smartlearn
virtual learning. real success.

AIB_2026_SERIE_0_FR_CHALLENGE-01_VÉRIFIER LE POC

ICT-Formation professionnelle Suisse

Table des matières

Situation initiale **3**

Mandat **6**

Question Q1 6

Question Q2 7

Question Q3 7

Question Q4 7

Question Q5 8

Question Q6 9

Corrigé type **9**

Documents de référence 9

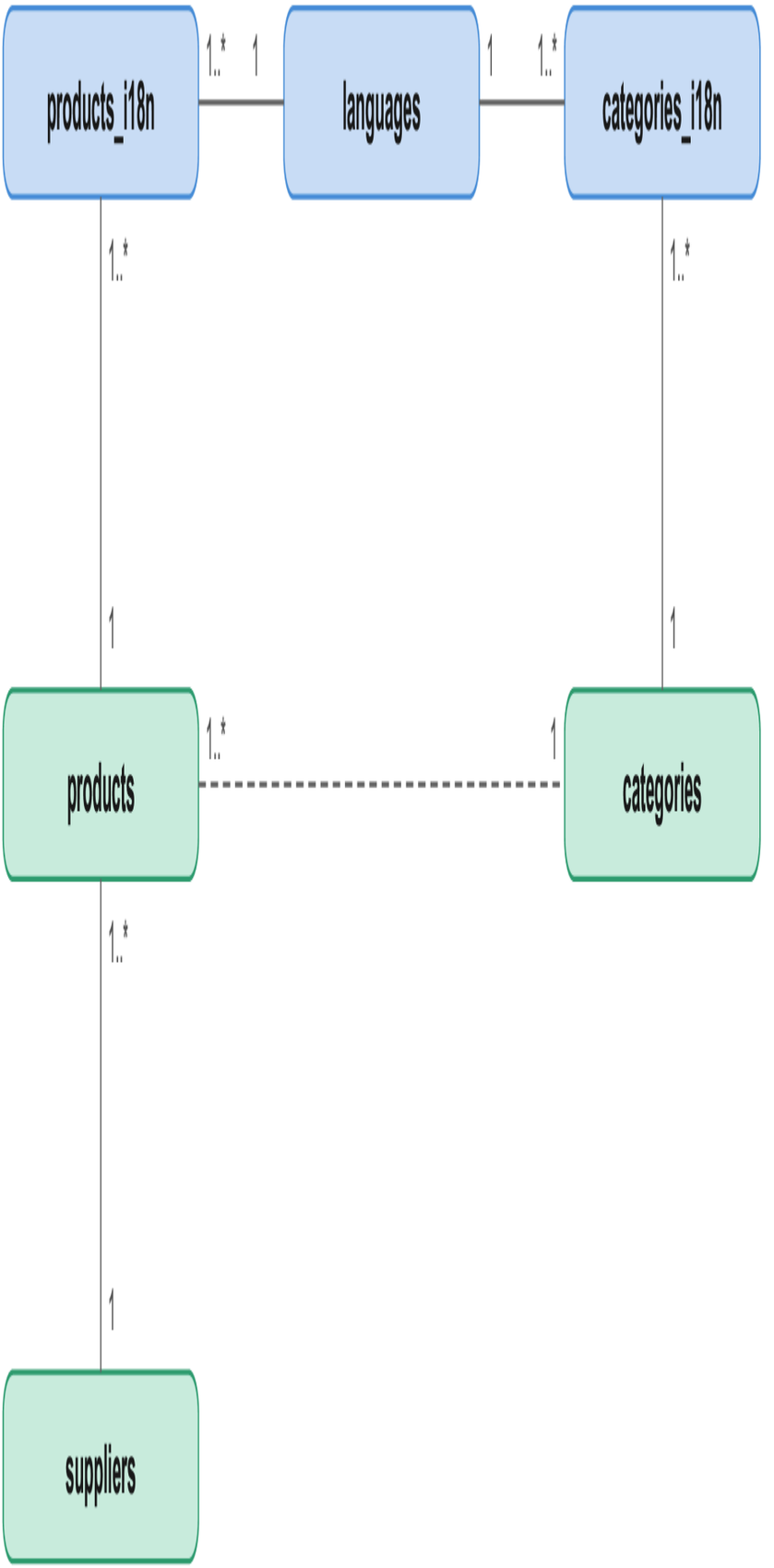
Situation initiale

Une entreprise de commerce de gros de fruits et légumes met déjà à disposition une API REST permettant de consulter des données relatives à son assortiment. Elle examine actuellement si, à l'avenir, les données pourraient également être rendues accessibles via des requêtes textuelles en langage naturel.

Une approche prometteuse à cet effet est le RAG (Retrieval-Augmented Generation). Dans ce cadre, un LLM est enrichi de manière ciblée avec des données provenant d'une source contrôlée. Le LLM se charge exclusivement du traitement du langage, tandis que la base de données structurée existante de l'entreprise demeure la source de données.

Afin d'évaluer cette approche, un Proof of Concept (PoC) a été développé, dans lequel le modèle de langage `phi3:mini` est utilisé. Ce modèle a été choisi en raison de sa faible taille et de la possibilité de l'exécuter localement sur un serveur de test.

Dans le PoC, quatre points de terminaison (*endpoints*) de l'API REST existante ont été utilisés à titre d'exemple comme source de données. Ces endpoints renvoient des réponses JSON structurées. La documentation complète de l'API est accessible via l'interface Swagger UI de l'environnement de test. La base de données du PoC comprend les tables et relations suivantes:


products_i18n
languages
categories_i18n
products
categories
suppliers
1..*
1

Pour l’intégration du LLM, deux approches différentes ont été implémentées et mises à disposition sous forme d’endpoints dans l’interface Swagger UI:

- **`POST /llm/v1`**: le LLM connaît les requêtes SQL, qui sont aussi utilisées dans les endpoints REST. À partir d’un prompt textuel, le LLM sélectionne une requête SQL prédéfinie. Cette approche est déterministe, car elle repose sur des requêtes SQL codées en dur.
- **`POST /llm/v2`**: le LLM connaît la structure de la base de données avec ses tables et ses relations. À partir d’un prompt textuel, le LLM génère lui-même la requête SQL. Cette approche est plus flexible, mais plus sujette aux erreurs.

Swagger UI est une interface utilisateur basée sur un navigateur permettant de décrire et de tester des API REST. Elle est disponible sur le serveur de test `ubu-srv-exam-01` à l’adresse `192.168.110.20` et accessible via le navigateur d’un client de test à l’adresse **`http://192.168.110.20/docs`** (http://192.168.110.20/docs). Dans l’environnement de test, un client équipé de Windows 11 (`win-cli-exam-01`) ou un client Ubuntu (`ubu-cli-exam-01`) sont mis à disposition au choix.

Remarque: *veuillez noter que le traitement des requêtes envoyées aux endpoints du LLM peut, en raison des ressources limitées du serveur, prendre entre 30 et 90 secondes. En l’absence d'une réponse après 180 secondes, l’endpoint s’interrompt avec une erreur de timeout.*

Environnement et données d’accès

Fonction	Hôte	Adresse IP	OS	Applications et outils	Utilisateur	Mot de passe
Testserver	`ubu-srv-`					

	exam-01	192.168.110.20/24	Ubuntu 24.04.2 LTS	Swagger UI sur serveur web (Uvicorn) Ollama avec phi3:mini Base de données MySQL avec données de l'assortiment	pas d'accès	
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	

Mandat

Ouvrez l'interface Swagger UI dans le navigateur d'un client de test à l'adresse `http://192.168.110.20/docs`. Testez les endpoints disponibles et répondez aux questions Q1 à Q6 dans les champs de réponse correspondants.

Question Q1

Combien y a-t-il de catégories de produits différentes dans la base de données? Saisissez le nombre de catégories dans le champ de réponse.

Nombre de catégories différentes

Question Q2

Quel est le produit le plus cher de l’assortiment? Saisissez le nom du produit dans le champ de réponse.

Nom du produit le plus cher

Question Q3

Quelle catégorie comporte le plus grand nombre de produits? Saisissez le nom de la catégorie dans le champ de réponse.

Nom de la catégorie

Question Q4

Comme chacun sait, les réponses d’un LLM ne sont pas déterministes.

Lors d’un test, l’endpoint **`POST /llm/v1`** a répondu au prompt `"Show me all the products in the berries category."` par la réponse JSON

enregistrée dans le fichier suivant:

- q4-response.json (https://ict-bb-ch.smartlearn.cloud/media/f1a31aa74ed6e6f62db39b.json)

Analysez la réponse enregistrée et cochez l'affirmation correcte.

Single-Choice

- ☐ Le bon endpoint JSON a été choisi et les produits ont été correctement renvoyés. (HTTP 200)
- ☐ Le bon endpoint JSON a été choisi, mais les produits renvoyés appartiennent à une autre catégorie. (HTTP 200)
- ☐ Le bon endpoint JSON a été choisi, mais la requête a échoué en raison d'un paramètre manquant. (HTTP 400)
- ☐ Le bon endpoint JSON a été choisi, mais le LLM a halluciné des noms de produits qui n'existent pas dans la base de données. (HTTP 200)
- ☐ Un mauvais endpoint JSON a été choisi et les données renvoyées sont incomplètes ou inadaptées. (HTTP 200)
- ☐ La requête n'a pu être associée à aucune requête connue. (HTTP 400)

Question Q5

Quelle information manque dans le prompt `"Show me all the products in the berries category."` pour que l'endpoint `POST /llm/v1` puisse exécuter la requête de manière fiable? Justifiez brièvement.

Conseil: testez directement votre réponse sur l'endpoint `POST /llm/v1` dans Swagger UI.

Question Q6

Lors d'un test, l'endpoint **`POST /llm/v2`** a répondu au prompt `"Show me the details of bananas."` par la réponse JSON enregistrée dans le fichier suivant:

- q6-response.json (https://ict-bb-ch.smartlearn.cloud/media/57e564746b37a396418adf.json)

Analysez la réponse enregistrée. Nommez la table hallucinée et expliquez, en vous référant au schéma de base de données fourni dans la situation initiale, pourquoi cette table n'existe pas.

Corrigé type

Documents de référence

Module: 619 – Vérifier la faisabilité des solutions basées sur l'IA

Objectif opérationnel: OO5: Évalue les solutions basées sur l'IA à l'aide des données, des informations et des expériences recueillies, en fonction des critères

de faisabilité définis et des critères généraux liés à l'IA.

Compétence opérationnelle: C4 - Vérifier la faisabilité d'une solution basée sur l'IA

Critères de performance: Les AIBS peuvent…

- définir des critères d'évaluation des solutions basées sur l'IA en tenant compte des perspectives techniques, [*organisationnelles, juridiques et sociétales*].
- définir la procédure appropriée pour le test, ainsi que les outils nécessaires à cet effet, [*en collaboration avec des spécialistes*] (p. ex. développement de prototypes, réalisation de tests en laboratoire ou sur le terrain).
- étudier et évaluer la faisabilité d'une solution basée sur l'IA [*en impliquant les parties prenantes et les spécialistes concernés*], conformément aux critères définis.
- reconnaître les signes d'éventuels effets de bias dans les données ou les modèles et évaluer leur importance pour la solution en coopération avec des experts.