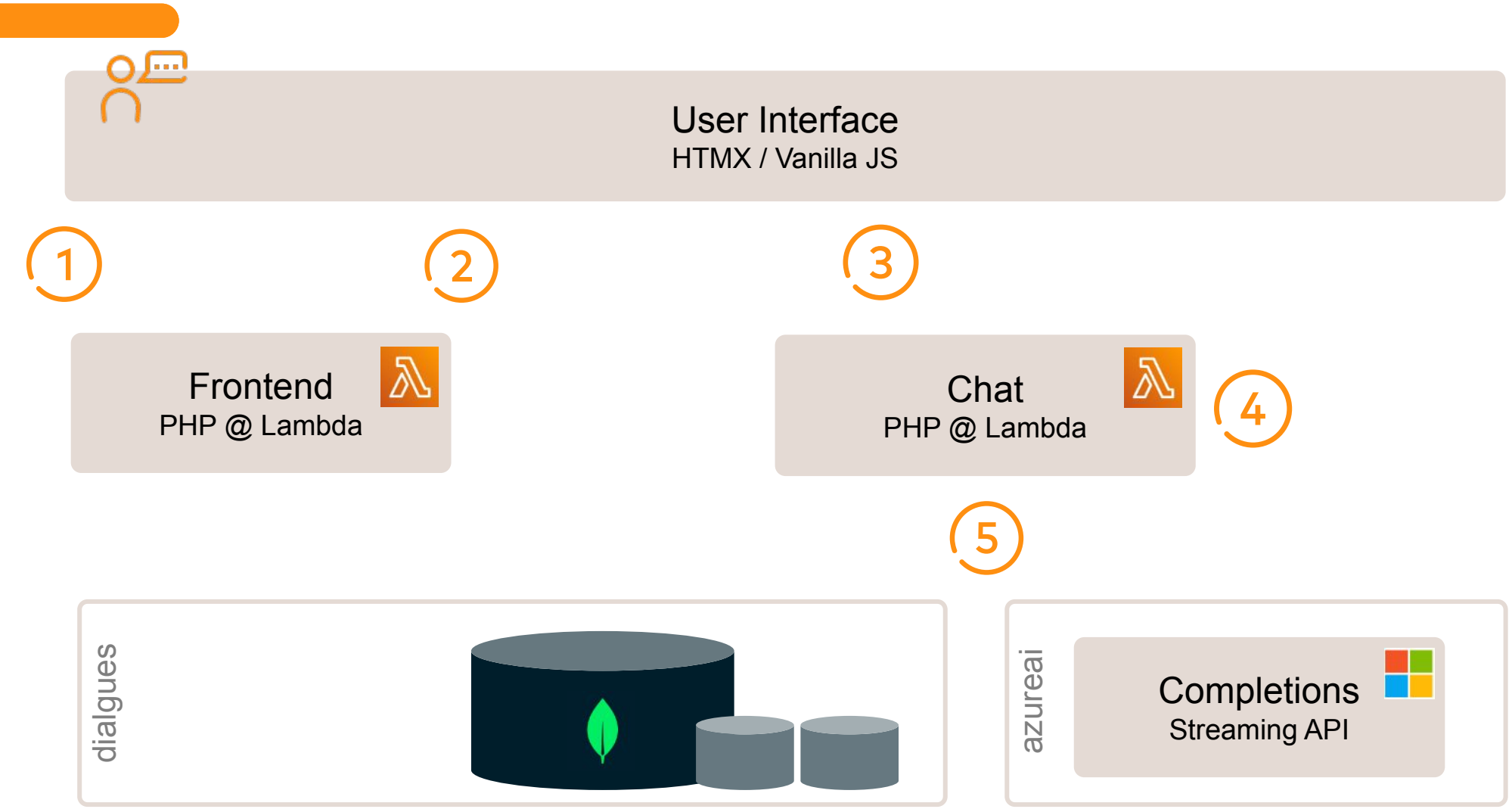


Architecture



Simple yet scaleable setup using carefree serverless approach in AWS

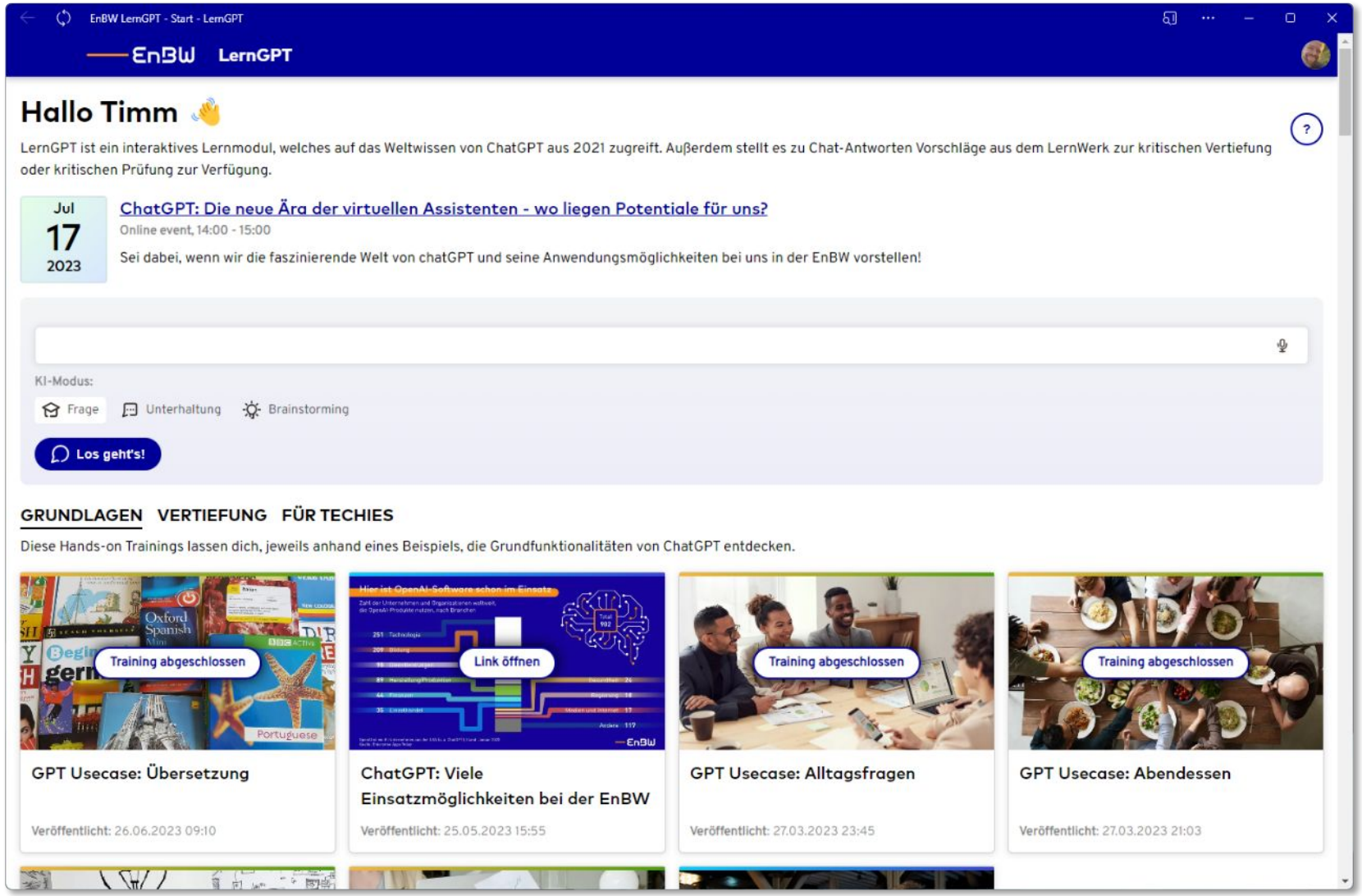
Anatomy of a chat



1. User agent sends a POST request to the frontend. The frontend creates a *dialogue* and redirects to /chat/{id}
2. Frontend renders an empty chat page including the WebSocket connection
3. User agent connects to chat lambda
4. Chat lambda invokes completions API and starts streaming completions
5. Chat lambda stores result in *dialogues* collection

Welcome to LernGPT

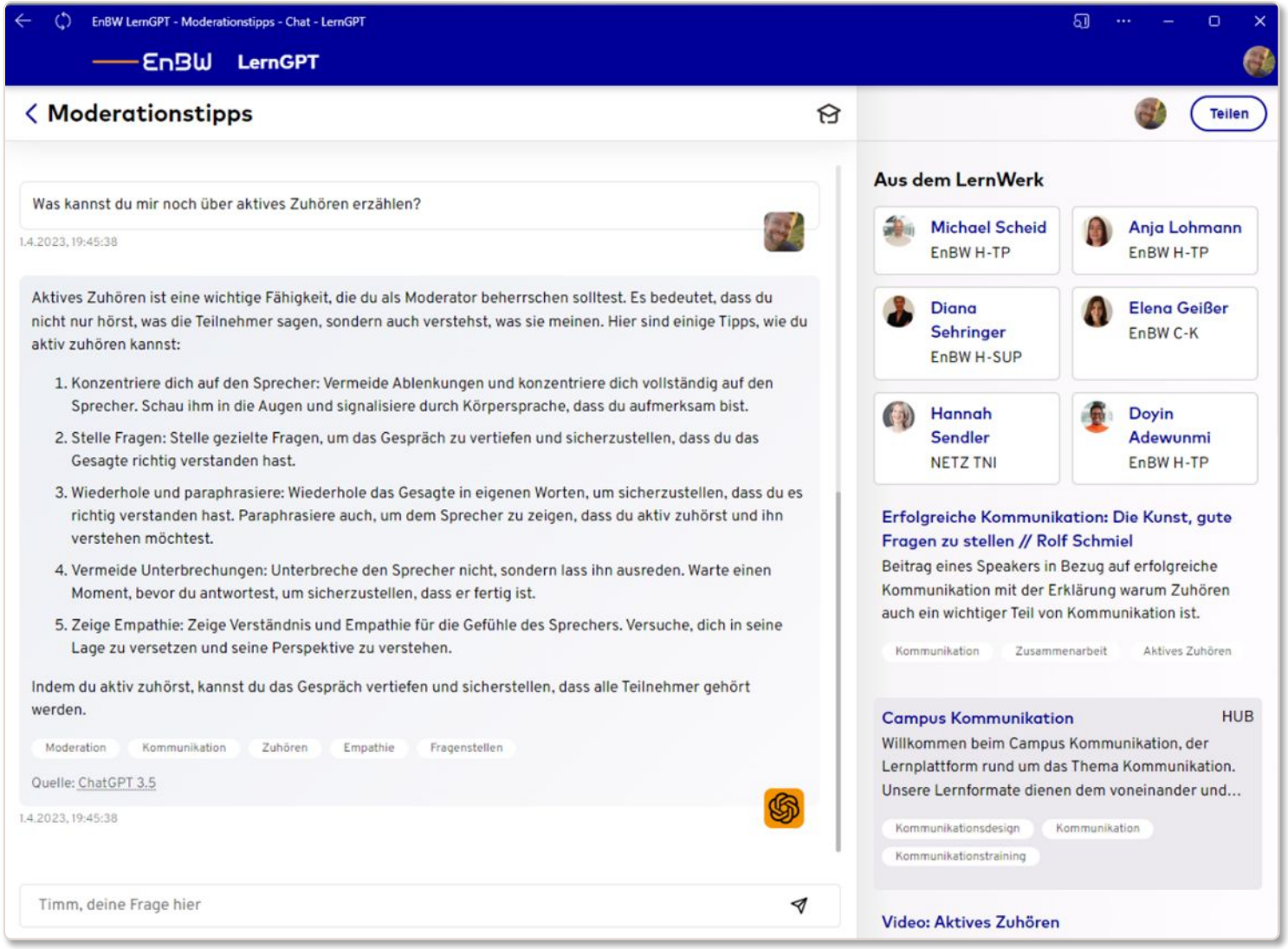
First official release after being approved by all necessary internal committees ;)



Contents from our LXP

EnBW LERNWERK

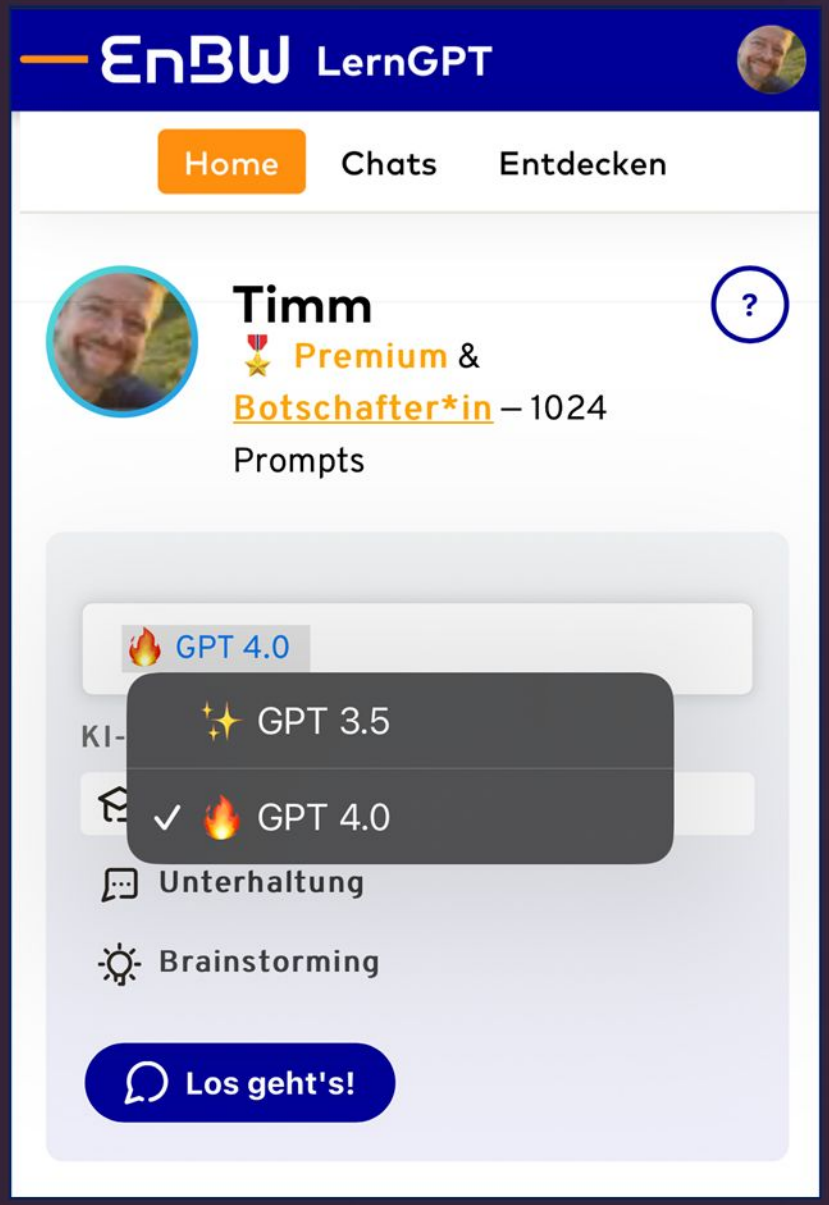
Never go alone



…and then invokes a search on colleagues with the relevant areas of expertise.

After the answer has been streamed, the LLM is asked to extract keywords…

Powered by Atlas Search



December 2023

GPT 4

Introducing GPTs

"You can now create custom versions of ChatGPT that combine instructions, extra knowledge, and any combination of skills".

OpenAI announcement
November 2023

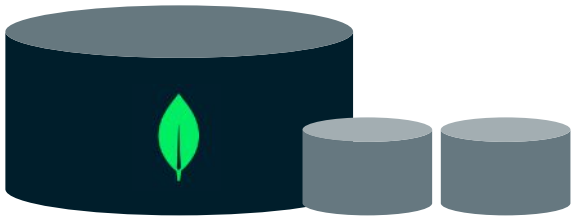
Including the system prompt

When starting a chat, we additionally query the *assistants* collection and read the instructions

User Interface
Browser

Chat
PHP @ Lambda

dialgues
assistants



azureai

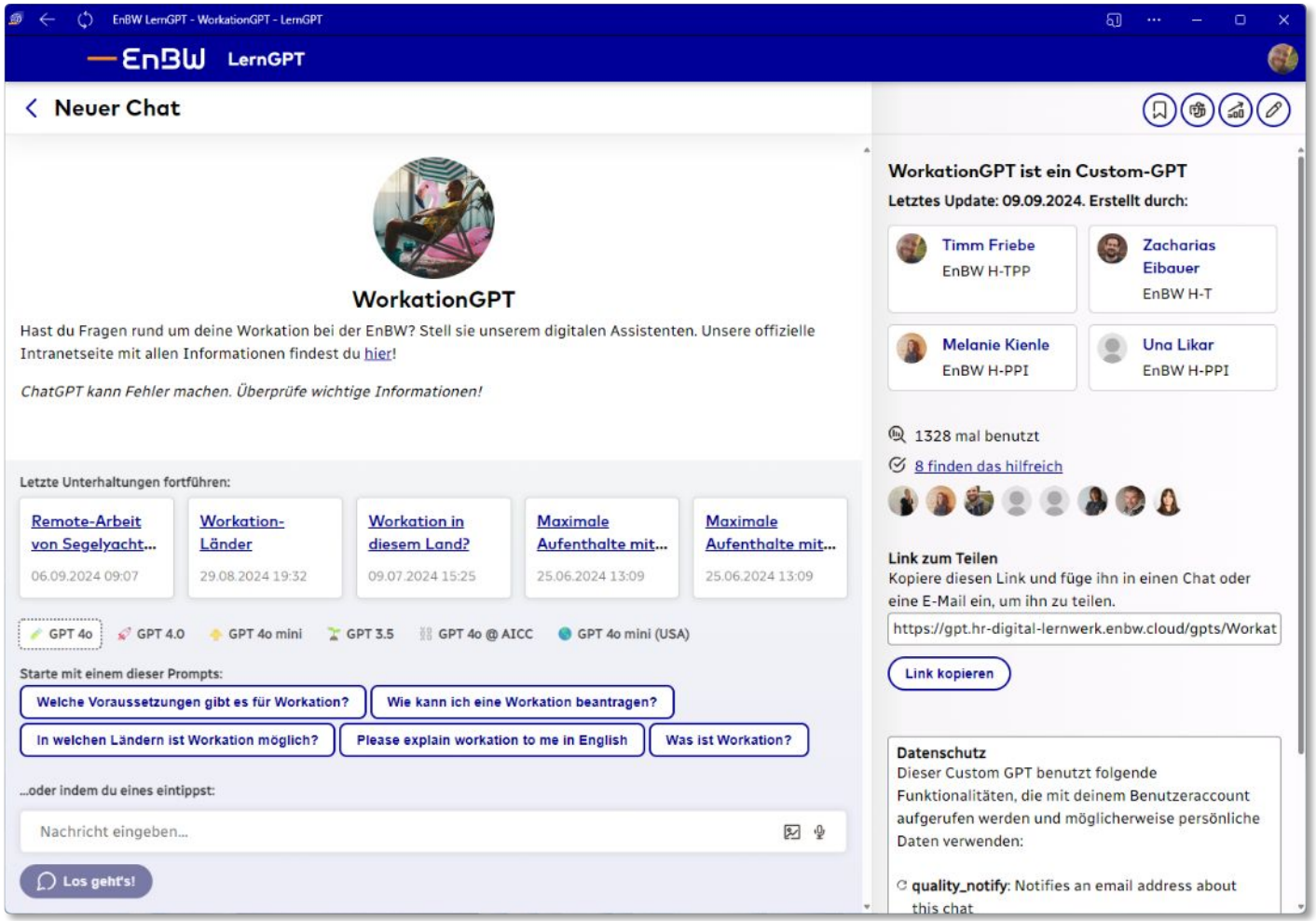
Completions
Streaming API

```
$ curl https://api.openai.com/v1/completions \
  -H "Content-Type: application/json" \
  -H "Authorization: Bearer $OPENAI_API_KEY" \
  -d '{
    "model": "gpt-3.5-turbo-instruct",
    "messages": [
      {"role": "system", "content": "You are …"},
      {"role": "user", "content": "$PROMPT"}
    ],
    "stream": true
  }'
```

The system prompt is simply prepended to the messages array

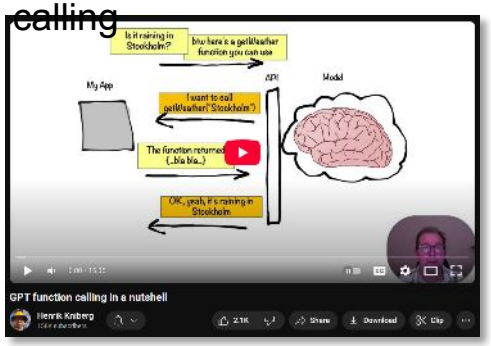
Custom GPTs – no-code „AI apps“

CustomGPTs started out as the possibility to edit the system prompt and a couple of smaller UI settings.



Anyone can create a CustomGPT in a matter of seconds – no coding necessary.

GPTs are created by a GPT – and function calling

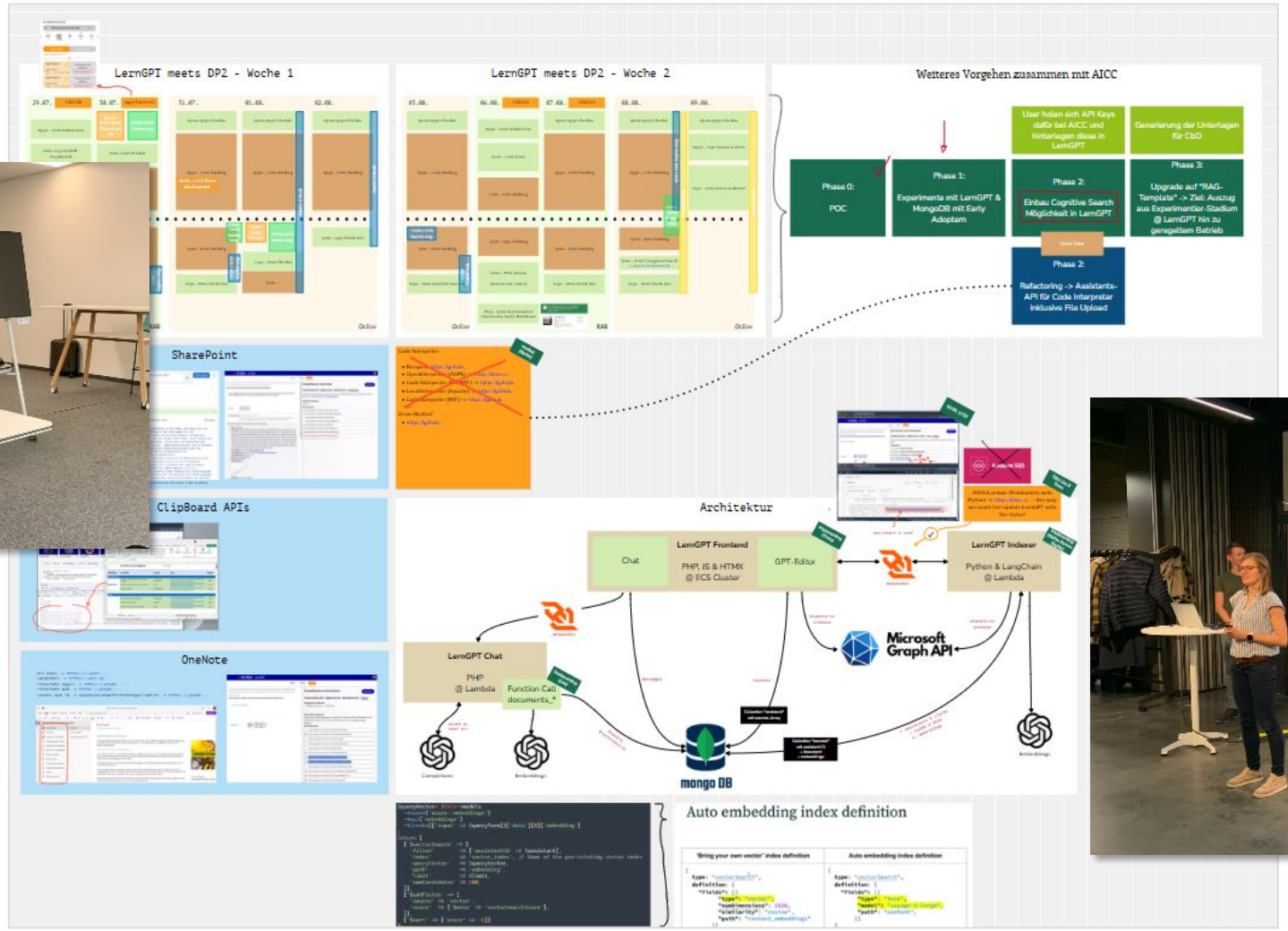


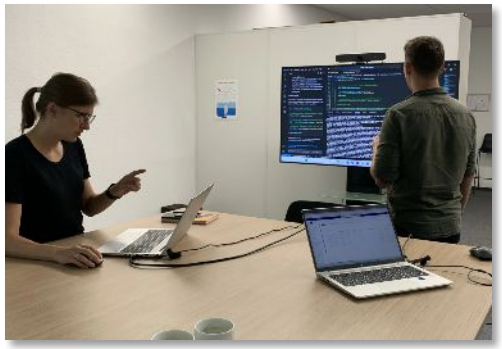
August 2024

LernGPT meets DP²

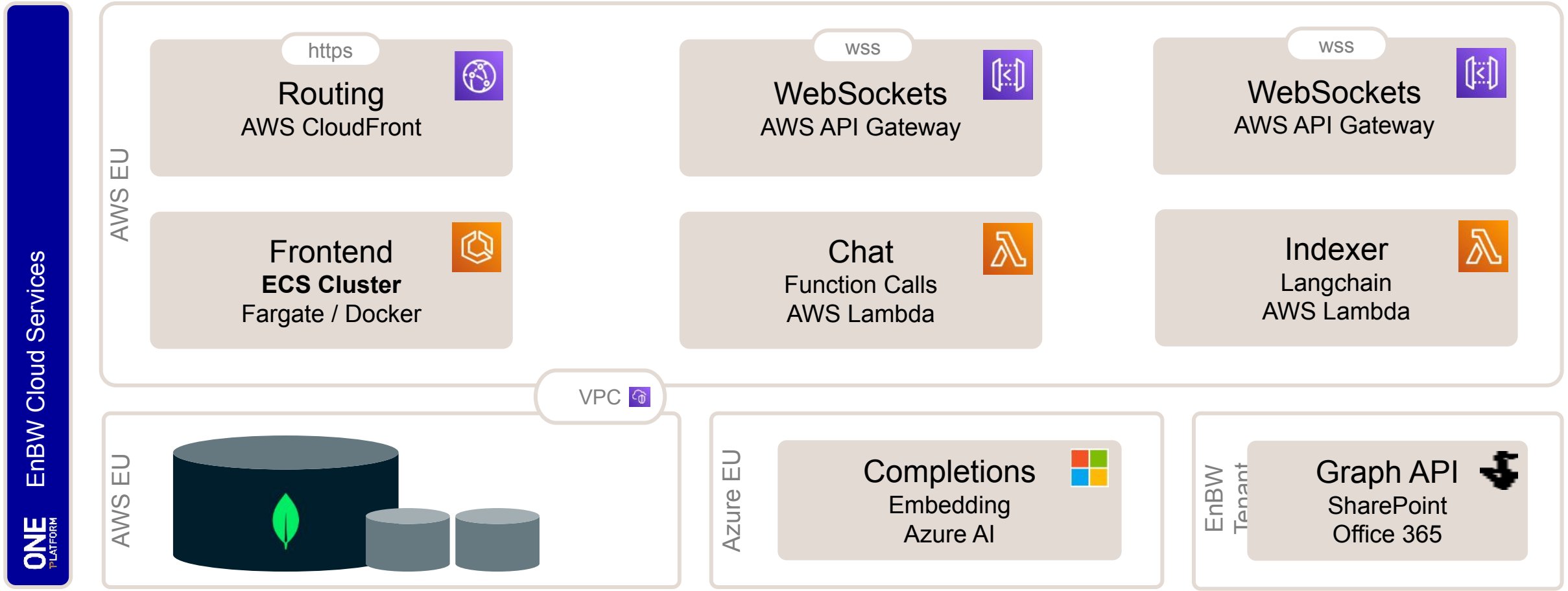
2 Weeks in August

…using $vectorSearch





Architecture



Our simple RAG system

EnBW

4 The function is called

…creates embeddings for the documents

2

…creates embeddings for the search term(s)

5

Indexer
Langchain
AWS Lambda

Completions
Embedding
Azure AI

Chat
Function Calls
AWS Lambda

The indexer fetches the selected documents by using the Graph API

1

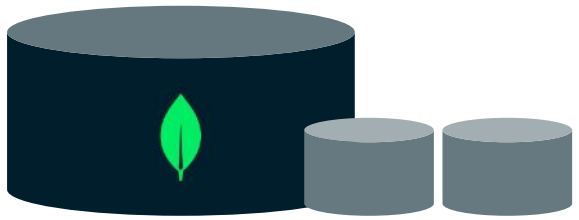
…and stores them alongside with the documents

3

6

…and issues a search query, returning the documents to the text function call. We use $search, $vectorSearch and a crazy pipeline for hybrid search:

Graph API
SharePoint
Office 365

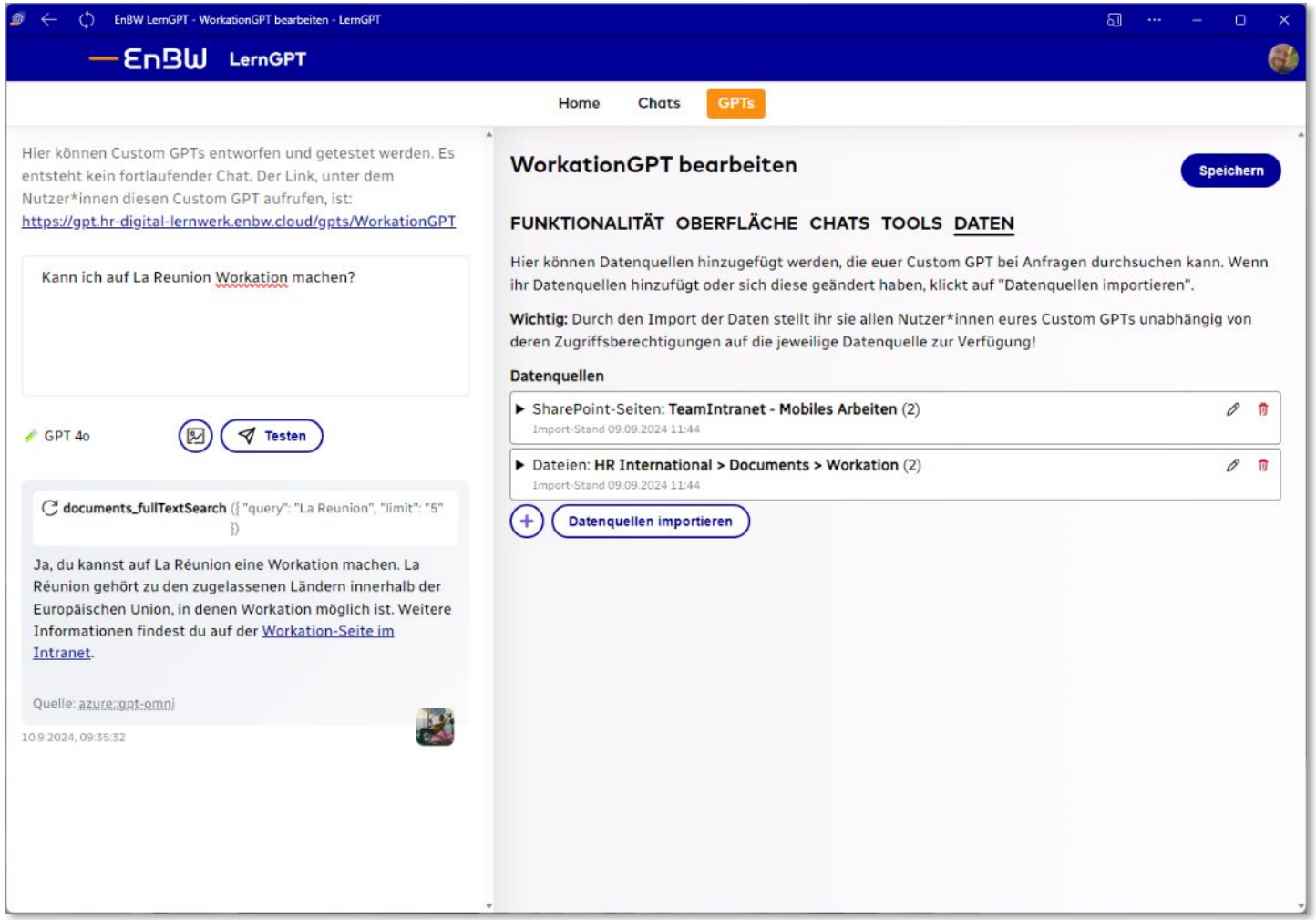


documents collection

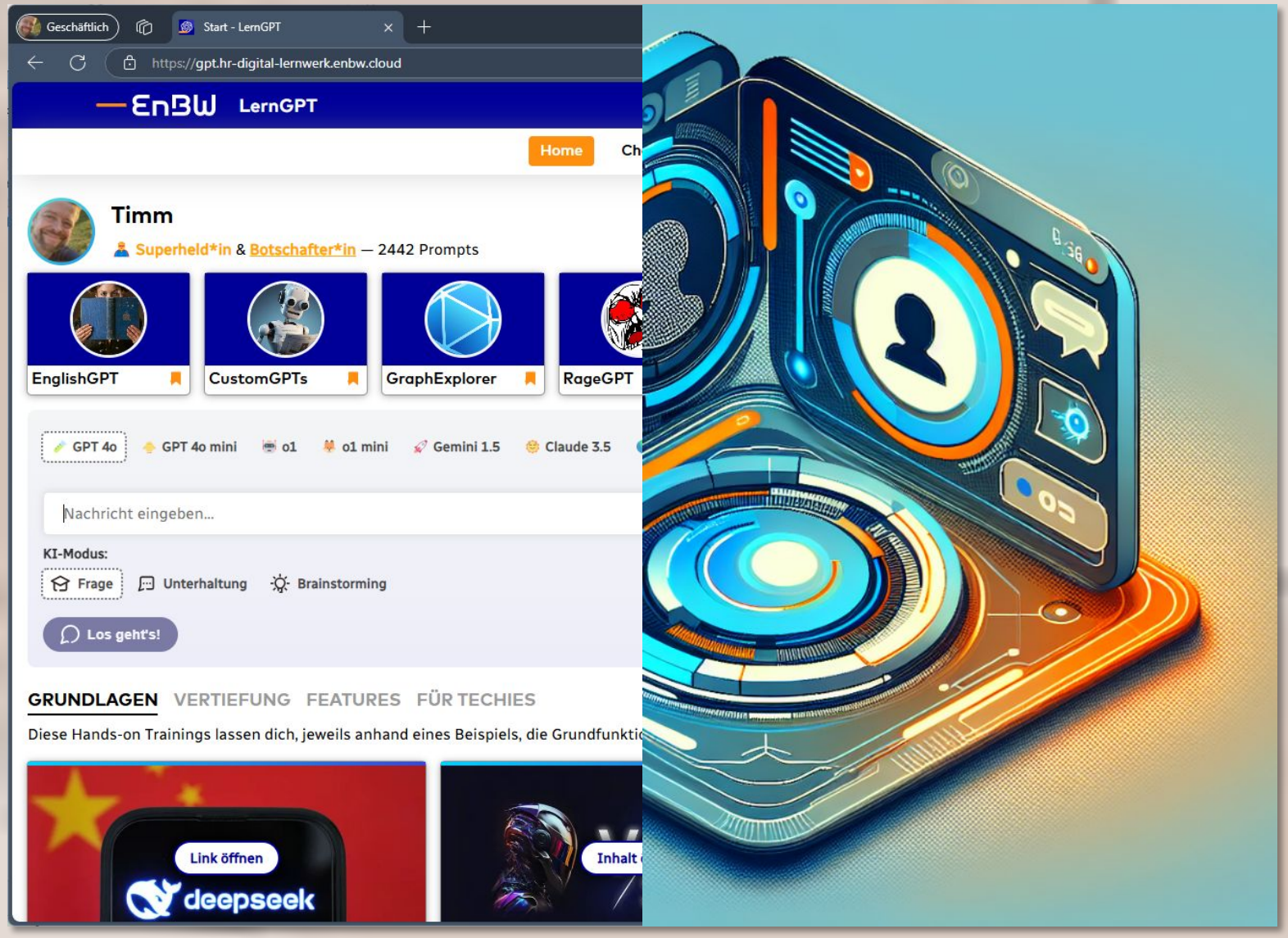
```
return $this->search([
  $vectorStage[0],
  ['$group' => ['_id' => null, 'vector' => ['$push' => '$$ROOT']]],
  ['$unwind' => ['path' => '$vector', 'includeArrayIndex' => 'rank']],
  ['$addFields' => [
    'source'    => 'vector',
    'vs_score' => ['$divide' => [1.0, ['$add' => ['$rank', $vectorPenalty, 1]]]],
  ]],
  ['$project' => ['vs_score' => 1, ...$select('$vector.')]],
  ['$unionWith' => [
    'coll' => 'documents',
    'pipeline' => [
      $fullTextStage[0],
      ['$addFields' => [
        'source' => 'fulltext',
        'fts_score' => ['$divide' => [1.0, ['$add' => ['$rank', $fullTextPenalty, 1]]]],
      ]],
    ]
  ]],
  ['$group' => ['_id' => $select('$'), 'vs_score' => ['$max' => '$vs_score'], 'fts_score' => ['$max' => '$fts_score']]],
  ['$project' => [
    ...$select('$_id.'),
    'vs_score' => ['$ifNull' => ['$vs_score', 0]],
    'fts_score' => ['$ifNull' => ['$fts_score', 0]]
  ]],
  ['$project' => [
    'score' => ['$add' => ['$fts_score', '$vs_score']],
    ...array_map(fn($value) => 1, $select('')),
  ]],
  ['$limit' => 10],
  ['$sort' => ['score' => -1]],
]);
```

Importing documents & testing search

Here, you can see the function call to *fullTextSearch* in action



Users select the data source(s) they want to import from. These are stored in the assistant's *sources* – the „import" button then controls the indexer via WebSocket.


Geschäftlich
Start - LernGPT
https://gpt.hr-digital-lernwerk.enbw.cloud
EnBW LernGPT
Home
Timm
Superheld*in & Botschafter*in — 2442 Prompts
EnglishGPT
CustomGPTs
GraphExplorer
RageGPT
GPT 4o
GPT 4o mini
o1
o1 mini
Gemini 1.5
Claude 3.5
Nachricht eingeben...
KI-Modus:
Frage
Unterhaltung
Brainstorming
Los geht's!
GRUNDLAGEN
VERTIEFUNG
FEATURES
FÜR TECHIES
Diese Hands-on Trainings lassen dich, jeweils anhand eines Beispiels, die Grundfunkti
Link öffnen
deepseek
Inhalt