smartlearn
virtual learning. real success.

AIB_2026_SERIE_0_DE_CHALLENGE-01_POC VERIFIZIEREN

ICT-Berufsbildung Schweiz

Inhaltsverzeichnis

Ausgangslage **3**

Auftrag **6**

Frage F1 6

Frage F2 7

Frage F3 7

Frage F4 7

Frage F5 8

Frage F6 9

Musterlösung **9**

Verortung 9

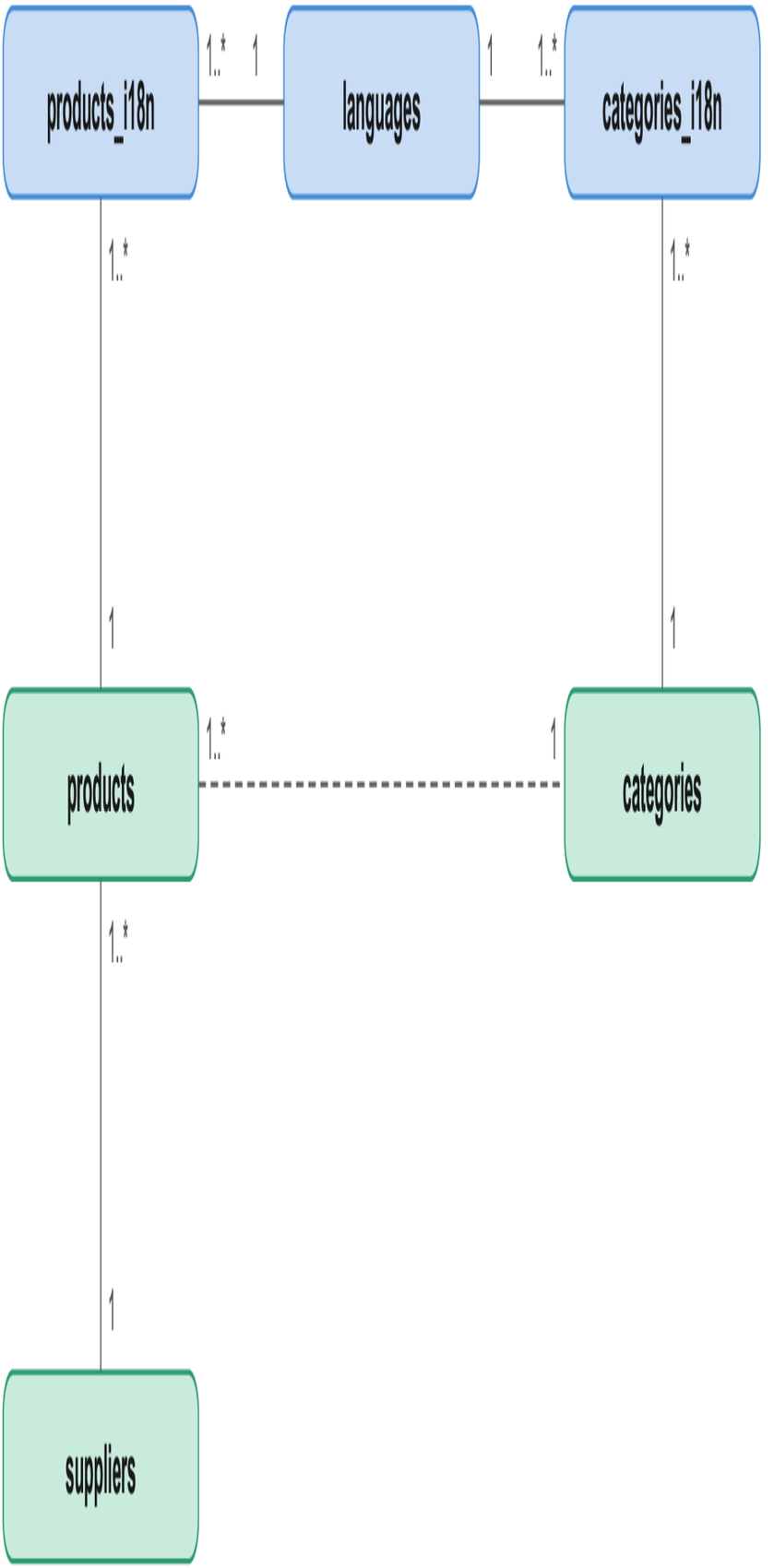
Ausgangslage

Ein Unternehmen im Grosshandel für Früchte und Gemüse stellt bereits heute eine REST-API zur Verfügung, über welche Daten zum Sortiment abgefragt werden können. Aktuell wird untersucht, ob diese Informationen künftig auch über natürlichsprachliche Textabfragen zugänglich gemacht werden können.

Ein vielversprechender Ansatz dafür ist RAG (Retrieval-Augmented Generation): Dabei wird ein LLM gezielt mit Daten aus einer kontrollierten Quelle angereichert. Das LLM übernimmt ausschliesslich die Sprachverarbeitung und die Datenbasis bleibt die bestehende, strukturierte Datenbank des Unternehmens.

Zur Evaluation dieses Ansatzes wurde ein Proof of Concept (PoC) entwickelt, in welchem das Sprachmodell `phi3:mini` zum Einsatz kommt. Das Modell wurde wegen seiner geringen Grösse und der Möglichkeit des lokalen Betriebs auf einem Versuchsserver ausgewählt.

Im PoC wurden exemplarisch 4 Endpunkte der bestehenden REST-API als Datenquelle genutzt, welche strukturierte JSON-Antworten zurückgeben. Die vollständige API-Dokumentation ist über die Swagger UI der Testumgebung zugänglich. Die Datenbank des PoC umfasst folgende Tabellen und Beziehungen:


products_i18n
languages
categories_i18n
products
categories
suppliers
1..*
1

Für die LLM-Integration wurden zwei unterschiedliche Ansätze implementiert und als Endpunkte im Swagger UI zur Verfügung gestellt:

- **`POST /llm/v1`**: Das LLM kennt die SQL-Abfragen, die auch in den REST-Endpunkten verwendet werden. Aufgrund eines textuellen Prompts wählt das LLM eine bekannte SQL-Abfrage aus. Dieser Ansatz ist deterministisch, weil er fest codierte SQL-Abfragen verwendet.
- **`POST /llm/v2`**: Das LLM kennt die Struktur der Datenbank mit den Tabellen und Bezeihungen. Aufgrund eines textuellen Prompts generiert das LLM die SQL-Abfrage selbst. Dieser Ansatz ist flexibler, aber fehleranfälliger.

Swagger UI ist eine browserbasierte Benutzeroberfläche zur Beschreibung und zum Testen von REST-APIs. Es ist auf dem Testserver `ubu-srv-exam-01` mit IPv4-Adresse `192.168.110.20` verfügbar und kann im Browser eines Testclients unter **`http://192.168.110.20/docs`** aufgerufen werden. In der Testumgebung steht wahlweise ein Client mit Windows 11 (`win-cli-exam-01`) und ein Ubuntu Client (`ubu-cli-exam-01`) zur Verfügung.

Hinweis: Beachten Sie, dass die Verarbeitung der Abfragen auf die LLM-Endpunkte wegen den beschränkten Ressourcen des Servers durchaus 30 bis 90 Sekunden dauern können. Bei Ausbleiben einer Antwort nach 180 Sekunden bricht der Endpunkt mit einem Timeout-Fehler ab.

Umgebung und Zugangsdaten

Funktion	Host	IP-Adresse	OS	Applikationen & Werkzeuge	Benutzer	Passwort
Testserver	ubu-srv-exam-01	192.168.110.20/24	Ubuntu 24.04.2 LTS	Swagger UI auf Webserver (Uvicorn) Ollama mit phi3:mini MySQL-Datenbank mit Sortimentsdaten	kein Zugriff	
Testclient	ubu-cli-exam-01	192.168.110.yyy/24	Ubuntu 24.04.2 LTS	Browser Firefox	vmadmin	sml12345
Testclient	win-cli-exam-01	192.168.110.101/24	Windows 11 Client	Browser Edge	vmadmin	sml12345

Auftrag

Rufen Sie die Swagger UI im Browser eines Testclients unter `http://192.168.110.20/docs` auf. Testen Sie die verfügbaren Endpunkte und beantworten Sie die Fragen F1 bis F6 in den entsprechenden Antwortfeldern.

Frage F1

Wie viele unterschiedliche Kategorien von Produkten gibt es in der Datenbasis? Tragen Sie die Anzahl der Kategorien ins Antwortfeld ein.

Anzahl unterschiedlicher Kategorien

Frage F2

Welches ist das teuerste Produkt im Sortiment? Tragen Sie den Namen des Produkts ins Antwortfeld ein.

Name des teuersten Produkts

Frage F3

Welche Kategorie hat die meisten Produkte? Tragen Sie den Namen der Kategorie ins Antwortfeld ein.

Name der Kategorie

Frage F4

Die Antworten von einem LLM sind bekanntlich nicht deterministisch.

In einem Versuch antwortete der Endpunkt **`POST /llm/v1`** auf das Prompt `"Show me all the products in the berries category."` mit der

JSON-Response, die in folgender Datei gespeichert wurde:

- q4-response.json (https://ict-bb-ch.smartlearn.cloud/media/b87fadc4549b9d7b3a55ae.json)

Analysieren Sie die gespeicherte Antwort und kreuzen Sie die korrekte Aussage an.

Single-Choice

- [ ] Der richtige JSON-Endpunkt wurde gewählt und die Produkte wurden korrekt zurückgegeben. (HTTP 200).
- [ ] Der richtige JSON-Endpunkt wurde gewählt, aber die zurückgegebenen Produkte gehören einer anderen Kategorie. (HTTP 200)
- [ ] Der richtige JSON-Endpunkt wurde gewählt, aber die Abfrage schlug wegen eines fehlenden Parameters fehl. (HTTP 400)
- [ ] Der richtige JSON-Endpunkt wurde gewählt, aber das LLM hat Produktnamen halluziniert, die nicht in der Datenbasis existieren. (HTTP 200)
- [ ] Ein falscher JSON-Endpunkt wurde gewählt und die zurückgegebenen Daten sind unvollständig oder unpassend. (HTTP 200)
- [ ] Die Abfrage konnte keiner bekannten Query zugeordnet werden. (HTTP 400)

Frage F5

Welche Information fehlt im Prompt `"Show me all the products in the berries category."`, damit der Endpunkt `POST /llm/v1` die Abfrage zuverlässig ausführen kann? Begründen Sie kurz.

Tipp: Testen Sie Ihre Antwort direkt am Endpunkt `POST /llm/v1` im Swagger UI.

Frage F6

In einem Versuch antwortete der Endpunkt **`POST /llm/v2`** auf das Prompt `"Show me the details of bananas."` mit der JSON-Response, die in folgender Datei gespeichert wurde:

- q6-response.json (https://ict-bb-ch.smartlearn.cloud/media/c9c8d121f8de0ae8368338.json)

Analysieren Sie die gespeicherte Antwort. Nennen Sie die halluzinierte Tabelle und erklären Sie anhand des Datenbankschemas in der Ausgangslage, weshalb diese Tabelle nicht existiert.

Musterlösung

Verortung

Modul: 619 - KI-basierte Lösungsansätze auf Machbarkeit prüfen

Handlungsziel: HZ5: Bewertet KI-basierte Lösungsansätze mit Hilfe der erhobenen Daten, Informationen und Erfahrungen anhand der definierten Kriterien zur

Machbarkeit sowie der allgemeinen KI-bezogenen Kriterien.

Handlungskompetenz:: C4 - KI-basierte Lösung auf Machbarkeit prüfen

Leistungskriterien: FAAIB können…

- Kriterien zur Bewertung von KI-basierten Lösungen unter Berücksichtigung der technischen, [*organisatorischen, rechtlichen und gesellschaftlichen*] Perspektive festlegen.
- das für die Prüfung geeignete Vorgehen, sowie die dafür notwendigen Werkzeuge [*zusammen mit Fachpersonen*] definieren (z.B. Entwicklung von Prototypen, Durchführung von Labor- oder Feldtests).
- die Machbarkeit einer KI-basierten Lösung [*unter Einbezug von relevanten Anspruchsgruppen und Fachpersonen*] gemäss den festgelegten Kriterien untersuchen und bewerten.
- Anzeichen für mögliche Bias-Effekte in Daten oder Modellen erkennen und in Zusammenarbeit mit Fachpersonen deren Bedeutung für die Lösung einschätzen.