

Herzlich Willkommen zum inovex Meetup!



inovex ist ein innovations- und qualitätsgetriebenes IT-Projekthaus mit dem Leistungsschwerpunkt „Digitale Transformation“.

Aktuelle Schwerpunkte:

- Agile Transformation
- Product Development Workshops
- E-Health
- Recommender Systems
- Generative AI

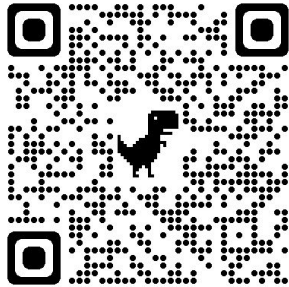


Schön, dass du hier bist

- bleib in Kontakt!



We're hiring!



Podcast



Überall, wo es Podcasts gibt: Spotify,
Overcast, Pocket Casts, ...

Soziale Medien

Insta: @inovexlife

LinkedIn: @inovex GmbH

Agenda

Sprachmodelle, Haystack und Pinselstriche: Ein Abend der Generativen KI

18:00 Uhr | Doors open and greeting

18:30 Uhr | Spielend leicht: LLM-Anwendungen bauen mit Haystack
(Christian Gill)

19:15 Uhr | Pizza Break

19:45 Uhr | Master of Art Forgery (Pascal Fecht)

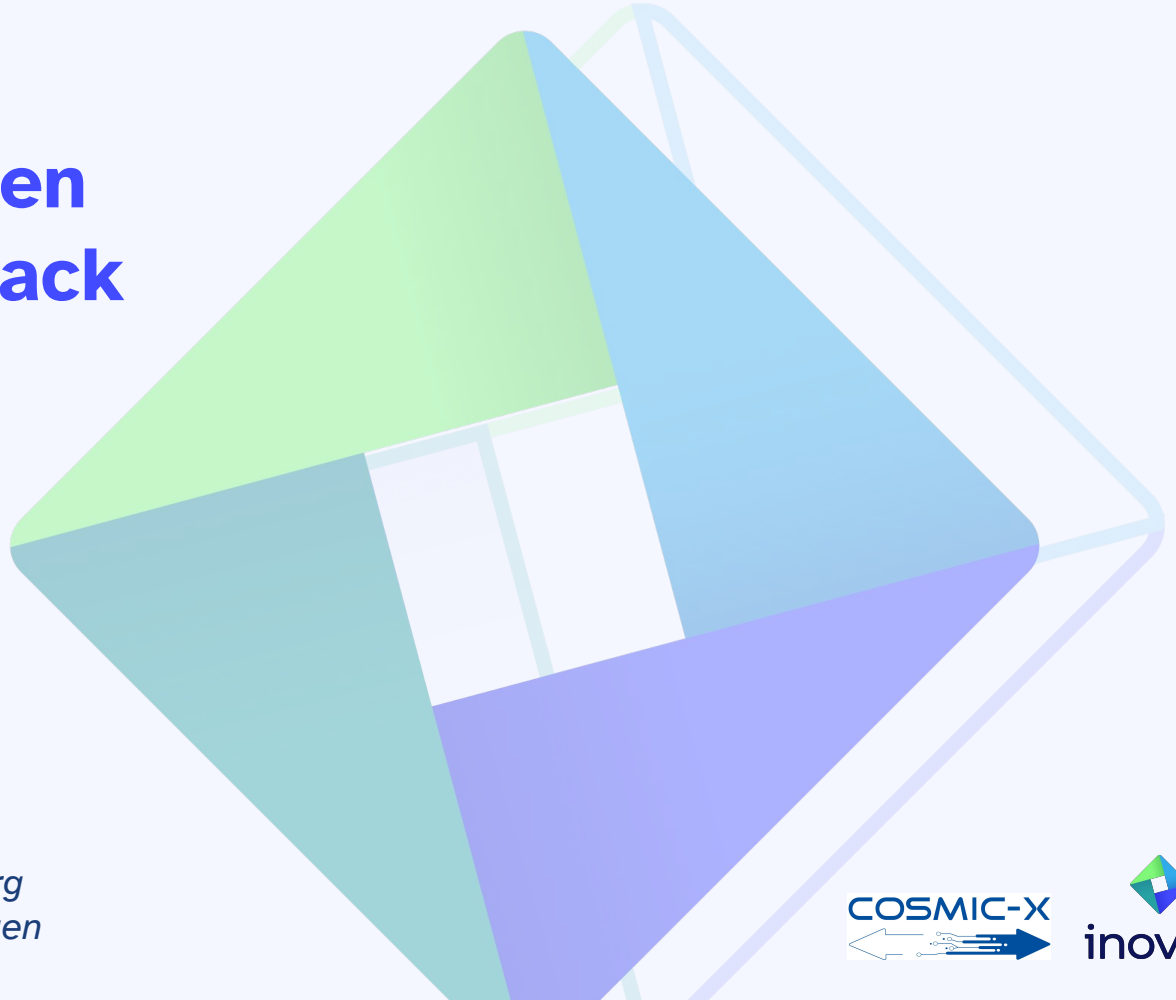
20:30 Uhr | Closing, drinks & networking

Spielend leicht: LLM-Anwendungen bauen mit Haystack

22. Februar 2024

Team inovex

*Karlsruhe · Köln · München · Hamburg
Berlin · Stuttgart · Pforzheim · Erlangen*



Motivation

- LLMs in aller Munde
 - GPT, AlephAlpha, Claude, LLaMa
- Beinahe unendliche Möglichkeiten
 - Erstellung neuer Texte
 - Generierung von Code
 - Zugang zu Wissen erleichtern
 - und vielem mehr...

Kunden Use Case



“Maschine X12 gibt merkwürdige Geräusche von sich”

Unternehmensdaten



internes Wissen



Dokumentation



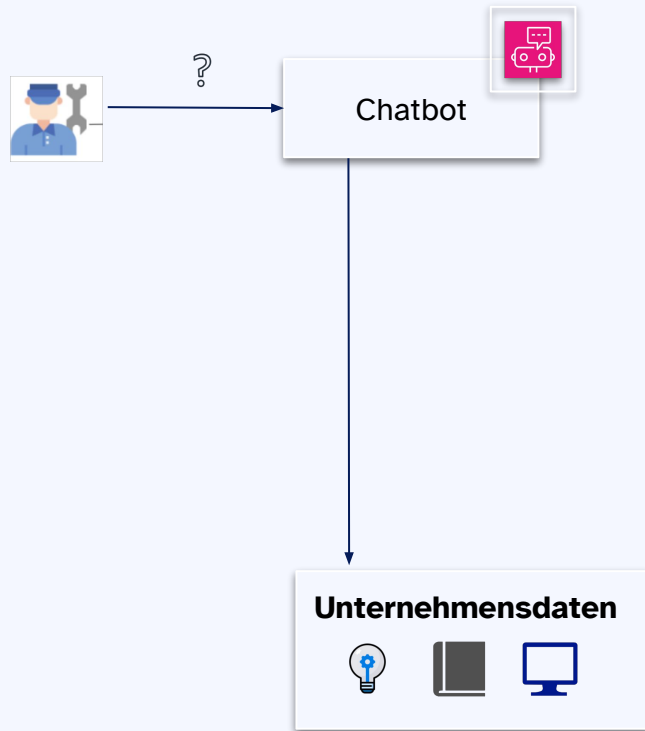
Unternehmensquellen



“Maschine X12 gibt in der zweiten Produktionsphase ein hochfrequentes Piepen von sich.”

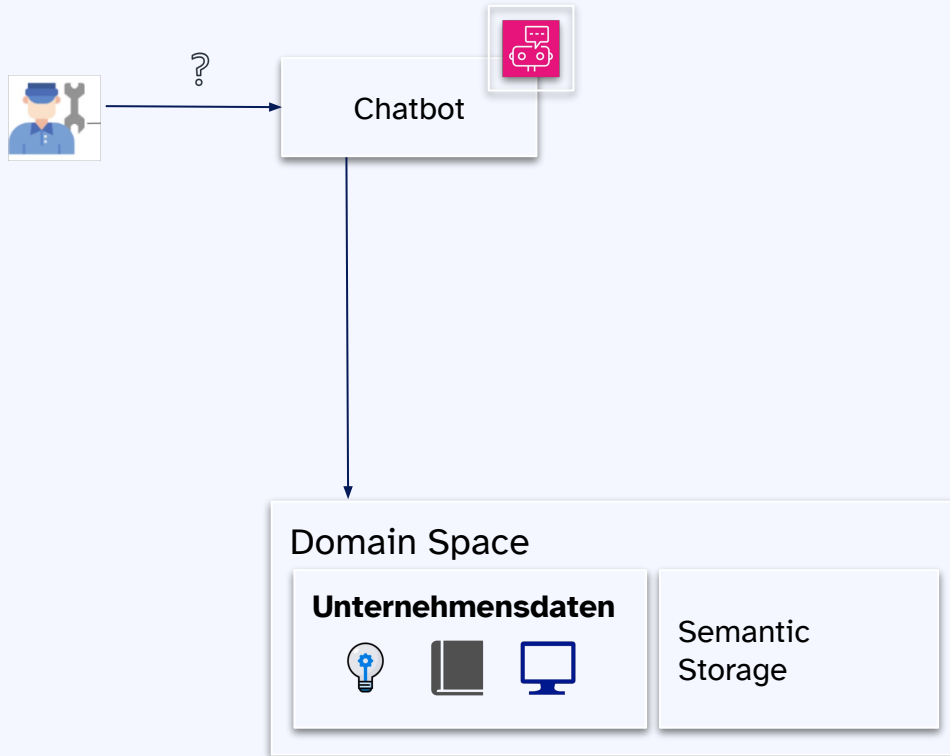
→ Wir benötigen **semantische Suche** um eine gute Antwort auf Domänenfragen zu erhalten

Nutzung von Fachwissen



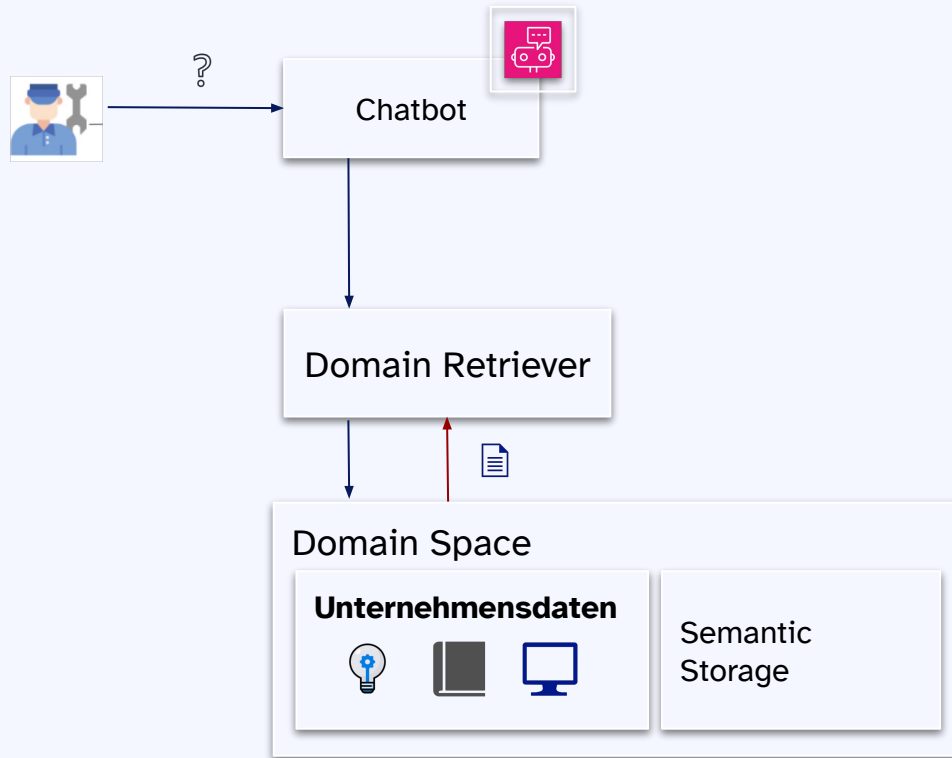
→ Eingabe

Nutzung von Fachwissen



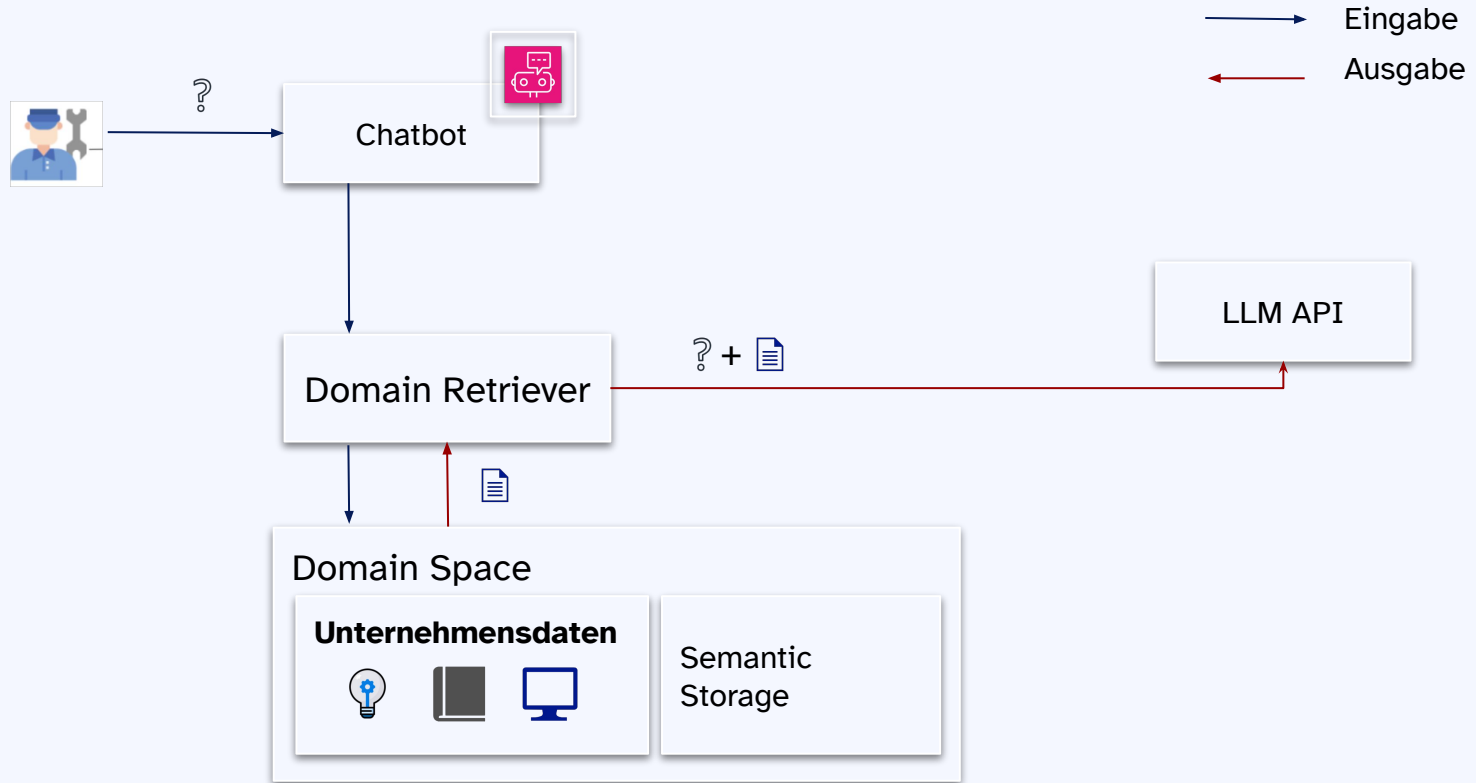
→ Eingabe

Nutzung von Fachwissen

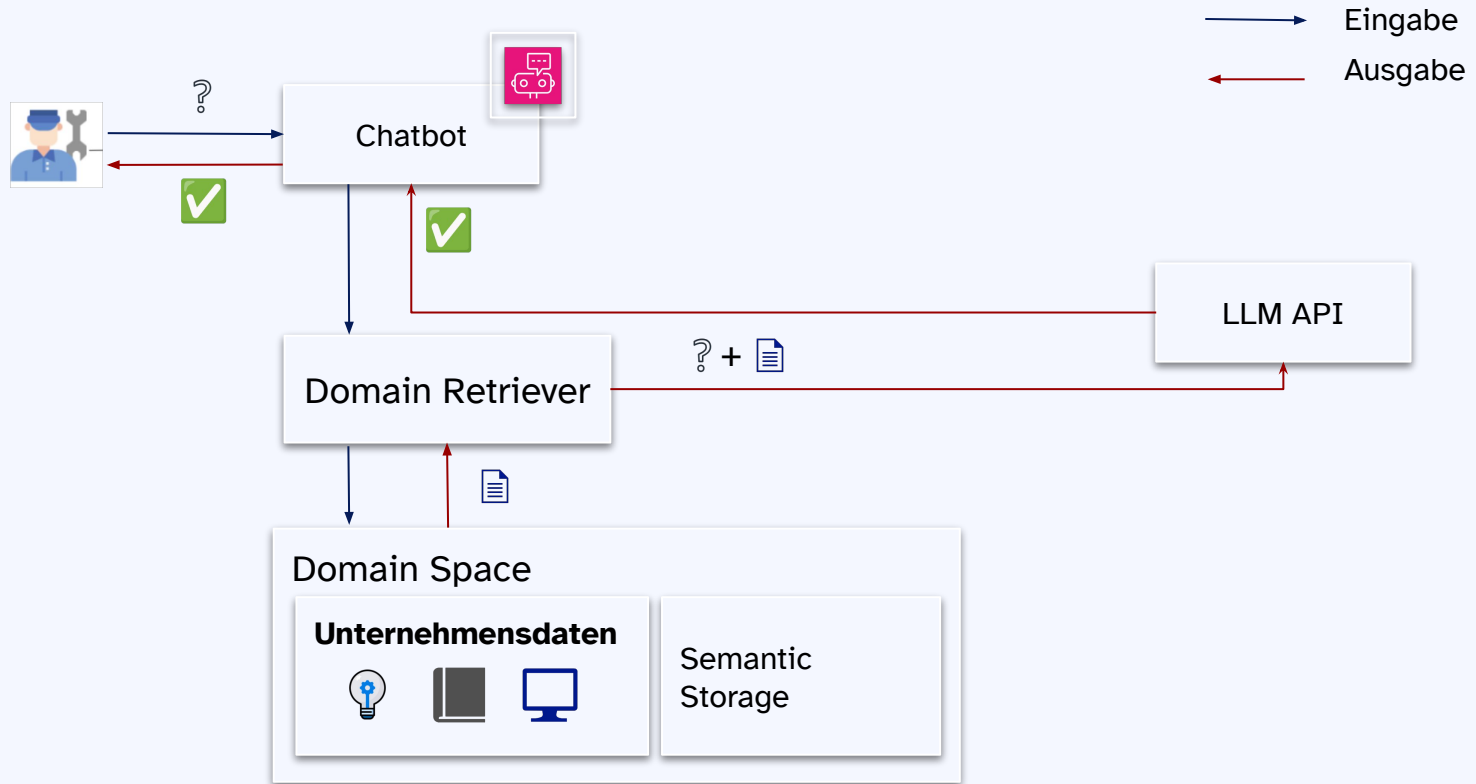


→ Eingabe
← Ausgabe

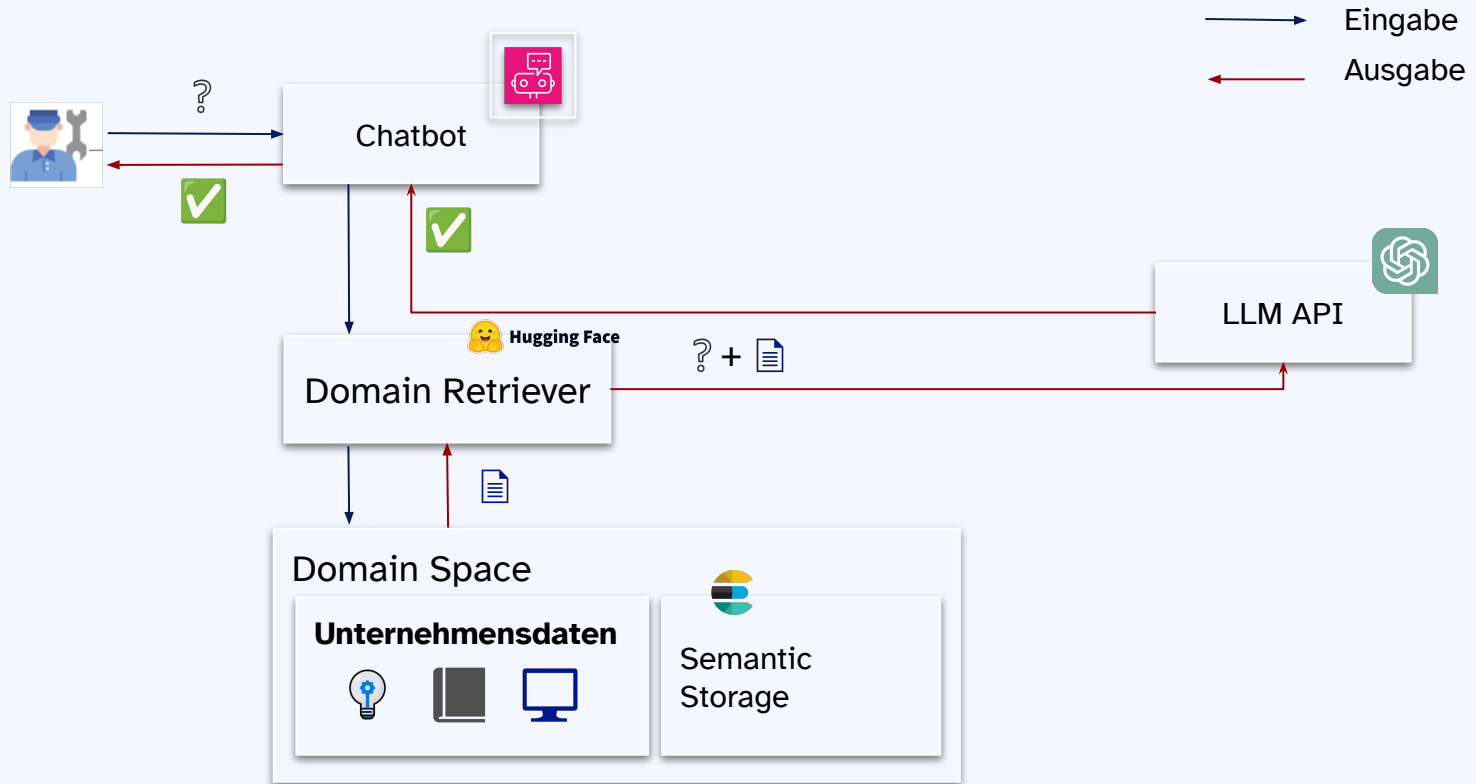
Nutzung von Fachwissen



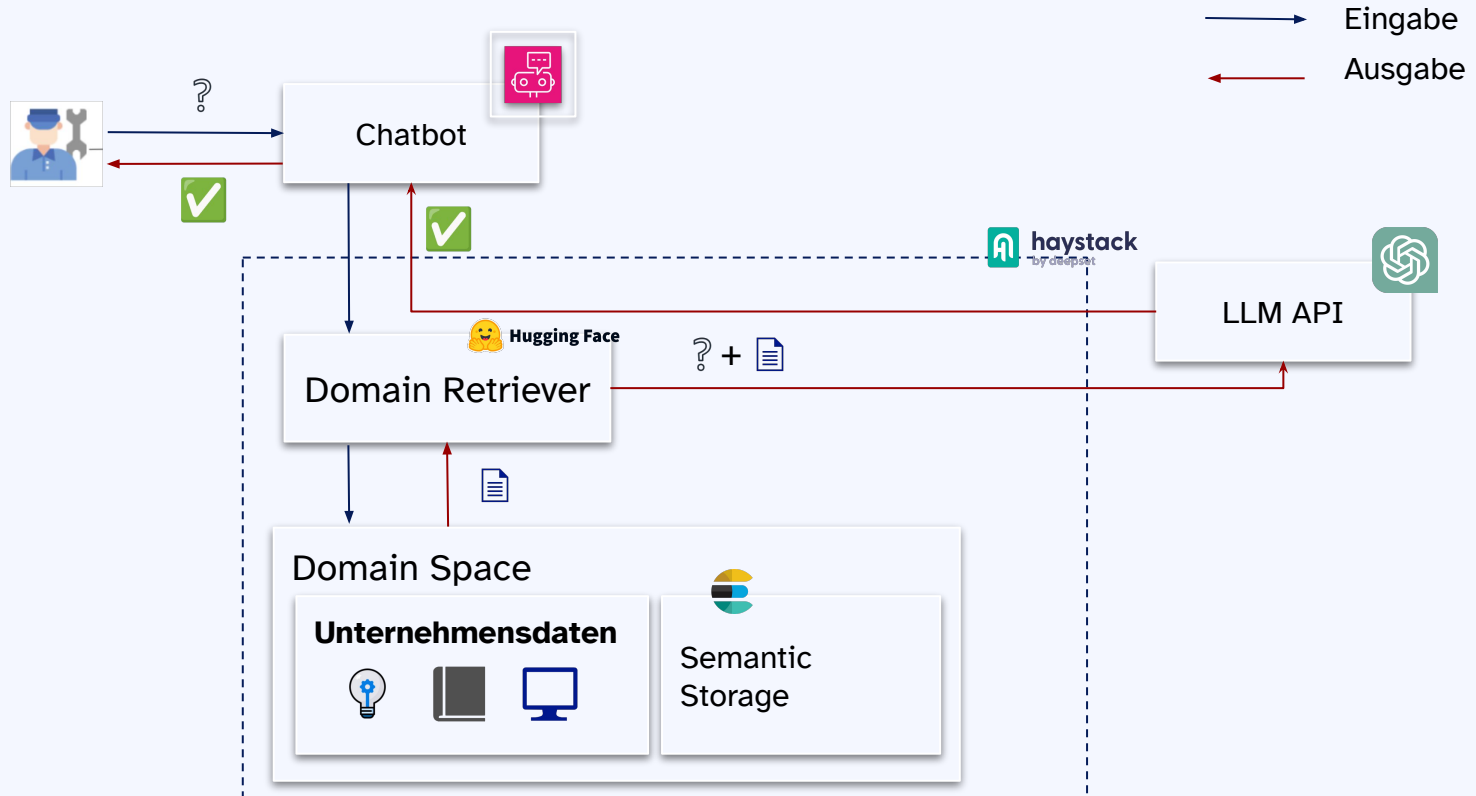
Nutzung von Fachwissen



Nutzung von Fachwissen



Nutzung von Fachwissen



Haystack: Ein Framework für LLM Anwendungen

All content was based on Haystack version 1.21.2

Was ist Haystack?



Open-Source **LLM Orchestrierungs-Framework** von [Deepset](#)

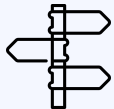
- 12.8k Sterne auf Github
- Verwendung der neuesten LLMs: OpenAI, Anthropic, Cohere, Open-Source LLMs
- Integration in gängige Datenbanken wie z.B. ElasticSearch, Vektordatenbanken (z.B. Qdrant, Milvus)



Semantische Suche, QA, Chatbots



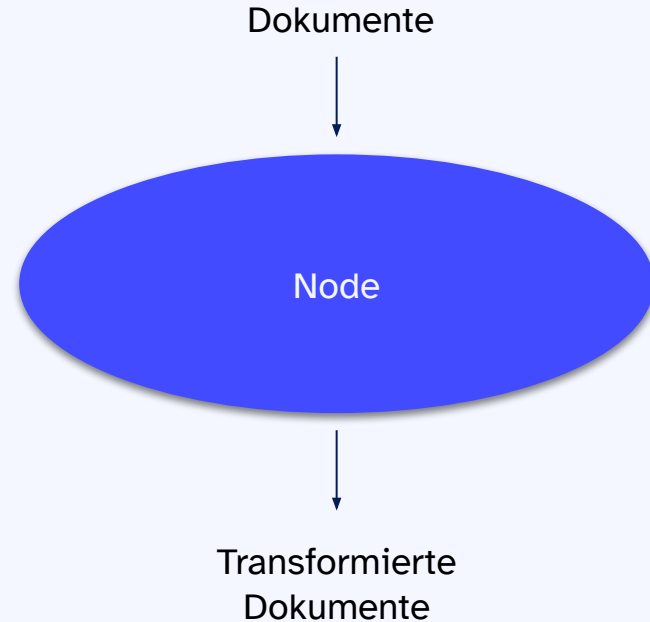
Leichtgewicht: schnelle Entwicklung von POCs



Alternativen: LangChain, LlamaIndex

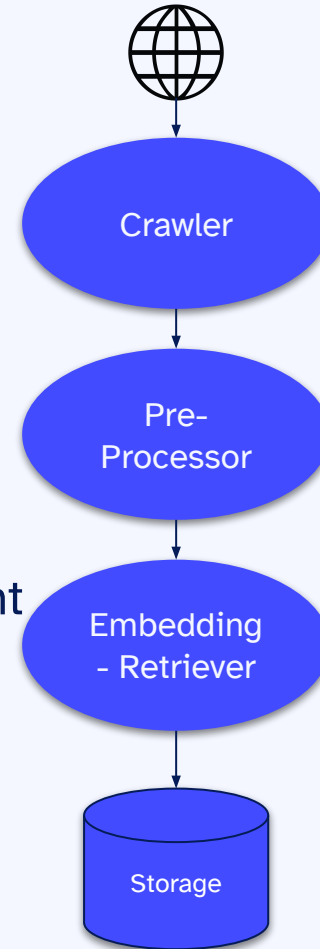
Haystack Nodes

- Node implementiert nur eine Funktion
- Node passt den erhaltenen Inhalt an
- Plug & Play



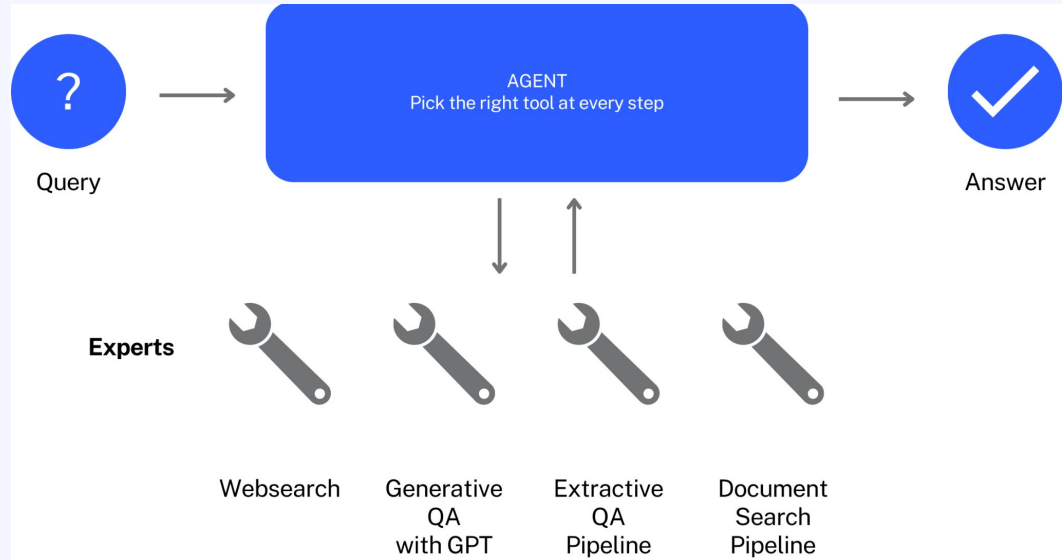
Pipelines: Zusammenspiel der Nodes

- Jede **Node** umfasst eigene Funktion
 - Crawler
 - Preprocessor
 - Embedding-Retriever
 - Prompt-Node
- **Daten** werden zwischen den Nodes ausgetauscht
- Ergebnis wird im **Document-Store** gespeichert

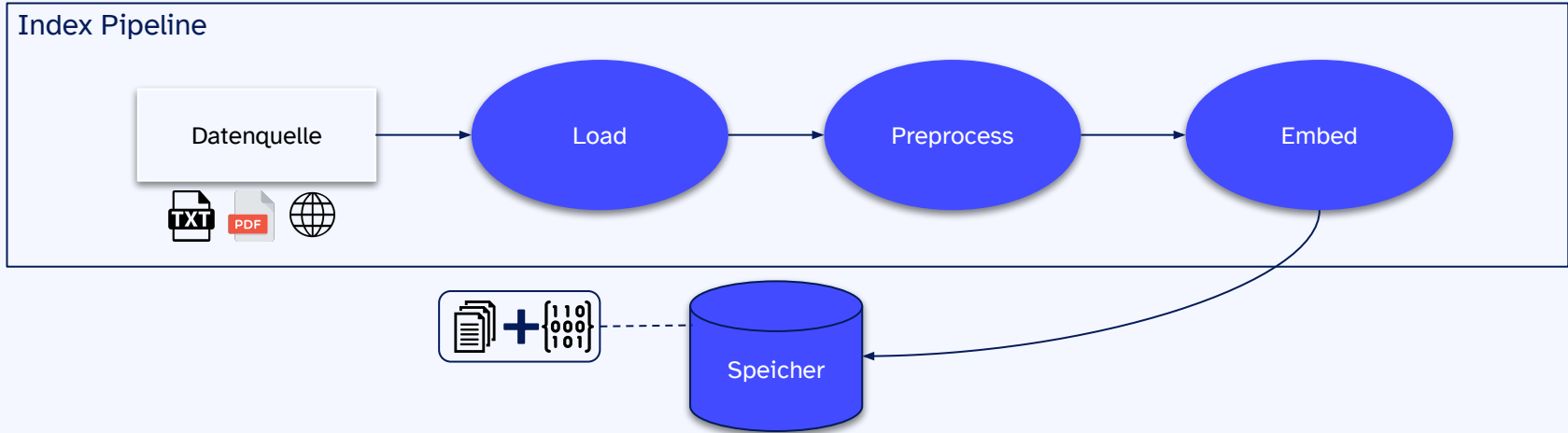


Agents: Intelligente Entscheidungsträger

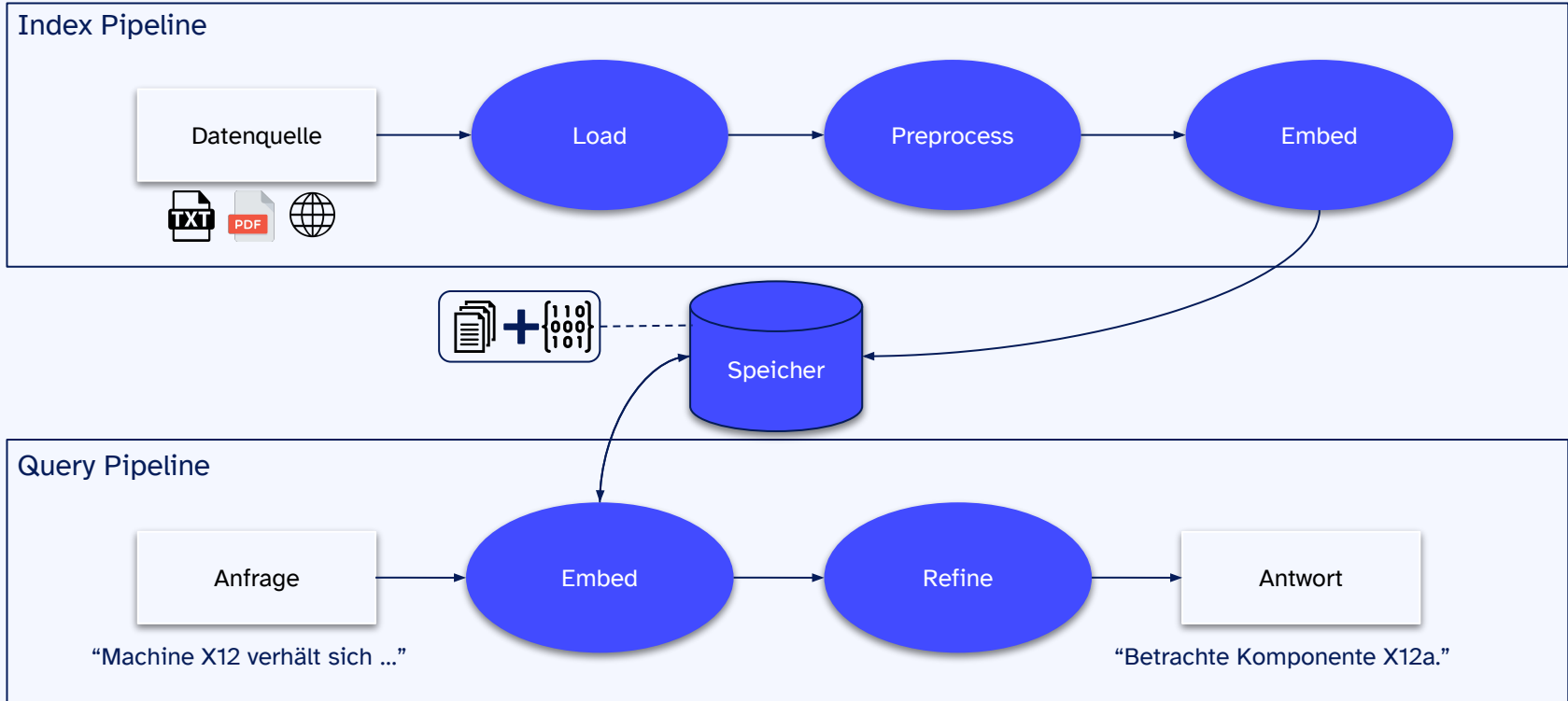
- **Entscheidende** Komponente um flexible Pipelines zu bauen
- Agent wählt das **richtige Werkzeug** um die beste Antwort zu geben



Zugang zum Wissen schaffen



Das Wissen abrufen



Demo

Fazit

Um LLM Anwendungen zu bauen..



Starte simpel, z.B. mit LLM Frameworks wie Haystack, die...

- Integrationen für gängige Tools mitbringen
- die Einbindung von Domänenwissen vereinfachen



Verwende Prompt Engineering

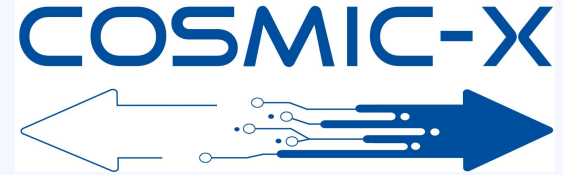
- verwende präzise Prompts und Leitfäden
- verwende Kontext um mehr Wissen bereitzustellen



Informiere dich über mögliche Einschränkungen

- Copyright & Privatsphäre, Sicherheit, Bias, ..
- bleib auf dem Laufenden mit aktuellen Gesetzen

Collaboration Project: cosmic-x



Der Inhalt dieser Veröffentlichung stammt aus dem Forschungsprojekt “**Collaborative Smart Services for Industrial Value Chains in GAIA-X (COSMIC-X)**”.

Dieses Forschungs- und Entwicklungsprojekt wird / wurde mit Mitteln des Bundesministeriums für Bildung und Forschung (BMBF) im Forschungsprogramm „Zukunft der Wertschöpfung (evtl. Förderkennzeichen ...)“ gefördert und vom Projektträger Karlsruhe (PTKA) betreut. Die Verantwortung für den Inhalt dieser Veröffentlichung liegt bei der Autorin/dem Autor



Vielen Dank!



Christian Gill

Product Owner & Data/ML Engineer

christian.gill@inovex.de

inovex ist ein innovations- und qualitätsgetriebenes IT-Projekthaus mit dem Leistungsschwerpunkt „Digitale Transformation“.

- gegründet 1999
- 500+ Mitarbeitende
- 8 Standorte in ganz Deutschland



www.inovex.de



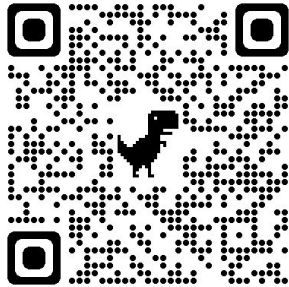
Noch Fragen?

Schön, dass du hier bist

- bleib in Kontakt!



We're hiring!



Podcast



Überall, wo es Podcasts gibt: Spotify,
Overcast, Pocket Casts, ...

Soziale Medien

Insta: @inovexlife

LinkedIn: @inovex GmbH

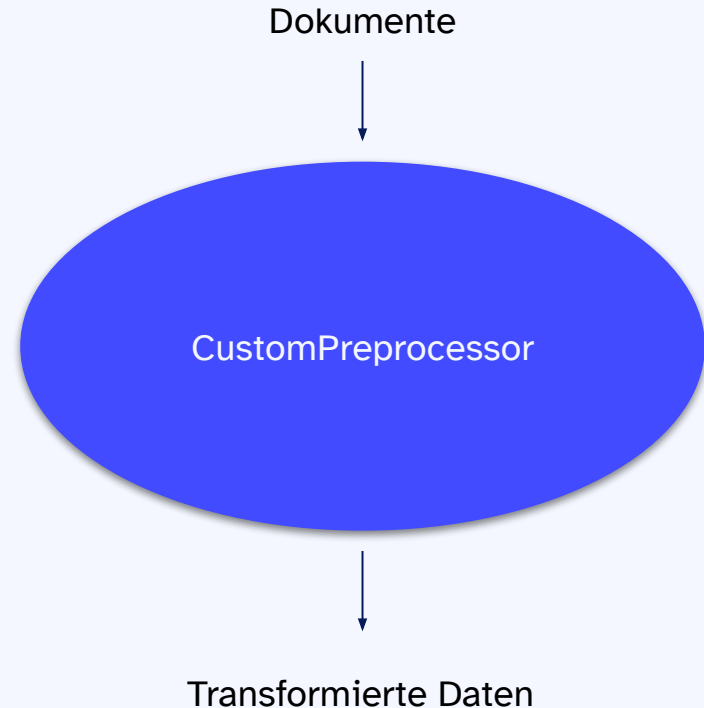
Custom Nodes: Flexibilität in der Verarbeitung

Eigene Aufgaben wie z. B.

- Vorverarbeitung: Auffinden von Informationen in PDFs
- Nachverarbeitung: Filtern der Suchergebnisse

Umsetzung:

- Erbt von `BaseComponent`
- Erstelle `run()`, `run_batch()`



Langchain vs Haystack

Langchain

- komplex
- Komponenten werden durch Chains zusammengebaut
- Anbindung an sehr viele DBs und LLMs
- Umfasst zusätzliche Services für Deployment, Experimentation

Haystack

- leichter Einstieg
- einzelne Nodes können alleine verwendet werden
- Anbindung an gängige DBs und LLMs
- Stellt “nur” REST API bereit