

Ach, deshalb!

Tech hinter KI-Chatbots

Tilman Scheel & ASK

Rechtsanwalt, Kunsthistoriker, KI-Dozent

IT für Filmfestivals und Bibliotheken seit 2004

CEO ava-library.com und **AboutSomethinK.org**



ASK ist der weltweit erste KI-Chatbot für öffentliche Bibliotheken. Er ist seit Juni 2024 online auf der Seite des Verbundes der Berliner Bibliotheken voebb.de



Mit wem?

The screenshot shows the Zonnedorp library website. At the top, there's a navigation bar with links: ANMELDEN, MEIN KONTO, SITZUNG BEENDEN, English, DGS, Leichte Sprache, Häufige Fragen, and Kontakt. Below this is a banner for VÖBB (Verbund der Öffentlichen Bibliotheken Berlins) with the text 'WIR BRINGEN MEDIEN IN BEWEGUNG'. The main header features the 'de Bib' logo and 'testbib Zonnedorp'. A navigation menu includes Home, Nieuws, Activiteiten, Collectie, Doelgroepen, and Chatbot-test. A search bar is present with the text 'Überall suchen' and buttons for 'Suchen' and 'Erweiterte Suche'. The chatbot test interface is highlighted with a green border. It contains the text 'de Bib tipt' and 'Test tot 15 december'. Below this, it says 'Aanbeveling voor een boek, e-book, muziek, film of iets anders uit de Bib-catalogus? Stel je vraag via de knop rechtsonder. Vergeet zeker ook geen feedback te geven!'. The URL 'https://zonnedorp.bibliotheek.be/test-chatbot' is provided. On the right side of the chatbot interface, there's a 'Druckversion' button and a 'Bild' icon. The URL 'voebb.ava.watch' is also visible.

de Bib tipt

Test tot 15 december

Aanbeveling voor een boek, e-book, muziek, film of iets anders uit de Bib-catalogus? Stel je vraag via de knop rechtsonder. Vergeet zeker ook geen feedback te geven!

URL: <https://zonnedorp.bibliotheek.be/test-chatbot>

Druckversion

URL: voebb.ava.watch

datronicsoft

koha

axiell

a|S|tec|
angewandte Systemtechnik eG

ASK

Terminologie & Themen

Empfehlungs-Bot

→ Ähnlichkeitssuche, Wahrscheinlichkeit

Service-Bot

→ Präzision und Aktualität, Halluzinationen

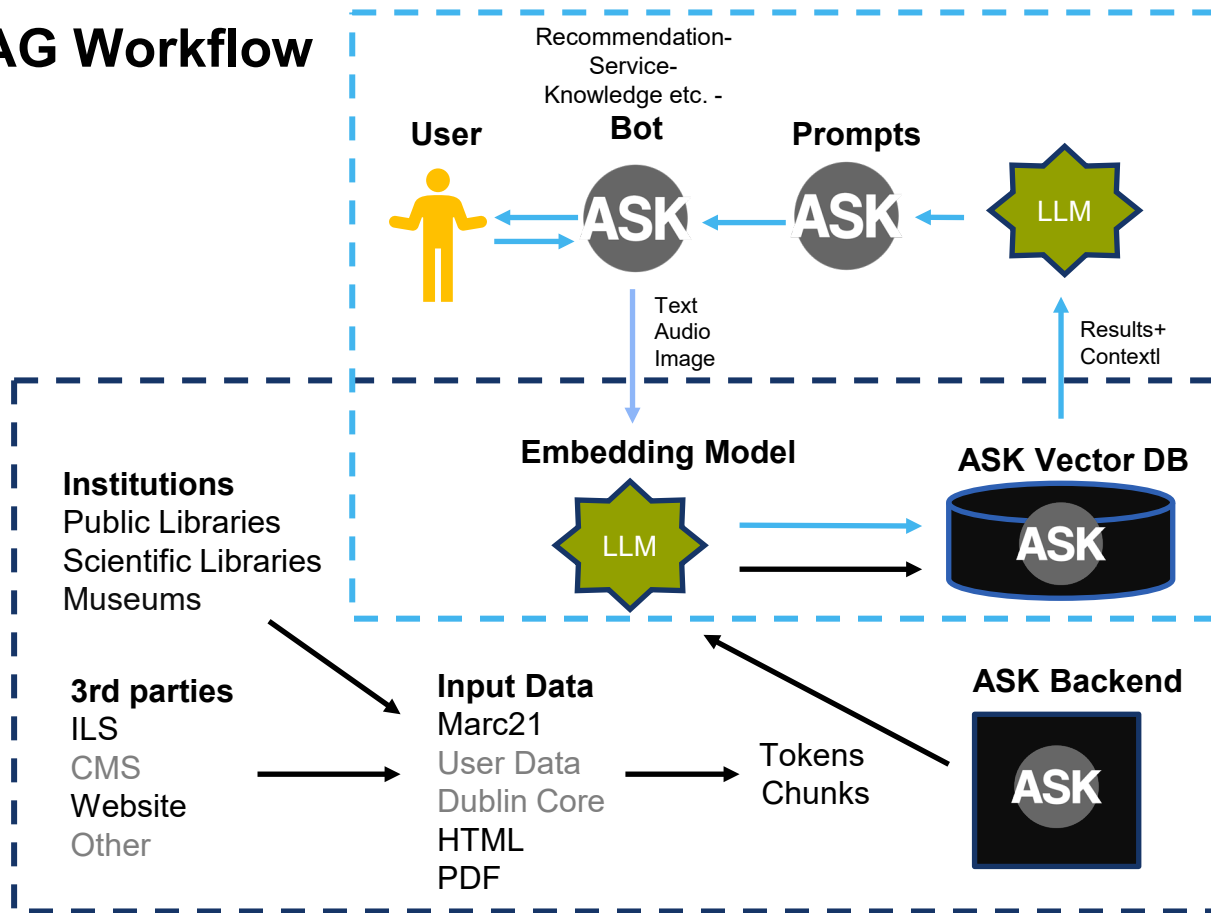
Wissens-Bot

→ Sicherheit, versch. Datenformate

Medien-Bot

→ Zusammenfassungen, Quellen, Kosten

RAG Workflow



Conversations

Terms:

- Tokens
- Chunking
- Dimension
- Embedding
- RAG
- graphRAG
- Hybrid Search
- Context
- Hallucinations
- Prompting
- Metrics

Metadata

Alternative: RAG vs Fine Tuning



Prompt: A librarian holds a book and appears to teach a child. Style of Abd al-Samad and Mir Sayyid Al.



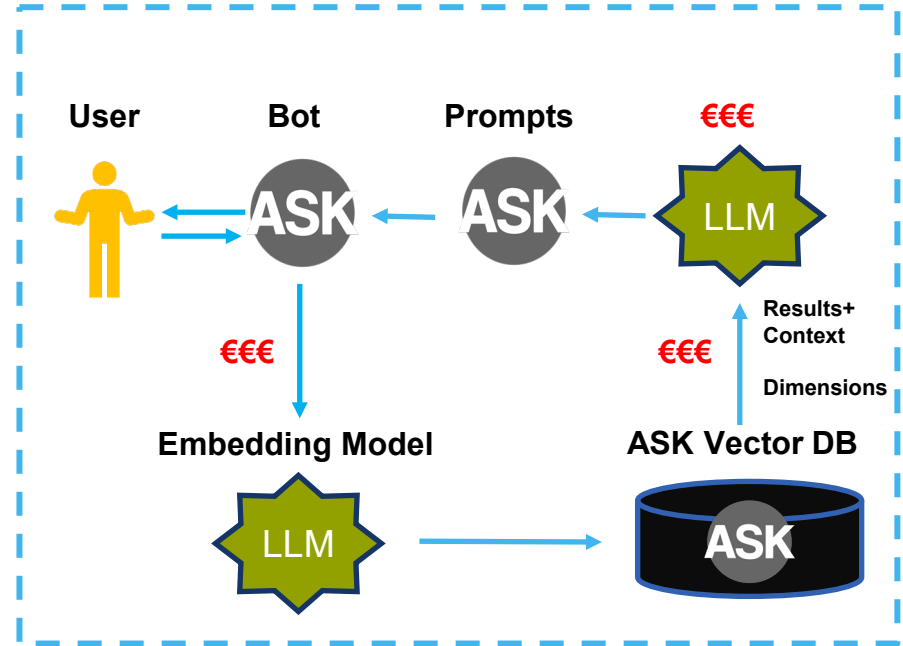
Prompt: Take the same image and let the child hold the book

Kompromisse & Eigenheiten

(die technologisch bedingt sind)

Die Kompromisse

- **Größe des LLMs** vs Kosten vs Geschwindigkeit vs. Nutzbarkeit der Ergebnisse (nicht: Präzision!)
- **Kontrolle, Datenschutz** vs. Kosten und Aufwand (Selbst hosten – API-Nutzung)
- Eierlegende Wollmilchsau vs. KI-Team
- **Prompting, Dimensionen & Kontext** vs. Kosten vs Geschwindigkeit vs. Nutzbarkeit der Ergebnisse



Trade Offs - Die Kompromisse - Live

```
HTTP 200 OK
Allow: GET, PUT, PATCH, DELETE, HEAD, OPTIONS
Content-Type: application/json
Vary: Accept

{
  "id": "bf04723e-8ee3-46ee-92c8-2cda7a3e20bb",
  "response": "Adolf Hitler was the leader of the National Socialist German Workers' Party (NSDAP) and served as Chancellor of Germany from 1933 to 1945. He is best known for his role in instigating World War II and for the Holocaust, which led to the systematic extermination of six million Jews and millions of others deemed undesirable by the Nazi regime. Hitler's aggressive expansionist policies and totalitarian rule had a profound impact on Europe and the world, leading to widespread destruction and loss of life. His ideology was rooted in extreme nationalism, anti-Semitism, and the belief in Aryan racial superiority.

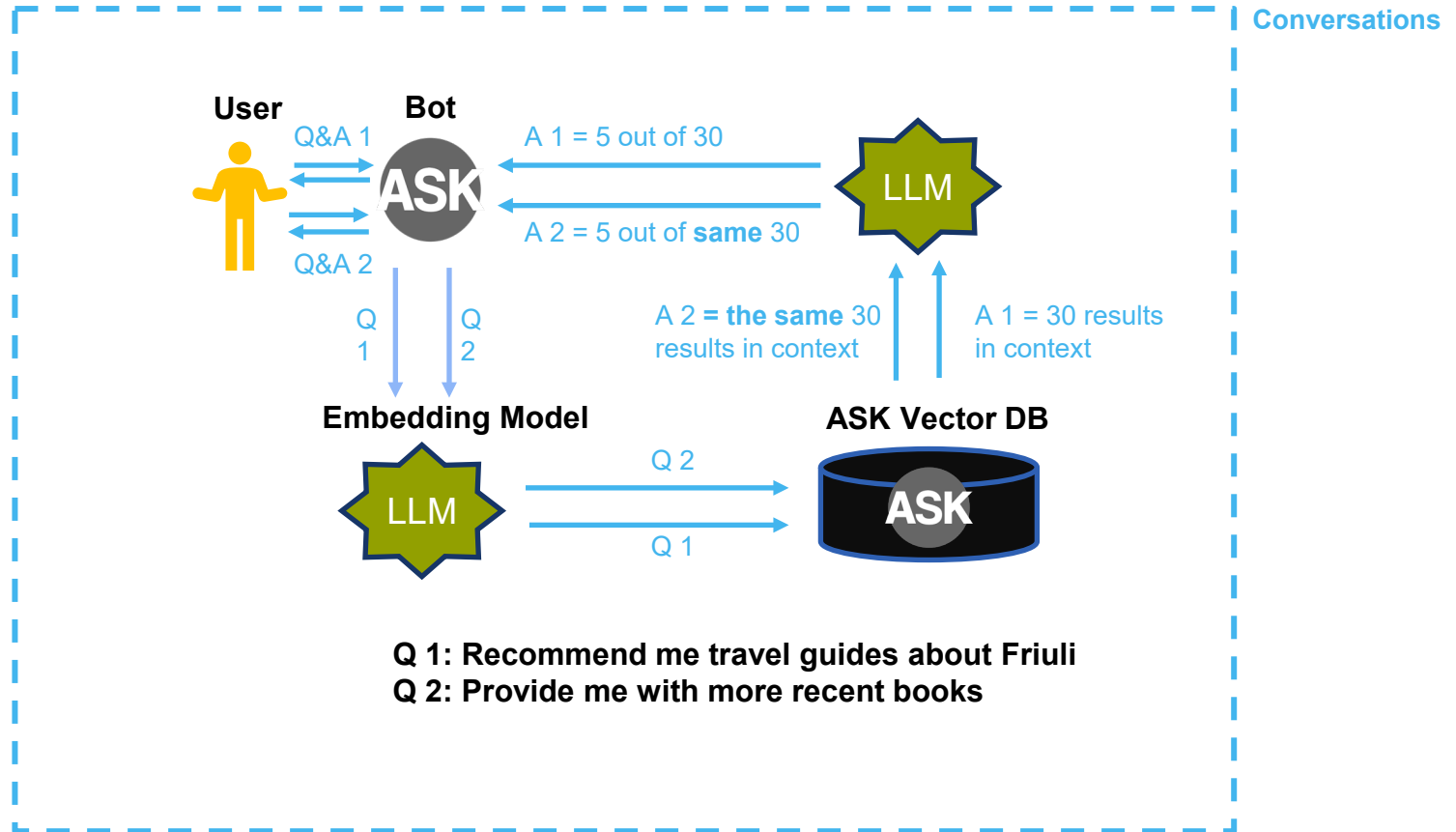
  If you're looking for books or films about Hitler or related topics, I can help you find some recommendations. Would you like that?",
  "state": {
    "messages": [
      {
        "role": "assistant",
        "content": "Hallo! Ich bin dein hilfreicher KI-Empfehlungsassistent für den Verbund der Öffentlichen Bibliotheken Berlins. Ich erweitere die klassische Suchfunktion, indem ich dir gezielte Empfehlungen aus unserem umfangreichen Katalog anbiete.

        Stell dir vor, du planst eine Reise und möchtest die besten Reiseführer für dein Ziel finden, aber auch spannende Romane, die dort spielen, um dich in die Kultur einzutauchen. Ich kann dir helfen, eine Liste von Reiseführern und Romanen zu erstellen, die perfekt auf dein Reiseziel abgestimmt sind.

        Was für eine Art von Empfehlung suchst du heute?"
      },
      {
        "role": "user",
        "content": "Tell me something about Hitler"
      }
    ]
  }
}
```

Umfang von Prompt & Kontext

Kompromiss Kontext



Eigenheiten KI - Lösungsansätze

- Wahrscheinlichkeit Expectation Management, Suche (-)
- Ähnlichkeitssuche Function Call „Exclude“
- Seltene Namen Hybrid Search
- Antwortmaschine Vorgaben, Links, Metadaten
- Strukturierte Daten (-) Metadaten aufräumen, Vektorisierung, Finetuning
- Unerwünschte Inhalte Prompts verbessern

Wahrscheinlichkeit vs Präzision und Vorhersagbarkeit

Die Katze jagt

Samurai 5%
Maus 70%
Eichhörnchen 20%
Tempel 5 %
Fuji 0%



Prompt: Eine Katze, die sich überlegt, ob sie einen Hund, eine Maus, ein Eichhörnchen, einen Jungen oder ein Haus jagen soll. Stil Hokusai

Ähnlichkeitssuche

Frage:

Kannst du Reiseführer für den Friaul empfehlen?

Problem:

Das Ähnlichste zu einem Reiseführer über den Friaul ist derselbe Reiseführer in anderer Auflage

Lösung:

Exclude Function



Prompt: an image with 4 boys, looking and dressing similar because they are quadruplets. Michelangelo style

Seltene Eigennamen

Frage:

Gibt es Bücher von Mosolff?

Problem:

- LLMs speichern Vektoren bzw Bedeutung, keine Buchstabenfolgen
- Eigennamen haben keine Bedeutung, und wenn sie sehr selten sind, gibt es auch keinen Kontext, in den sie eingebettet sein könnten.

Lösungen:

Hybride Suche, Zeit



Prompt: design an image of A crazy scientist who is writing a large volume of a book, Sumi-e style

Seltene Eigennamen - Live

☒ Aktiviere den `text\_search\_input` Parameter der Funktion `load\_embeddings` *

Aktiviert den `text\_search\_input` Parameter der Funktion `load\_embeddings` der es dem Sprachmodell erlaubt Volltextsuchen auszuführen.

Sprache für Volltext-Suche *

Deutsch

Normalisiere die Suchanfragen von `text\_search\_input` entsprechend der ausgewählten Sprache.

Beschreibung des Parameters `text\_search\_input` der Funktion `load\_embeddings` *

Search for entries that have a text matching the given input. These are provided in addition to the entries found by the parameter `query`. The search input is handled by PostgreSQL's `websearch\_to\_tsquery`. The input should be in German.

Eine Beschreibung des Parameters `text\_search\_input` den wir dem Sprachmodell zur Verfügung stellen

☐ Aktiviere die Funktion `find\_similar` *

Stellt dem Sprachmodell eine Funktion `find\_similar` zur Verfügung, die es ermöglicht, ähnliche RAG Einträge zu einem gegebenen RAG Eintrag zu suchen

Beschreibung der Funktion `find\_similar` *

Search for entries similar to a given catalog entry id returned by load_embeddings. Make sure to include the title or other distinguishing elements of the original entry in the exclude list to avoid retrieval of identical works. Activate this whenever a user

Antwortmaschine

Frage:

Bis wann kann ich wo das Buch heute zurückgeben

Problem:

Antwort muss aktuell und Präzise sein

Lösungen:

Prompten und / oder verlinken



Prompt: a khmer sculpture representing a know-it-all

Antwortmaschine – Live

Prompt für den Chatbot *

Information About VÖBB Libraries

- Annual membership typically costs 10 €, free for persons under 18.
- No library card is needed to use the desks.
- Accounts can be managed online at [Mein Konto](adisintern:*SBK).
- Users can register in person or online at [Online-Anmeldung](/ausweis).
- Fees can be paid online or in person.
- For contact info, opening hours, library cafés, accessibility, and locations, refer users to [Kontakt & Standorte](adisintern:*SW320).
- An overview of all VÖBB's online media is available at [Digitale Angebote](adisintern:*SW2). These include Onleihe, Genios, OverDrive, Pressreader, AVA, Filmfriend, Tigerbooks, Brockhaus, Munzinger, Duden, Freegal Music, Naxos Music, and phase6

Der Prompt der zur Instruktion des Chatbots genutzt wird. Wird als Jinja2 Template ausgewertet.



Sprache des Nutzers erkennen

Erkennt die Sprache der Nachrichten des Nutzers. Stellt die erkannte Sprache als Template Variable `detected\_language` bereit

Strukturierte Daten vs. Natürliche Sprache

Frage:

Ich suche Zeitschriften zur Gesundheit

Problem:

Metadaten sind - nach dem Verständnis eines LLMs - nicht immer schlüssig.

Lösungen:

- Prompten der konkreten Metadaten
- Umfassendere Metadaten
- Importer anpassen

→ Forschung!
zB Gewichtung, Finetuning



Prompt: Two women discussing with each other. One in a baroque costume the other dressed strictly in a kimono, looking a bit in a bad mood. Rubens style.

Strukturierte Daten vs. Natürliche Sprache

The title is Bewegungstherapie und Gesundheitssport B & G ; offizielles Organ des Deutschen Verbandes für Gesundheitssport und Sporttherapie e.V. (DVGS).

....

A variation of the title is "B & G".

A variation of the title is "B und G".

The link id is 990125318930206441.

This catalog entry refers to an **electronic resource**.

This catalog entry has been produced in Germany.

The physical description is **Online-Ressource**.

The content type is text.

The media type is computer.

The carrier type is online resource.

[Link](<https://doi.org/10.1055/s-00000119>).

[Link](<https://ezb.ur.de/?2141948-6>).

Index terms are **Zeitschrift**.

Vision – Netzwerk Wissen und Kultur

Select a role

☒ Tourist

☐ Child

☐ Scholar

Artwork History

 Marble statue of a youthful Hercules

+ Start New Chat



 **Marble statue of a youthful Hercules -**



Welcome to The MET! Enter the ID of a particular work. The bot will take it from there. For those who want to know more, we will suggest literature from the Watson library.

Please choose your role and get started.

☒ Tourist

☐ Child

☐ Scholar

 **Marble statue of a youthful Hercules -**



The "Marble Statue of a Youthful Hercules" is a Roman sculpture from the Early Imperial period, specifically the Flavian era (69–96 CE). This impressive marble piece stands over 97 inches tall and was originally part of the Giustiniani Collection in Rome before becoming part of the Metropolitan Museum of Art in 1903, thanks to the generosity of Mrs. Frederick F. Thompson. The statue underwent significant restorations in the early 17th century, with

Ask anything



ASK

Hausaufgaben

Metaprompts des VÖBB-Chatbots sind frei verfügbar unter <https://github.com/voebb-dev/voebb-chatbot>

1. Testen Sie den VÖBB-Chatbot und identifizieren Sie Probleme, Fehlfunktionen, Schwächen.
2. Liegt hier eine prinzipielle Schwäche vor (s.o.), oder besteht die Chance, es per Prompting besser zu machen?
3. Falls zweiteres: Versuchen sie den Metaprompt anzupassen, um die Probleme zu beheben.

In einer weiteren Sitzung werden wir ausgewählte Ansätze live im Bot testen. Sie können sich hier anmelden:

12 März um 11.00 Uhr!

Anmeldung hier:

<https://us06web.zoom.us/join/zoom/register/z73CsiVWQb-ILQgWUVbdWQ>



Umfassende Antworten vs. Unerwünschte Inhalte

Frage:

Ich suche Bücher über den Nationalsozialismus

Problem:

Durch Prompten ist es verboten, Bücher von und mit Hitler zu empfehlen. Damit werden die Empfehlungen im Kontext vom Nationalsozialismus insgesamt zu stark eingeschränkt.

Mögliche Lösungen:

Feiner Granuliertes Prompting

Und Jetzt?

**Spielen, Testen, Lernen, Machen,
Teilen,**

<https://aboutsomethink.org/de/preise/>

Danke !

tilman.scheel@aboutsomethink.org

<https://www.linkedin.com/in/tilman-scheel-2221297/>



<https://aboutsomethink.org/webinars>