

A close-up photograph of a black and silver robotic arm reaching out from the left side of the frame. It is positioned to touch or just above the tip of a human hand that enters from the bottom right. The human hand has a tattoo on the forearm. The background is a plain, light color.

ChatAI Solution enriched with your own data

iits

Powered by
iits-consulting

Your Speaker



Victor Getz

CTO & Co-Founder

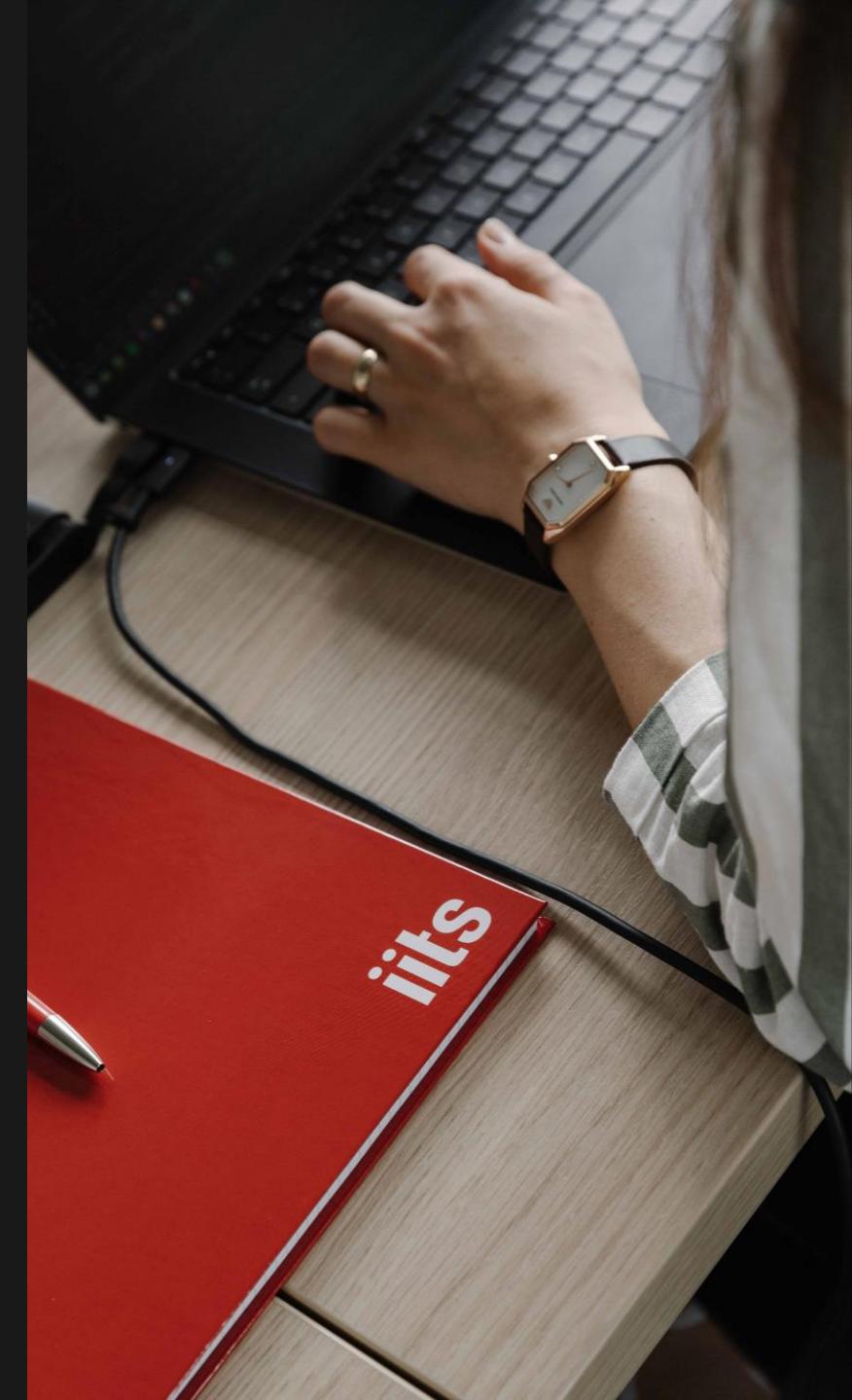
Victor.Getz@iits-consulting.de



Dr. Tim Delbrügger

Head of AI Department

Tim.Delbruegger@iits-consulting.de



iits-consulting

iits

Founded in 2019

We provide agile development teams that help
businesses automate their complex processes

130+ people are working for us



...

Agenda

- Use Cases and Success Stories
 - AI and LLMs
 - What are LLMs good for?
 - Selected Use Cases
 - Q&A
- How to setup Ollama with RAG
 - Architecture
 - Live coding demo
 - Q&A



iits

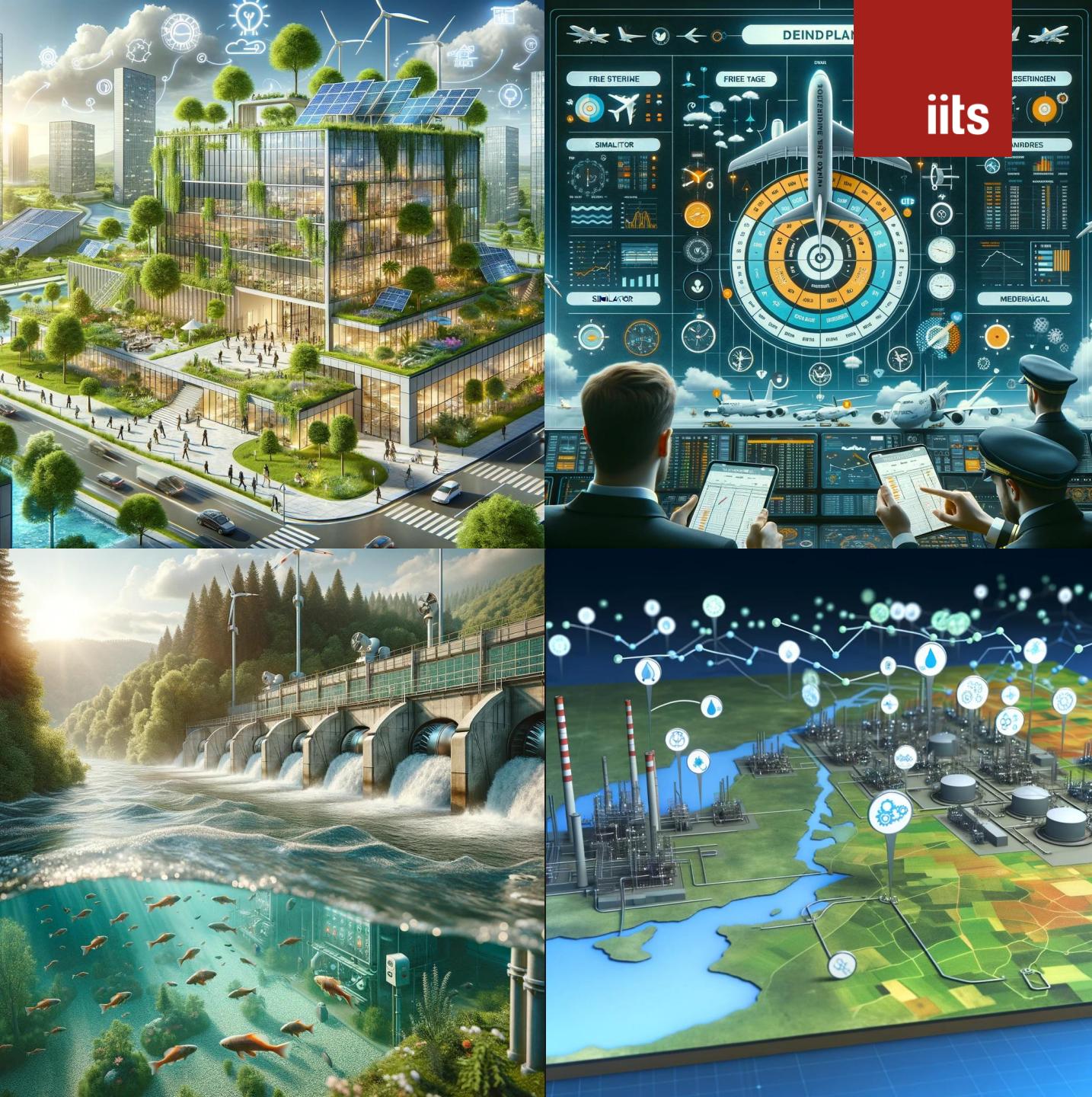
“

„It is difficult to think of a major industry that AI will not transform. This includes healthcare, education, transportation, retail, communications, and agriculture. There are surprisingly clear paths for AI to make a big difference in all of these industries.“

– Andrew Ng



AI is more than LLM's



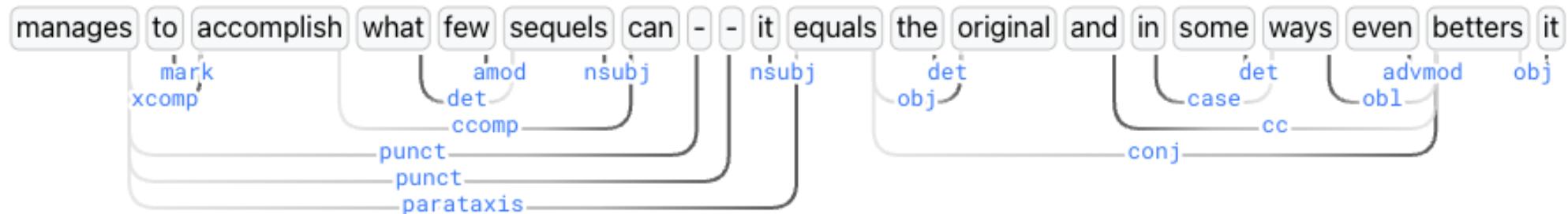
Large Language Models: Birds-Eye View

Maximum size of the prompt
is called „context window“



The best thing is: LLMs come ready to go – no training required

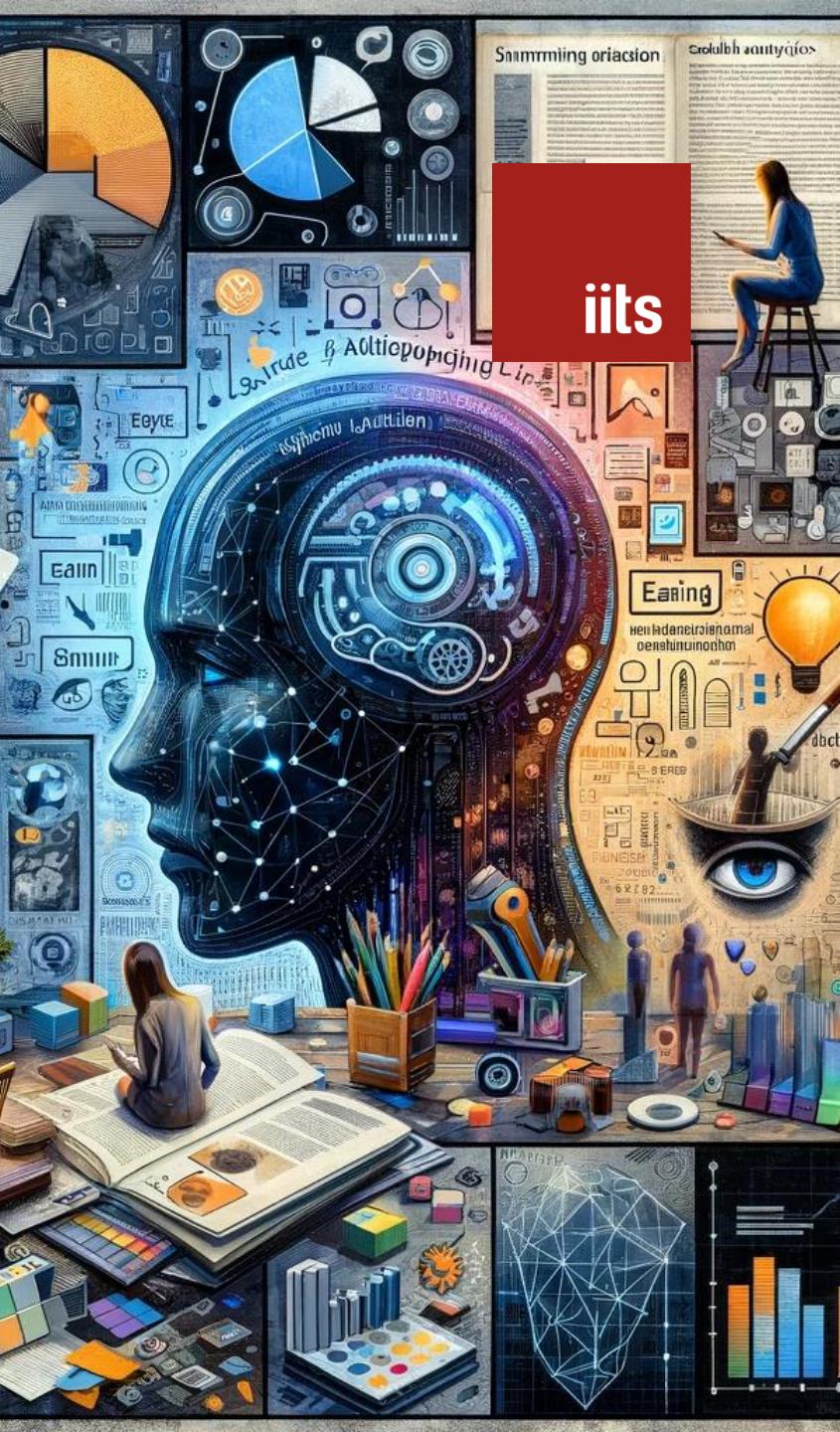
Attention Mechanism – explained



Source: Dodrio

What are LLMs good for?

- Textual tasks that a human can do with just a short explanation and examples
- Understand and extract Information from texts
- Summarize and compare texts
- Mood, sentiment and general style detection
- Generate good-looking content
- Style adjustment
- Code Suggestions



Don't use LLMs directly for...



Calculations / Equation solving

It is not a calculator!



Medical Advice

It is not a physician!



Decision Making

LLMs have been trained
only on language, not on
outcomes in the real world
and their implications.

=> Use optimization
algorithms with carefully
crafted models and expert
knowledge



Database queries

An LLM is not a SQL
engine.

**It will deceive you with
good-looking but wrong
results!**



Forecasting

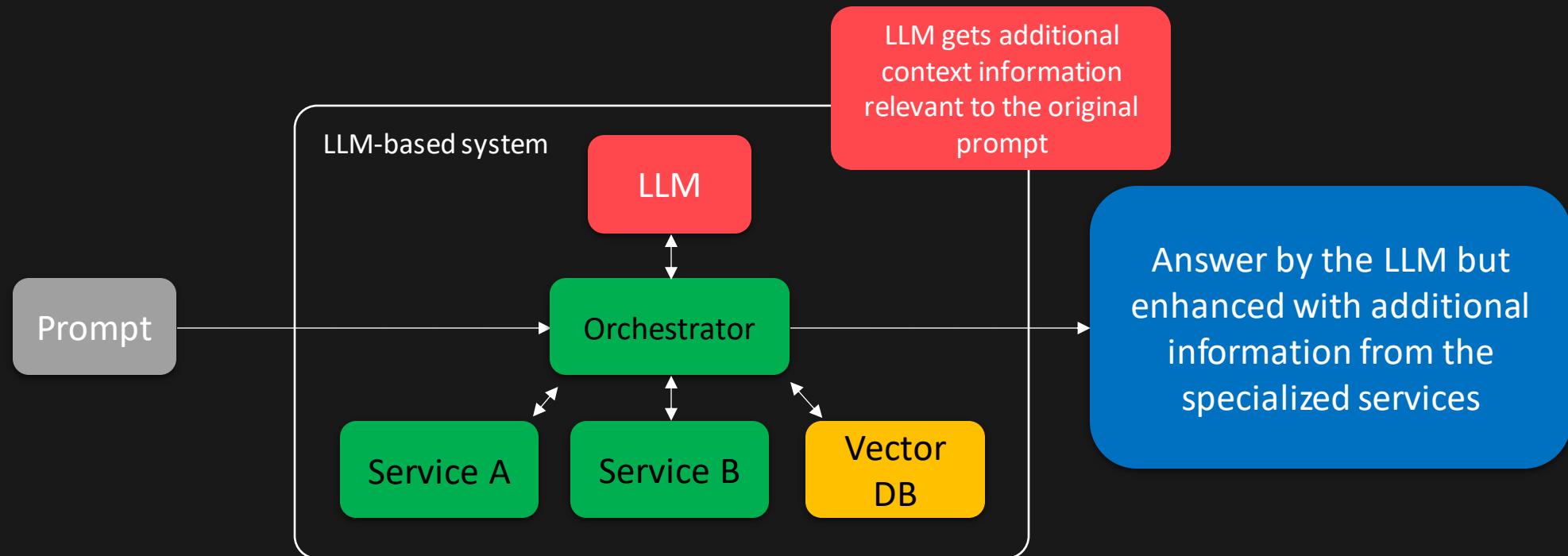
LLMs are no crystal ball,
they don't know the future
and are not trained to do
data-based predictions.

=> Use regression models,
maybe in combination with
simulations

**BUT we can combine LLMs with specialized
systems for such tasks!**

Augmented Generation

- Mitigate limitations of LLMs with a system around the LLM



Use Case: Broadband Tenders

- In Germany, each local authority (city, district, etc.) must create their own broadband expansion tenders to get funded.
- Each tender is a collection of documents with references to funding programs
- Tenders are screened for **key facts** to prepare an evaluation and further processes
- One good employee needs **multiple hours** to extract the key facts of **one tender**

Result – Use Case: Broadband Tenders

- Tender documents are uploaded
- ~ 40 Key Facts are extracted automatically,
no chat required
- Manual correction possible
- The system can show relevant
sources for each key fact

The screenshot displays two main sections of the KI Assistent application:

- Dokumentenmanagement:** A section titled "Ausschreibungsunterlagen" (Bidding documents) with a large dashed area for file upload. It includes a cloud icon with an upward arrow and the text "Datei hierhin ziehen oder klicken". Below it, a list of supported formats is shown: Unterstütztes Format: .pdf. The list includes:
 - Microsoft Word Dokument (.doc, .docx)
 - Microsoft Word Dokument (.doc, .docx)
- Extraktion der Keyfacts:** A section titled "Extraktion der Keyfacts" (Extraction of Keyfacts) containing a form with various fields and their descriptions. Fields include:
 - * Name der Ausschreibung: [Input field]
 - Name der Gebietskörperschaft: [Input field]
 - Anzahl Lose: [Input field]
 - Förderprogramm: Das Bundesförderprogramm zur Unterstützung des Gigabitausbaus in Deutschland ist in den Dokumenten genannt. [Input field]
 - Fördermodell: Das Fördermodell, das hier vorliegt, ist die "Richtlinie zur Förderung des Ausbaus von gigabitfähigen Breitbandnetzen" beim L [Input field]
 - Abgabeformat: Die Angebote sollen ausschließlich elektronisch über das Vergabeportal <https://eu.eu-supply.com> eingereicht werden. Die ele [Input field]
 - eVergabe-Portal: Die Angebote sollen ausschließlich über das Vergabeportal <https://eu.eu-supply.com> eingereicht werden. [Input field]
 - Abgabedatum: Die Angebote müssen spätestens am 10.03.2023 abgegeben werden. [Input field]
 - Abgabezeit: Die Angebote müssen spätestens um 12:00 Uhr abgegeben werden. [Input field]Buttons at the top right of this section include "Automatisches Ausfüllen" and "Exportieren".

Result – Use Case: Broadband Tenders

Unterstütztes Format: .pdf und .docx

Relevante Textpassagen für Anzahl Lose X

F002_CSX 23 - Verfahrensinformationen.pdf / Seite 4

Die Ausschreibung erfolgt technologieneutral. **Das Projektgebiet ist entsprechend der Anzahl der Förderantragsgegenstände in 4 Lose untergliedert:** Los 1 832.5/3-19 09NI300066 Landkreis Holzminden (weiße Flecken und Schulen) Los 2 832.5/3-20 11NI300481 Stadt Holzminden und Flecken Bevern (Gewerbe) Los 3 832.5/3-20 11NI300482 Stadt Stadtoldendorf, Gemeinden Deensen und Arholzen (Gewerbe) Los 4 832.5/3-20 11NI300641 Stadt Eschershausen, Gemeinde und Hehlen und Flecken Lauenförde (Gewerbe) Dienstleistungskategorie CPV-Code Bezeichnung Zusatzteil Haupt- gegen stand64210000-1 Ergänzende Gegenstände32412000-4 32571000-6 Leistungsorte NUTS-Code DE926 Hauptleistungsort Ort 37603 Holzminden Ausführungsfristen Dauer (ab Auftragsvergabe) 84 Monat(e)

* Name der Ausschreibung

Name der Gebietskörperschaft (L) (C) (I) (Q)

Anzahl Lose (L) (C) (I) (Q)

Show information sources in the documents

Use Case: POC Logistics

- For another client, we had to build a POC for an **internal knowledge base** about logistics.
- Example:
 - The employee wants to know the **DHL standard business hours** based on terms & conditions from this document [here](#)

Result – Demo: Logistics

 New Chat

mistral:7b



Set as default



You

What are the standard working hours for dhl? based on the terms&conditions

< 1 / 1 >



Ollama mistral:7b

Based on the question i chose the custom data input: Logistic_documents, reason=The question asks for DHL's standard working hours based on the terms & conditions.

The standard working hours for DHL, as specified in the provided terms and conditions, are from Monday to Friday, 08:00 – 18:00 hrs. However, upon specific Customer request and when the booking is made before 12:00pm, shipments can be collected on the day of booking. The same day pick-up areas and respective latest booking times are determined by each DHL Freight branch. Recipients must accept delivery of the shipment during these standard working hours, immediately after truck arrival at their premises.



**ChatGPT changed
our way of working
and our lives**

But it has its flaws

- ChatGPT has a **problem** with data privacy
- It is not possible to **add private data** from **internal sources** on a big scale

OpenSource LLM & iits-consulting

- That's where OpenSource LLM and iits-consulting come into play
- We offer similar the same functionality as ChatGPT (but with your data)
- We will setup a private OpenSource LLM Tool and add your data
- Your data is stored secure, encrypted and private
- Your data is not shared with US Companies
- Customizable data access rules based on your company account

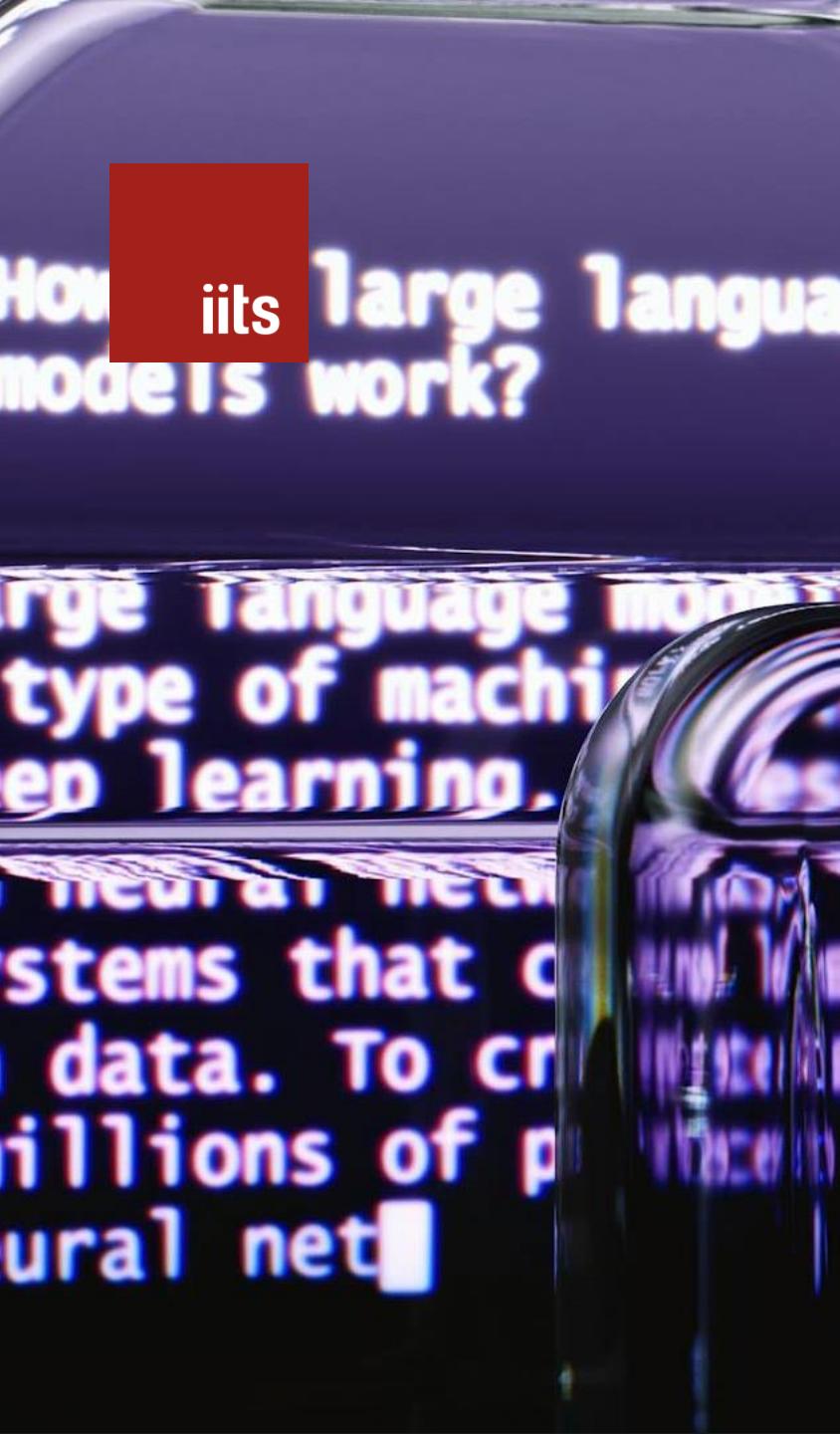
Data Ingestion

- Our setup builds on top of **Airbyte** and offer **300+ pre-built connectors**
- Add **your data** on your own through the UI; **no software development needed**
- If you need a **special connector**, we can implement it for you
- PDF, Word, Excel, CSV, Databases are **supported by default**

Select the type of source you want to connect

The screenshot shows a user interface for selecting data sources. At the top right is a red button labeled "iits". Below it is a search bar with the placeholder "Search" and three filter checkboxes: "Generally Available" (checked), "Beta" (checked), and "Alpha" (unchecked). A list of connectors is displayed in rows, each with a logo, name, and status indicator (e.g., BETA). The connectors listed are: Airtable, Amplitude, Chargebee, Confluence, Facebook Pages, Freshsales, Google Ads, and Google Search Console.

Connector	Status
Airtable	
Amplitude	
Chargebee	
Confluence	BETA
Facebook Pages	BETA
Freshsales	BETA
Google Ads	
Google Search Console	

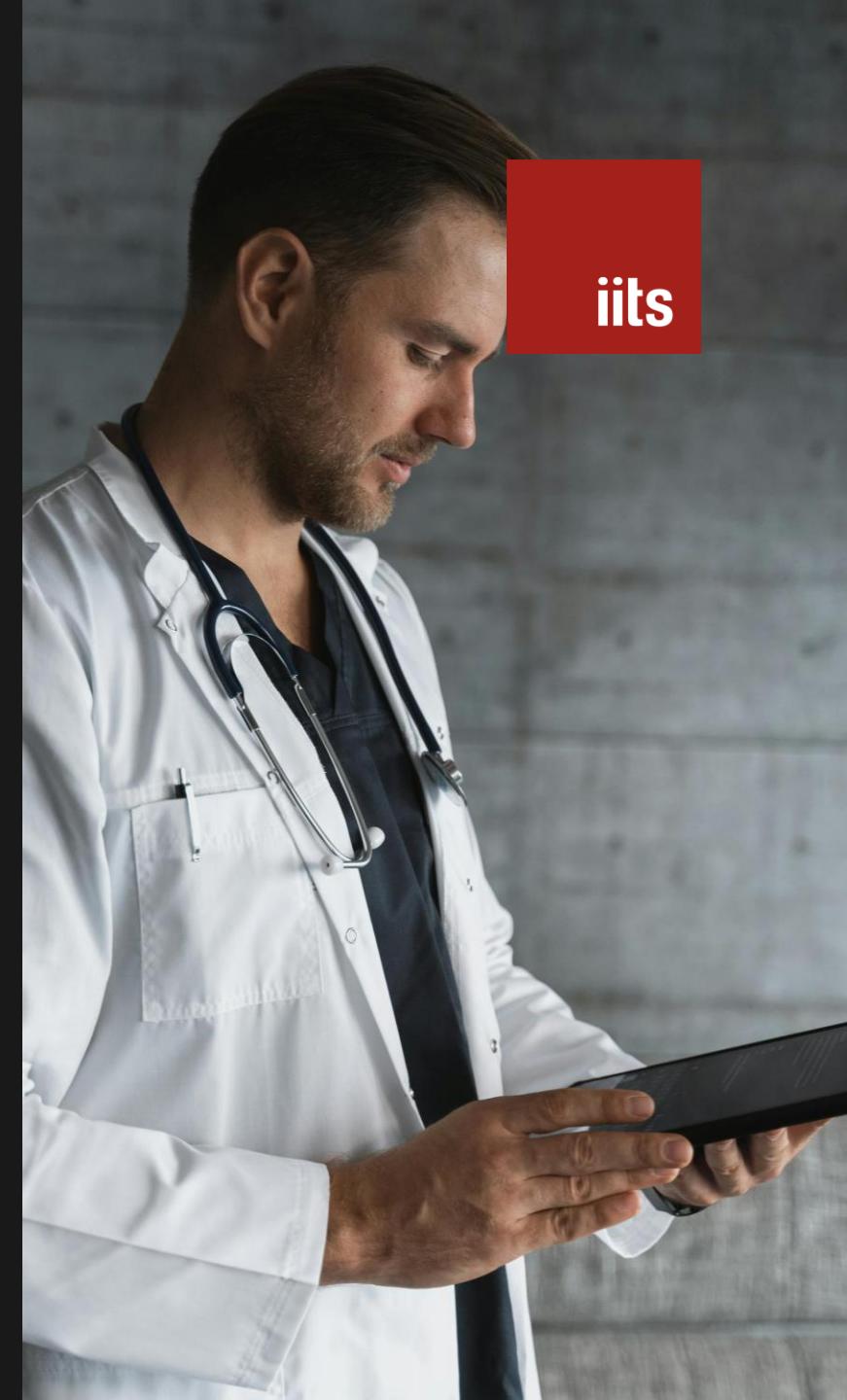


iiits AI middleware

- It does not load **all the data**, only what is **relevant** to the question
- It selects which **data** needs to be **analyzed** based on the question
- You can choose between **multiple LLM models** and decide which one would be better **suited** for your use case

Demo Health

- For one of our **clients**, we had to build an application that should allow doctors to access **patient information more quickly**
- Example:
 - Was Patient Pamela Rogers **diagnosed** with hypertension?
 - The original document can be found [here](#)



Result - Demo: Health

iits

The screenshot shows the Ollama Web UI interface. At the top left, it says "Ollama Web UI" and "mistral:7b (3.8GB)". Below that is a "Set as default" button. In the center, there's a large teal circular icon with the text "iits". Below the icon, the text "How can I help you today?" is displayed. At the bottom, there are four buttons: "Tell me a fun fact about the Roman Empire", "Show me a code snippet of a website's sticky header", "Help me study vocabulary for a college entrance exam", and "Give me ideas for what to do with my kids' art". At the very bottom, there's a text input field containing "How old is my Patient Pamela Rogers?", a microphone icon, and an upward arrow icon. A footer note at the bottom reads "LLMs can make mistakes. Verify important information."

History of Present Illness

This is the first admission for this 56 year old woman, who states she was in her usual state or good health until one week prior to admission. At that time she noticed the abrupt onset (over a few seconds to a minute) of chest pain which she describes as dull and aching in character. The pain began in the left para-sternal area and radiated up to her neck. The first episode of pain one week ago occurred when she was working in her garden in the middle of the day. She states she had been working for approximately 45 minutes and began to feel tired before the onset of the pain. Her discomfort was accompanied by shortness of breath, but no sweating, nausea, or vomiting. The pain lasted approximately 5 to 10 minutes and resolved when she went inside and rested in a cool area.

Since that initial pain one week ago she has had 2 additional episodes of pain, similar in quality and location to the first episode. Three days ago she had a 15 minute episode of pain while walking her dog, which resolved with rest. This evening she had an episode of pain awaken her from sleep, lasting 30 minutes, which prompted her visit to the Emergency Department. At no time has she attempted any specific measures to relieve her pain, other than rest. She describes no other associated symptoms during these episodes of pain, including dizziness, or palpitations. She becomes short of breath during these episodes but describes no other exertional dyspnea, orthopnea, or paroxysmal nocturnal dyspnea. No change in the pain with movement, no association with food, no GERD sx, no palpable pain. She has never been told she has heart problems, never had any chest pains before, does not have claudication. She was diagnosed with HTN 3 years ago,

She does not smoke nor does she have diabetes. She was diagnosed with hypertension 3 years ago and had a TAH with BSO 6 years ago. She is not on hormone replacement

Convey the acute or establish a chronology

onset
character
location
radiation
circumstances; exac

associated symptom
duration
resolution; alleviating

Describe the natural onset

Change or new circu
New duration
Reason she come in
What has patient tri

Relevant positive an

Review of systems fo

Relevant risk factor/e

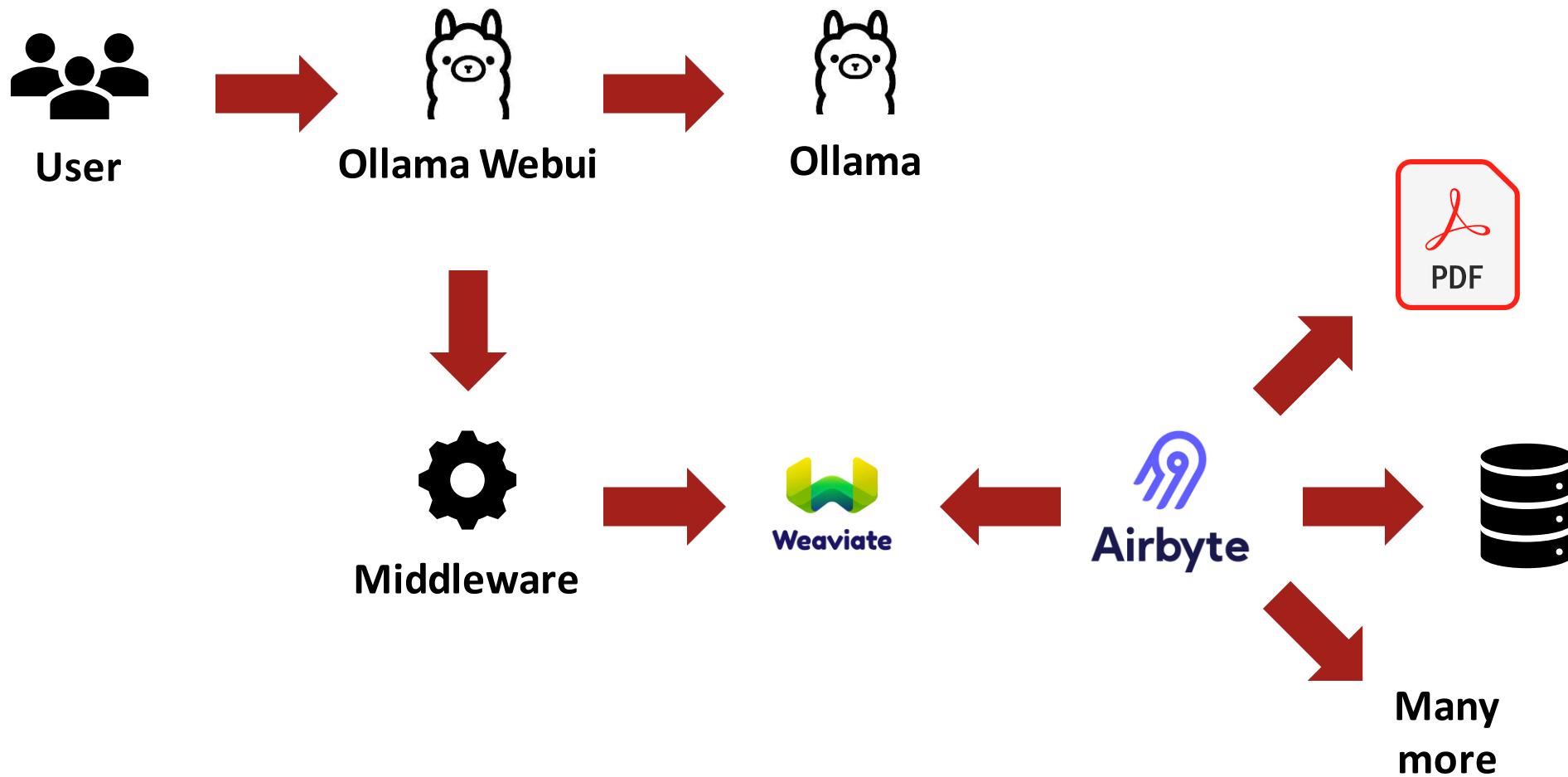
Demo



Time
for
Q&A

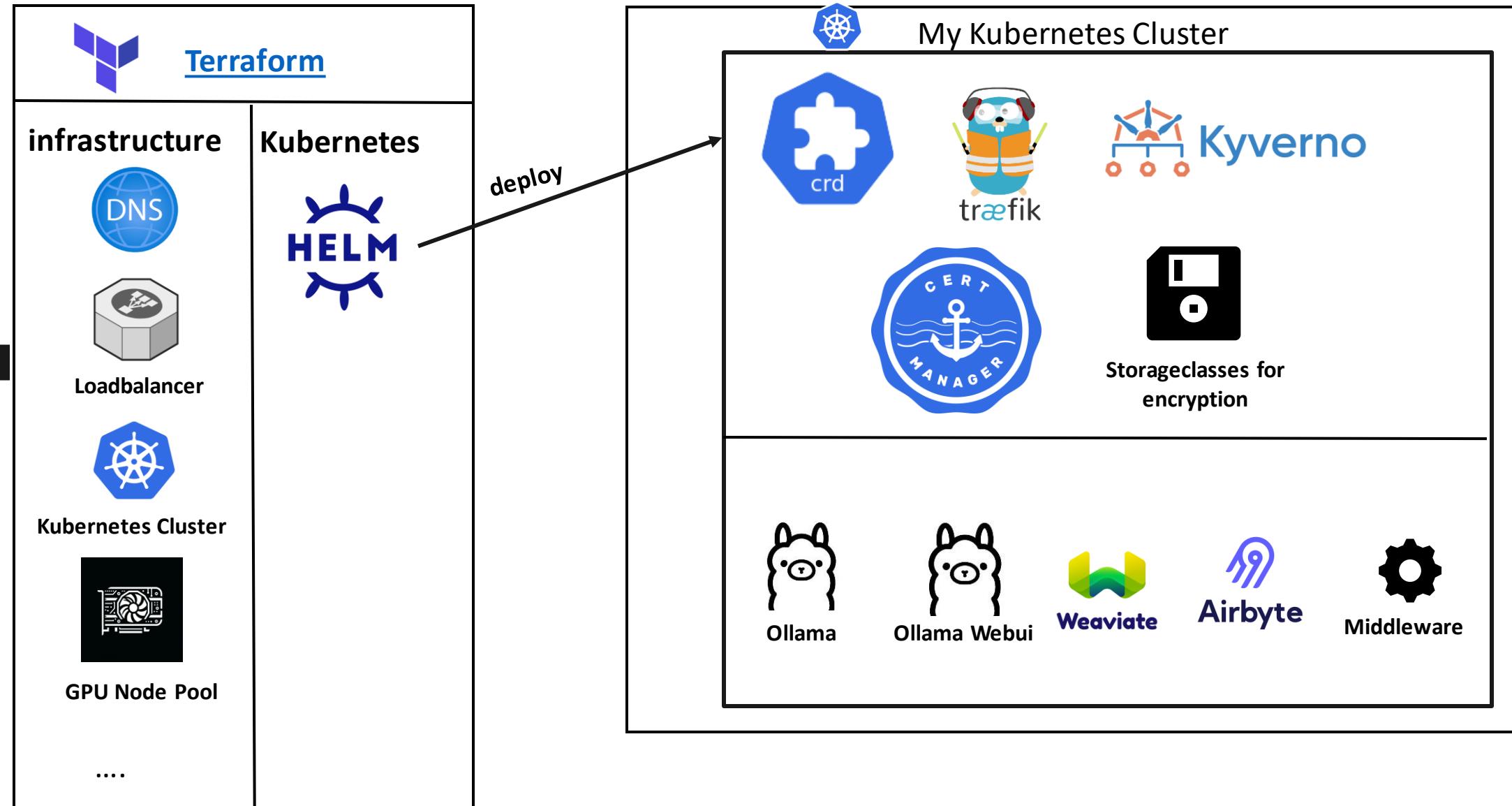


Ollama with RAG: service architecture



Ollama with RAG: infrastructure architecture

iits



Time to Code

```
8
9 require 'rspec/rails'
10
11 require 'capybara/rspec'
12 require 'capybara/rails'
13
14 Capybara.javascript_driver = :webkit
15 Category.delete_all; Category.create!
16 Shoulda::Matchers.configure do |config|
17   config.integrate do |with|
18     with.test_framework :rspec
19     with.library :rails
20   end
21 end
22
23 # Add additional requires below this line if you have additional dependencies
24
25 # Requires supporting ruby files with custom matchers and helpers
26 # in spec/support/ and its subdirectories. This file is now loaded
27 # in _spec.rb will both be required automatically.
28 # run twice. It is recommended that you do not name this file spec_helper,
29 # as it can interfere with devise's own spec_helper.
30
31 # run as spec files by default. This means you don't need to use
32 # `context` or `it` to run specs, but individual steps in scenarios
33 # still need to be run with `step` or `then`.
```

Time
for
Q&A



Part II

iits

Innovation
as a Team



**Let's create a success story
together!**

Your Contact Person:

Victor Getz

team-sales@iits-consulting.de