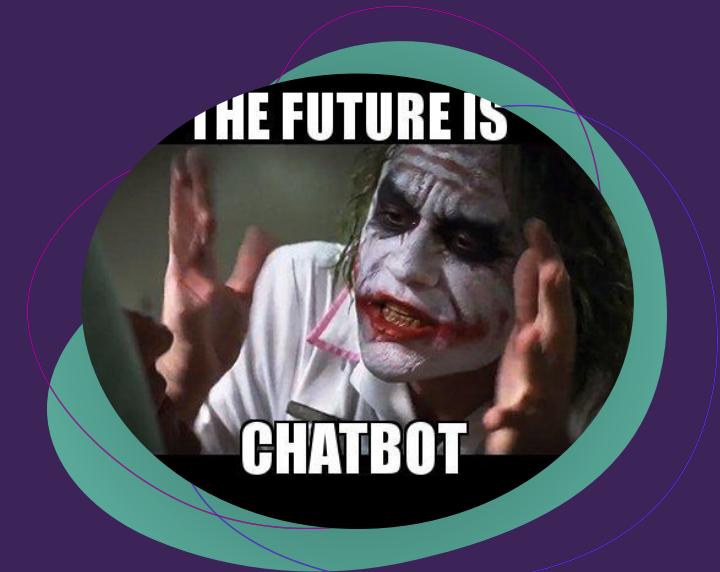


Flexible Chatbots with Elasticsearch



What we'll do today



Agenda

01

Introduction

Fabio & Christoph

02

Motivation, Requirements & Solution

Christoph

03

LLMs, Demo, Learnings & Takeaways

Fabio

04

Q&A

Fabio & Christoph

Team



**Fabio
Ziegler**
IT Consultant



**Christoph
Eder**
Senior IT Consultant /
Elastic Partner Manager

Your IT partner with over 12 years of Elastic experience



EBCONT & ELASTIC

- » Austrian IT service provider
- » > 25 years & 700 employees
- » ~ 120 customer
- » Full IT service provider
 - » Software development
 - » Consulting
 - » Operations
 - » Security

- » Elastic ELITE Partner
- » > 13 years if experience
- » > 50 Elastic customers
- » Full Scale Elastic Consulting
 - » Implementation
 - » Setup & Upgrade
 - » Operations
 - » Subscriptions



Flexible Chatbots with Elasticsearch

Motivation & Requirements

cas.ai-ml-playground.ebcont.com/c/5...  Confluence AI 

GPT-4o-mini GPT-4o-mini

Sie Wie beantrage ich Urlaub?

NEUEN CHAT STARTEN

GPT-4o-mini

Claude 3.5

Claude 3

EBCONT-Chatbot

Um Urlaub zu beantragen, müssen Sie den Urlaubsantrag in SAP erfassen. Hier sind die Schritte, die Sie befolgen sollten:

Schritte zur Beantragung von Urlaub:

- Antrag erfassen:** Erfassen Sie Ihren Urlaubsantrag in SAP. Stellen Sie sicher, dass es sich um einen regulären Urlaubstag handelt und nicht um einen Feiertag oder Sonderurlaubstag.
- Ganztägige Abwesenheit:** Beachten Sie, dass nur ganztägige Abwesenheitsanträge zulässig sind.
- Kommentare hinzufügen:** Falls Sie Sonderurlaub beantragen (z.B. für Hochzeit, Umzug oder Sterbefall), geben Sie den Grund im Kommentarfeld an und fügen Sie erforderliche Bestätigungen (z.B. Heiratsurkunde, Meldezettel) bei.

Wichtige Hinweise:

- Der Urlaubsstand wird am Eintrittsdatum um den im Vertrag festgelegten Wert erhöht. Achten Sie darauf, dass das Eintrittsdatum korrekt gesetzt ist, da sonst der Urlaubsstand nicht erhöht wird.

Falls Sie weitere Informationen oder Unterstützung benötigen, lassen Sie es mich wissen!

Nachricht 

Motivation #1

Lots of data in the intranet that can't be found...



BEREICHSVERKNÜPFUNGEN

[Willkommen bei EBCONT](#)[Good to know](#)[Vorlagen](#)[EBCONT Toolstack](#)[Technologien und Partnerschaften](#)[Software & Security](#)[Circles und Events](#)[Projekt Guidelines](#)[Knowlegebase Infrastructure / Support](#)[Tools und Links für den Arbeitsalltag](#)[EBCONT Maps](#)[EBCONT PERT-Tool](#)[EBCONT Talent Tool](#)[EPICON - Charity Marketing](#)[EBCONT Let's talk about Health](#)[Blogs](#)[SAP](#)[HR4you - CV Maintenance Guidelines](#)[Bereich konfigurieren](#)

Seiten

EBCONT HOME

Erstellt von Gregor Sieber, zuletzt aktualisiert von Markus Stangl am Dez. 11, 2024 • Lesedau...

welcome!

INTRO

Liebes EBCONT Team,

willkommen auf EBCONT HOME, unserer neuen Startseite auf Confluence.

Hier findet ihr wichtige Unternehmensinformation zu unserem **EBCONT Alltag** (...

Links zu Tools und Technologien, die wir bei EBCONT verwenden, Partnerschaft...

Auch [Blogs](#), [Circles](#) und [Events](#) haben hier ihr neues Zuhause gefunden.

Als zentrale Anlaufstelle für alle organisatorischen und administrativen Fragen st...

Bei Problemen mit Eurem IT Equipment könnt ihr beim [EBCONT Operations Support](#) ein persönliches Dashboard anlegen.

Viel Spaß beim Durchklicken.

 Kontaktpersonen: [@Ilse Merkinger-Boira](#) [@Markus Stangl](#) [@Bettina L...](#)

INHALTSVERZEICHNIS

Schließen

FILTERN NACH

Bereich

Beitragender

Typ

Datum

Stichwort

Bereichskategorie

Erweiterte Suche

urlaub beantragen

109 Suchergebnisse

Tipps zur Suche

Zeiterfassung

Support EBCONT IT Dez. 12, 2024

Möglichkeiten bestehen daher: Arbeitet man an beiden Tagen NICHT: Bitte **Urlaub** oder Zeitausgleich für 31.12. **beantragen** (bzw ... erfasse...)

EBCONT HOME

EBCONT HOME Dez. 11, 2024

Stundenaufzeichnung? Zur Stundenaufzeichnung (und auch zum **Beantragen** von z.B. **Urlaub** oder Eintragen von Krankenstand

2 - Erfassung von Mitarbeiterzeiten & Abwesenheiten

Support EBCONT IT Sept. 11, 2024

Abwesenheiten (**Urlaub**, Zeitausgleich etc.) erfassen und **beantragen**. Abwesenheiten anzeigen Übersicht über ... Zeitaufzeichnung (genannt...)

_Organisatorisches

Operations Aug. 12, 2020

Ausnahme und nach vorheriger Abstimmung mit dem Vorgesetzten möglich. **Urlaub Urlaub** über SAP rechtzeitig **beantragen**

Administratives

Welcome@EBCONT Mai 08, 2024

Arbeitszeiten Zur Stundenaufzeichnung (und auch zum **Beantragen** von z.B. **Urlaub** oder Eintragen von Krankenstand) IT Sicherheit im **Urlaub**

ISMS - Information Security Management System Juni 06, 2024

Viren kennen keine Grenzen, Cyber-Kriminelle machen keinen **Urlaub**. Damit man während der schönsten Zeit des Jahres nicht zum Opfer wi... Tipps für sicheren **Urlaub** mit Laptop und Smartphone

ISMS - Information Security Management System Juni 26, 2023

1542530901703331841-images-0.jpg Quelle BSI:
<https://www.bsi.bund.de/DE/Themen/Verbraucherinnen-und-...>

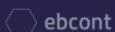
Willkommensmappe.pdf

Welcome@EBCONT / ... / Der erste Tag Juni 22, 2020

Urlaub bist. Arbeitszeiten Wir nutzen bei EBCONT für die ordnungsgemäße Erfassung der Leistungszeiten das System ... steht...

Willkommensmappe_2020.pdf

Welcome@EBCONT / ... / Der erste Tag Nov. 02, 2020



Bereiche ▾ Personen Analytics

Erstellen

...

Dashboard



EBCONT HOME



Seiten

Analytics

BEREICHSVERKNÜPFUNGEN

Willkommen bei EBCONT

Good to know

Vorlagen

EBCONT Toolstack

Technologien und Partnerschaften

Software & Security

Circles und Events

Projekt Guidelines

Knowlegebase Infrastructure / Support

Tools und Links für den Arbeitsalltag

EBCONT Maps

EBCONT PERT-Tool

EBCONT Talent Tool

EPICON - Charity Marketing

EBCONT Let's talk about Health

Blogs

SAP

HR4you - CV Maintenance Guidelines

Bereich konfigurieren

Seiten

EBCONT HOME

Erstellt von Gregor Sieber, zuletzt aktualisiert von Markus Stangl am Dez. 11, 2024 • Lesedau...

welcome!

INTRO

Liebes EBCONT Team,

willkommen auf EBCONT HOME, unserer neuen Startseite auf Confluence.

Hier findet ihr wichtige Unternehmensinformation zu unserem **EBCONT Alltag** (...

Links zu Tools und Technologien, die wir bei EBCONT verwenden, Partnerschaft

Auch **Blogs**, **Circles** und **Events** haben hier ihr neues Zuhause gefunden.

Als zentrale Anlaufstelle für alle organisatorischen und administrativen Fragen st...

Bei Problemen mit Eurem IT Equipment könnt ihr beim **EBCONT Operations Sup...** persönliches Dashboard anlegen.

Viel Spaß beim Durchklicken.

⚠ Kontaktpersonen: [@Ilse Merkinger-Boira](#) [@Markus Stangl](#) [@Bettina L...](#)

INHALTSVERZEICHNIS

Schließen

FILTERN NACH

Bereich

Beitragender

Typ

Datum

Stichwort

Bereichskategorie

Erweiterte Suche

urlaub beantragen

109 Suchergebnisse

Tipps zur Suche

Zeiterfassung

Support EBCONT IT Dez. 12, 2024

Möglichkeiten bestehen daher: Arbeitet man an beiden Tagen NICHT: Bitte **Urlaub** oder Zeitausgleich für 31.12. **beantragen** (bzw ... erfasse...

EBCONT HOME

EBCONT HOME Dez. 11, 2024

Stundenaufzeichnung? Zur Stundenaufzeichnung (und auch zum **Beantragen** von z.B. **Urlaub** oder Einträgen von Krankenstand

2 - Erfassung von Mitarbeiterzeiten & Abwesenheiten

Support EBCONT IT Sept. 11, 2024

Abwesenheiten (**Urlaub**, Zeitausgleich etc.) erfassen und **beantragen**. Abwesenheiten anzeigen Übersicht über ... Zeitaufzeichnung (genannt...)

Organisationsübersicht

Operations Aug. 12, 2020

Ausnahme und nach vorheriger Abstimmung mit dem Vorgesetzten möglich. **Urlaub Urlaub** über SAP rechtzeitig **beantragen**

Administratives

Welcome@EBCONT Mai 08, 2024

Arbeitszeiten Zur Stundenaufzeichnung (und auch zum **Beantragen** von z.B. **Urlaub** oder Einträgen von Krankenstand)IT Sicherheit im **Urlaub**

ISMS - Information Security Management System Juni 06, 2024

Viren kennen keine Grenzen, Cyber-Kriminelle machen keinen **Urlaub**. Damit man während der schönsten Zeit des Jahres nicht zum Opfer wi...Tipps für sicheren **Urlaub** mit Laptop und Smartphone

ISMS - Information Security Management System Juni 26, 2023

1542530901703331841-images-0.jpg Quelle BSI:
<https://www.bsi.bund.de/DE/Themen/Verbraucherinnen-und-...>

Willkommensmappe.pdf

Welcome@EBCONT / ... / Der erste Tag Juni 22, 2020

Urlaub bist. Arbeitszeiten Wir nutzen bei EBCONT für die ordnungsgemäße Erfassung der Leistungszeiten das System ... steht...

Willkommensmappe_2020.pdf

Welcome@EBCONT / ... / Der erste Tag Nov. 02, 2020



ebcont

Bereiche



Personen



Analytics



Erstellen



...



Dashboard



EBCONT HOME



Seiten Analytics

EBCONT HOME

Erstellt von Gregor Sieber, zuletzt aktualisiert von Markus Stangl am Dez. 11, 2024 • Lesedau...

INTRO

Liebes EBCONT Team,

willkommen auf EBCONT HOME, unserer neuen Startseite auf Confluence.

Hier findet ihr wichtige Unternehmensinformation zu unserem **EBCONT Alltag** (...

Links zu Tools und Technologien, die wir bei EBCONT verwenden, Partnerschaft...

Auch [Blogs](#), [Circles](#) und [Events](#) haben hier ihr neues Zuhause gefunden.

Alle zentrale Anlaufstelle für alle organisatorischen und administrativen Fragen st...

Bei Problemen mit Eurem IT Equipment könnt ihr beim [EBCONT Operations Support](#) ein persönliches Dashboard anlegen.

Viel Spaß beim Durchklicken.

Kontaktpersonen: [@Ilse Merkinger-Boira](#) [@Markus Stangl](#) [@Bettina L...\[@Bettina L...\]\(#\)](#)

Bereich konfigurieren



Schließen

FILTER NACH

Bereich

Beitragender

Typ

Datum

Stichwort

Bereichskategorie

Erweiterte Suche

urlaub buchen

115 Suchergebnisse

Tipps zur Suche

Jour Fixe - 2022-11-24 - Week 47

Operations Nov. 24, 2022

Urlaub Weihnachtszeit!! Bitte rechtzeitig **buchen** und im OP-Calendar auch bekannt geben. Danke. Type your task ... , GIT Repo) Resources...

IT Sicherheit im Urlaub

ISMS - Information Security Management System Juni 06, 2024

Viren kennen keine Grenzen, Cyber-Kriminelle machen keinen **Urlaub**. Damit man während der schönsten Zeit des Jahres nicht zum Opfer wi...

Tipps für sicheren Urlaub mit Laptop und Smartphone

ISMS - Information Security Management System Juni 26, 2023

1542530901703331841-images-0.jpg Quelle BSI:
<https://www.bsi.bund.de/DE/Themen/Verbraucherinnen-und-...>

Ebcont Avatar Weekly 2024.12.18

Ebcont LABS Jan. 23, 2025

milan.jovanovic@ebcont.com Rückfrage bei Valentin bezüglich Kommentar Noch offen, da Valentin auf **Urlaub** ist Milan ... Learning...

Stundenbuchungen interne Leistungen

EBCONT Führungskräfte März 05, 2024

durch interne Tätigkeiten beschäftigen, sondern vorrangig Abbau von ZA und **Urlaub** nutzen wenn sinnvoll, 50/50 Lösung für **Urlaub/ZA** anbiete...

Wien

Welcome@EBCONT Feb. 21, 2025

TELEPHONE VOTE AS THE FRONTDESK IS NOT PERMANENTLY EMPLOYED AT THE MOMENT! ABLAUF: 1) Rad **buchen**: Termin im ...

User-Dokumentation

Project-Factsheets Sept. 11, 2017

und beinhaltet folgende Tasks: **Urlaub** Arzt/Behördengang Krankenstand Sonderurlaub Es ist ein managed-project, somit ist es für Mitarbeiter...

2021-12-09 - Week 49

Operations Dez. 09, 2021

für Audit Termin: 13.12 - 14.12.2021 thomas.mistelbauer@ebcont.com Urlaubsplanung **Urlaub**, Bereitschaften ... Jira Ticket immer **buchen**.....

Daten BOA

Project-Factsheets Juni 16, 2017

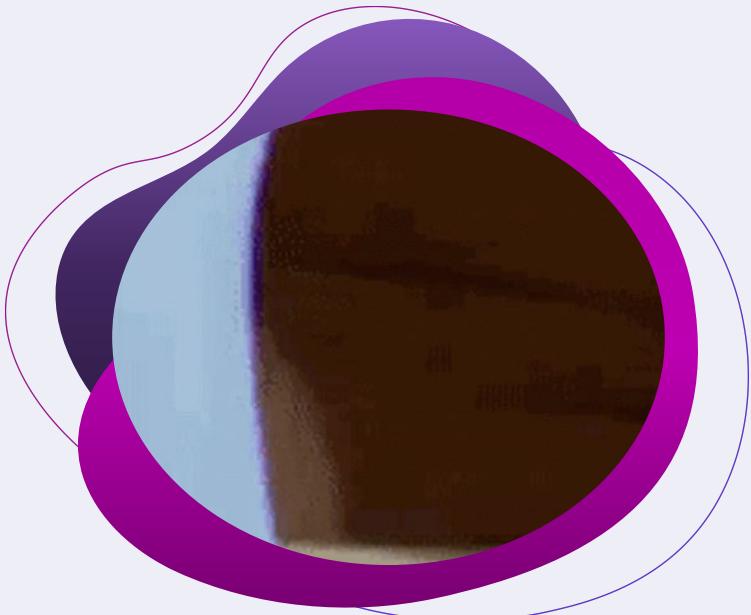
Motivation #2

*Our employees are building their own
solutions*



Motivation #3

Our customers have the same problem



Requirements

- Easy to use
- Good results
- “Enterprise-ready”
 - Connectors
 - Maintainable
 - Access permissions
 - Scalable and flexible
 - APIs
- Cheap 😊

Requirements

- Easy to use → Chatbot
- Good results → Prompt Engineering
- “Enterprise-ready” → Elastic
 - Connectors
 - Maintainable
 - Access permissions
 - Scalable and flexible
 - APIs
- Cheap 😊

Sie

Wie beantrage ich Urlaub?

GPT-4o-mini GPT-4o-mini

EBCONT-Chatbot

Um Urlaub zu beantragen, müssen Sie den Urlaubsantrag in SAP erfassen. Hier sind die Schritte, die Sie befolgen sollten:

Schritte zur Beantragung von Urlaub:

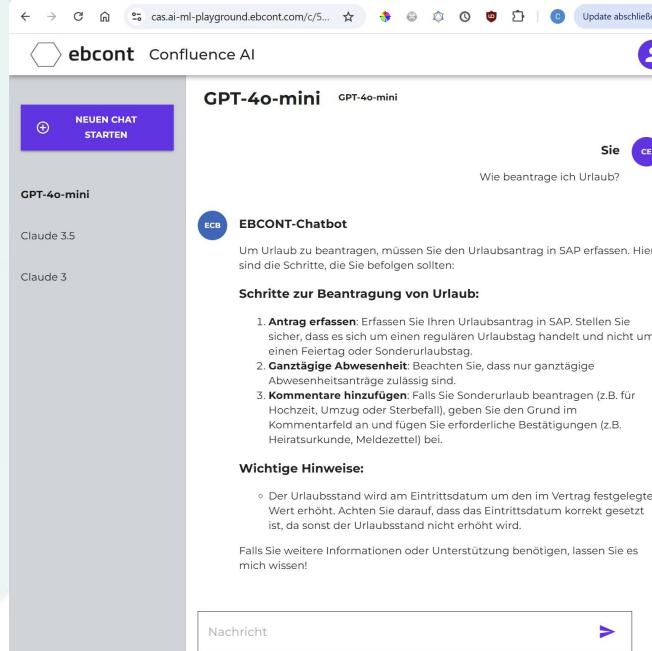
1. **Antrag erfassen:** Erfassen Sie Ihren Urlaubsantrag in SAP. Stellen Sie sicher, dass es sich um einen regulären Urlaubstag handelt und nicht um einen Feiertag oder Sonderurlaubstag.
2. **Ganztägige Abwesenheit:** Beachten Sie, dass nur ganztägige Abwesenheitsanträge zulässig sind.
3. **Kommentare hinzufügen:** Falls Sie Sonderurlaub beantragen (z.B. für Hochzeit, Umzug oder Sterbefall), geben Sie den Grund im Kommentarfeld an und fügen Sie erforderliche Bestätigungen (z.B. Heiratsurkunde, Meldezettel) bei.

Wichtige Hinweise:

- Der Urlaubsstand wird am Eintrittsdatum um den im Vertrag festgelegten Wert erhöht. Achten Sie darauf, dass das Eintrittsdatum korrekt gesetzt ist, da sonst der Urlaubsstand nicht erhöht wird.

Falls Sie weitere Informationen oder Unterstützung benötigen, lassen Sie es mich wissen!

Nachricht ➤





Why Elastic?

- » Widely used
- » Vector database
- » Machine learning/embedding
- » > 400 integrations

Enterprise-ready

- » Document Level Permissions
- » LDAP integration

The screenshot shows the Elastic Integrations page. At the top, there's a search bar labeled "Search Elastic". Below it, a navigation bar has "Integrations" selected. A sidebar on the left lists categories with counts: All categories (257), AWS (25), Azure (23), Cloud (38), Communications (3), Config management (2), Containers (13), Custom (22), Datastore (25), Elastic Stack (17), File storage (5), Geo (2), Google Cloud (3), Kubernetes (12), Language client (9), Message Queue (9), and Monitoring (5). A "Browse integrations" tab is active. The main area displays a grid of integration cards, each with a title, icon, and a brief description. Some cards have "Beta" status indicators.

Category	Count	Integration	Description
All categories	257	1Password Events Reporting	Collect events from 1Password Events API with Elastic Agent.
AWS	25	AbuseCH	Collect threat intelligence from AbuseCH API with Elastic Agent.
Azure	23	ActiveMQ Logs	Collect and parse logs from ActiveMQ instances with Filebeat.
Cloud	38	ActiveMQ Metrics	Collect metrics from ActiveMQ instances with Metricbeat.
Communications	3	Aerospike Metrics	Collect metrics from Aerospike servers with Metricbeat.
Config management	2	AlienVault OTX	Collect threat intelligence from AlienVault OTX with Elastic Agent.
Containers	13	Anomali	Collect threat intelligence from Anomali APIs with Elastic Agent.
Custom	22	Apache HTTP Server	Collect logs and metrics from Apache servers with Elastic Agent.
Datastore	25	Apache Tomcat	Collect and parse logs from Apache Tomcat servers with Elastic Agent.
Elastic Stack	17	API	Add search to your application with App Search's robust APIs.
File storage	5	APM	Collect performance metrics from your applications with Elastic APM.
Geo	2		
Google Cloud	3		
Kubernetes	12		
Language client	9		
Message Queue	9		
Monitoring	5		



Why Elastic?

Open & flexible

- » Embedding Models

Scalable

- » OnPremise, Cloud, Hybrid

Many additional features

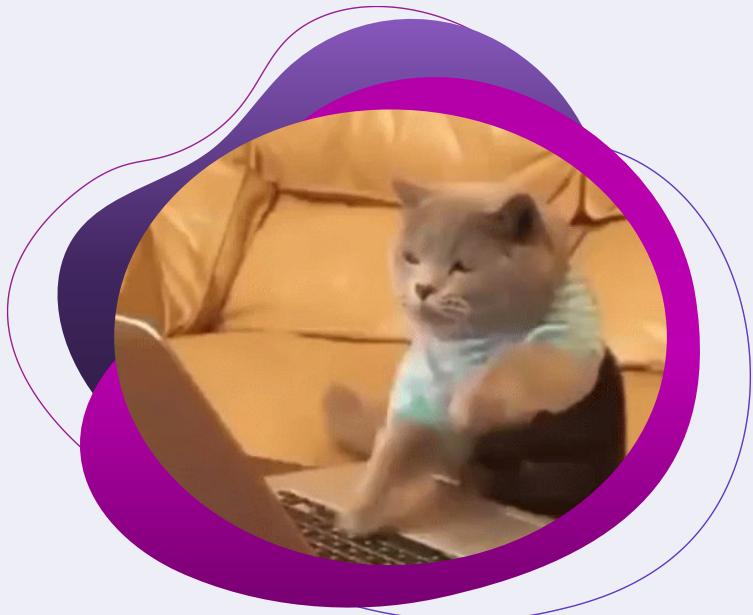
- » Semantic search, BM25,
etc.

The screenshot shows the Elastic Integrations interface. At the top, there's a navigation bar with the Elastic logo, a search bar labeled "Search Elastic", and a "View deployment details" button. Below the navigation is a section titled "Integrations" with the sub-instruction "Choose an integration to start collecting and analyzing your data." There are two tabs: "Browse integrations" (which is selected) and "Installed integrations". A sidebar on the left lists categories with counts: All categories (257), AWS (25), Azure (23), Cloud (38), Communications (3), Config management (2), Containers (13), Custom (22), Datastore (25), Elastic Stack (17), File storage (5), Geo (2), Google Cloud (3), Kubernetes (12), Language client (9), Message Queue (9), and Monitoring (5). To the right, there are several cards representing different integrations:

- 1Password Events Reporting: Collect events from 1Password Events API with Elastic Agent.
- AbuseCH: Collect threat intelligence from AbuseCH API with Elastic Agent.
- ActiveMQ Logs: Collect and parse logs from ActiveMQ instances with Filebeat.
- ActiveMQ Metrics: Collect metrics from ActiveMQ instances with Metricbeat.
- Aerospike Metrics: Collect metrics from Aerospike servers with Metricbeat.
- Akamai: Akamai Integration (Beta).
- AlienVault OTX: Collect threat intelligence from AlienVault OTX with Elastic Agent.
- Anomali: Collect threat intelligence from Anomali APIs with Elastic Agent.
- Apache HTTP Server: Collect logs and metrics from Apache servers with Elastic Agent.
- Apache Tomcat: Collect and parse logs from Apache Tomcat servers with Elastic Agent.
- API: Add search to your application with App Search's robust APIs.
- APM: Collect performance metrics from your applications with Elastic APM.

LLMs, Chats & RAG

Our Path – Challenges & Insights



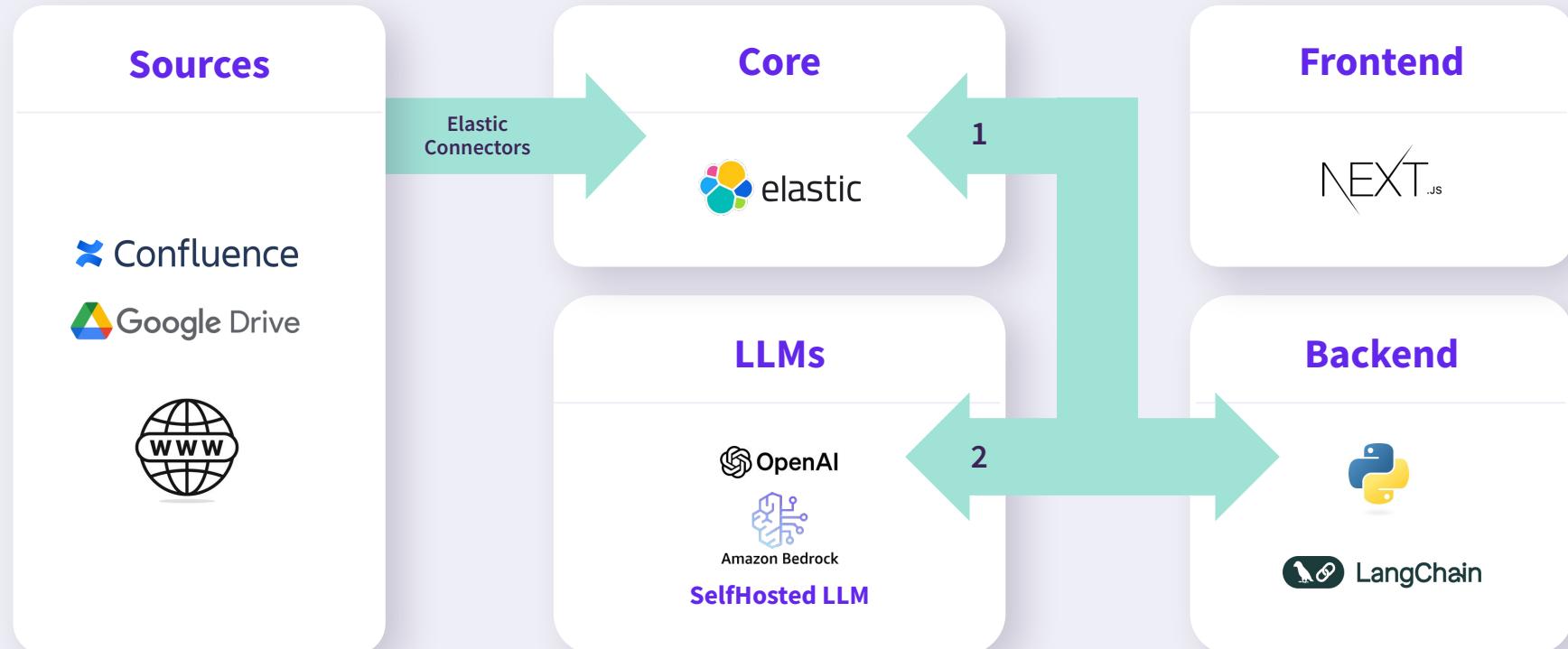
Tech-Stack

- » **Programming language: Python**
 - Standard for AI/Machine Learning applications
 - Wide range of libraries in the LLM area
- » **LLM processing: LangChain**
 - allows us to use LLMs in a structured form
 - supports the usage of agents for tool execution
- » **Data storage: Elastic**
 - Supports queries in vector databases (kNN-Search)
- » **Frontend/Backend:** React and FastAPI

The structure

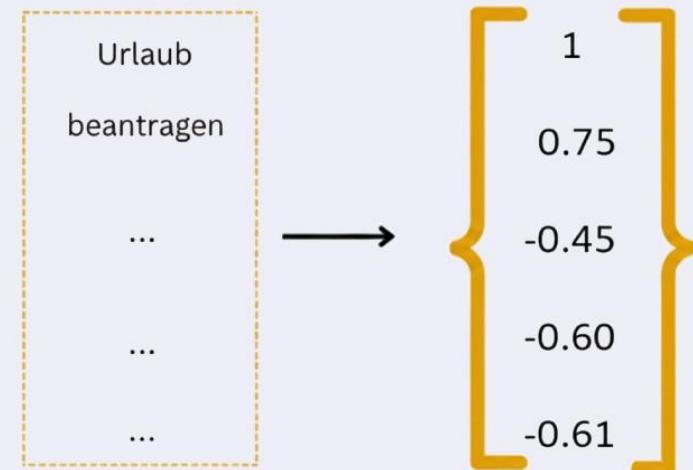


Architecture



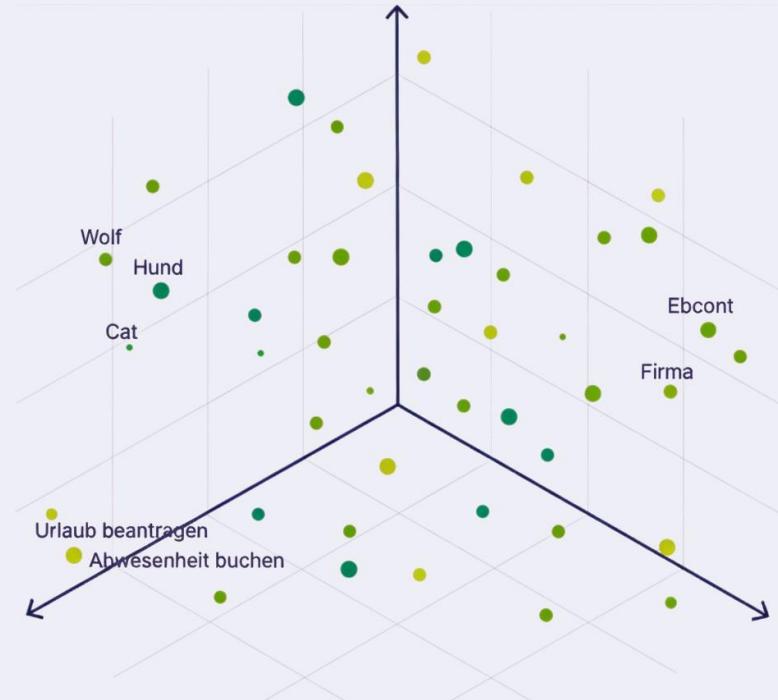
Vector Database

- » Sources are diverse:
Confluence, Google Drive, Homepage
- » Subdivision of source documents in
semantically related parts (“chunks”)
- » Generation of a vector representation from
each chunk, using an embedding model
- » Dedicated backend for chunking, embedding,
and storage in Elastic



RAG

- » Retrieves relevant context from Elastic vector database
- » Semantic search – “*Abwesenheit buchen*” vs. “*Urlaub beantragen*”
- » Combination with “traditional” BM25 search (matches in text)



Prompt Engineering / System Prompt

- » Instructions on what tasks the LLM should perform
- » Specification of requirements as detailed as possible
 - prevention of hallucinations / use of own knowledge
- » Description of available tools
- » Minor adjustments depending on the model (OpenAI, Claude, ...) may be useful

LLM

1. Input: User input + system prompt
2. Tool execution by agent:
→ Query Elastic vector database
3. Generation of a response by LLM
in the context of the information obtained



Learnings

- » Challenges in prompt engineering:
 - Consistent style in responses
 - Preventing the use of “internal knowledge” (!)
- » Quality of response primarily determined by quality of context
 - relevant search results: kNN + BM25
 - intelligent chunking, metadata
- » Local models – insufficient performance
- » Confluence connector
- » Data quality / old data



Q & A

Thanks for your attention!



THANKS
FOR WATCHING