

A l'attention de JORDAN REY-JOUANCHICOT

2024-2025 ING3-HPDA

**Projet - Application IA**

---

# Rapport projet LLM

---

CHAABOUNI Sarah

TROUBAT Victoria



# Table des matières

<b>1</b>	<b>Introduction</b>	<b>2</b>
1.1	Présentation de la base de donnée . . . . .	2
1.2	Objectifs . . . . .	3
1.2.1	Structure et concept . . . . .	3
1.3	Composition du système RAG . . . . .	3
1.4	Phase d'indexation des données . . . . .	4
1.4.1	Choix de la taille des segments . . . . .	4
1.5	Recherche et récupération des informations . . . . .	4
1.5.1	Mesures de similarité . . . . .	4
<b>2</b>	<b>Etude des modèles</b>	<b>6</b>
<b>3</b>	<b>Évaluation des Modèles Ollama en Quantification 1B</b>	<b>7</b>
3.0.1	Méthodologie . . . . .	7
3.1	Questions posées . . . . .	8
3.2	Résultats . . . . .	8
3.3	Questions posées et résultats . . . . .	11
3.4	Discussion . . . . .	18

Notre projet s'articule autour de l'intelligence artificielle générative. Nous avons ainsi souhaité concevoir un chatbot capable de répondre à des besoins spécifiques. C'est dans cette optique que nous avons choisi d'explorer le domaine des modèles de langage de grande envergure (LLM), un sujet vaste et en adéquation avec notre vision du projet.

L'objectif de notre projet est d'exploiter nos cours d'Ingénieur 3 afin de développer un chatbot capable de répondre aux questions relatives aux matières intégrées dans notre base de données. Ainsi, les réponses fournies seront directement issues de nos propres ressources pédagogiques, garantissant une meilleure adéquation avec notre apprentissage. Contrairement à un chatbot généraliste tel que ChatGPT, notre solution permettra de se référer précisément aux leçons que nous avons étudiées.

Dans ce rapport, nous examinerons les principaux concepts et méthodes liés à notre projet, en mettant particulièrement l'accent sur la performance du code et les différentes options choisies pour l'optimiser.

## **1.1 Présentation de la base de donnée**

Pour notre projet, nous avons choisi de traiter des cours qui nous intéressent particulièrement, à savoir la matière suscitant le plus de questions lors de nos entretiens, ainsi que celle liée à notre spécialité en HPDA. C'est dans cette optique que nous avons téléchargé un diaporama fourni par l'un de nos professeurs.

## 1.2 Objectifs

L'objectif est de générer une réponse basée sur nos cours. Lorsqu'une question est posée au chatbot, celui-ci identifie les passages pertinents dans les supports de cours et formule une réponse en s'appuyant sur ces références.

### 1.2.1 Structure et concept

La génération augmentée par récupération (RAG) combine la récupération d'informations et la génération de texte afin d'améliorer la performance des modèles linguistiques en intégrant des sources de connaissances externes. Cette approche permet aux modèles de langage d'accéder à des informations actualisées et spécifiques à un domaine, rendant ainsi leurs réponses plus précises et pertinentes.

Dans ce cadre, nous allons explorer le concept de RAG et construire un système RAG simple à partir de zéro en utilisant **Python et Ollama**. Ce système aura pour objectif de récupérer des informations à partir d'un ensemble de données prédéfini et de générer des réponses en fonction des connaissances récupérées.

## 1.3 Composition du système RAG

Notre système RAG est structuré autour de trois composants essentiels :

- **Modèle d'embedding** : Un modèle de langage pré-entraîné qui convertit le texte en vecteurs d'embedding, capturant le sens sémantique des phrases. Ces vecteurs serviront à rechercher des informations pertinentes dans la base de données.
- **Base de données vectorielle** : Un système de stockage des connaissances et de leurs vecteurs d'embedding correspondants. De nombreuses technologies existent, telles que **Qdrant, Pinecone et pgvector**, mais nous allons ici implémenter une base en mémoire simple.
- **Chatbot** : Un modèle de langage capable de générer des réponses en se basant sur les connaissances récupérées. Il pourra s'appuyer sur des modèles tels que **Llama, Gemma ou GPT**.

## 1.4 Phase d'indexation des données

La première étape de la mise en place d'un système RAG consiste à **indexer les données**. Cette phase implique de découper les documents en petits segments et de calculer un vecteur d'embedding pour chacun d'eux, afin de faciliter leur recherche ultérieure.

### 1.4.1 Choix de la taille des segments

La taille des segments dépend de l'application :

- Dans un **système de recherche documentaire**, chaque segment peut correspondre à un **paragraphe** ou une **phrase**.
- Pour un **assistant conversationnel**, un segment peut représenter un **tour de conversation**.

Une fois l'indexation effectuée, chaque segment ainsi que son vecteur d'embedding correspondant seront stockés dans la base de données vectorielle.

## 1.5 Recherche et récupération des informations

Les vecteurs d'embedding seront ensuite utilisés pour retrouver les informations les plus pertinentes en fonction d'une requête. Ce processus est similaire à une requête SQL avec un WHERE, à la différence que l'on ne cherche pas une correspondance exacte de texte, mais une **similarité sémantique** entre la requête et les segments indexés.

### 1.5.1 Mesures de similarité

Pour comparer la similarité entre deux vecteurs, plusieurs métriques peuvent être utilisées, notamment :

- **Cosinus similarité**
- **Distance euclidienne**
- **Autres mesures de distance vectorielle**

Dans notre implémentation, nous utiliserons la **cosine similarity**, qui permet d'évaluer à quel point deux vecteurs sont proches en termes d'orientation dans l'espace vectoriel. Sa formule est donnée par :

$$\text{similarité} = \frac{A \cdot B}{\|A\| \times \|B\|} \quad (1.1)$$

Ainsi, en combinant ces différentes étapes, notre système RAG sera capable d'enrichir les réponses du chatbot en s'appuyant sur des connaissances précises et actualisées, améliorant ainsi la pertinence et la fiabilité de ses réponses.

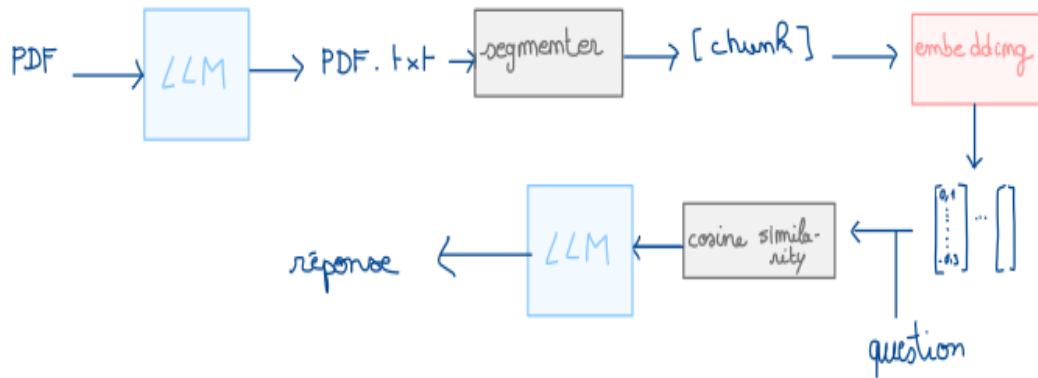


FIGURE 1.1 – Schéma récapitulatif

Pour le projet nous avons utilisé plusieurs modèles :

- **Transformation du PDF en texte** : Pour convertir notre PDF en un fichier texte, nous avons utilisé `langchain.text_splitter`. Le module `CharacterTextSplitter` de `LangChain` permet de diviser un texte en segments plus petits en fonction d'un caractère spécifique (par défaut, un espace ou un saut de ligne). Cela est particulièrement utile pour traiter de longs documents, comme un PDF, afin de les rendre plus facilement exploitables par des modèles de langage.
- **Embedding** : Nous avons utilisé le modèle `sentence-transformers/all-MiniLM-L6-v2`. Il s'agit d'un modèle *sentence-transformers* qui convertit des phrases et des paragraphes en un espace vectoriel dense de 384 dimensions. Cela permet des applications telles que le regroupement (*clustering*) ou la recherche sémantique.
- **Modèle de langage** : Nous avons choisi la collection **Llama 3.2**, qui regroupe de grands modèles de langage multilingues (LLMs). Cette collection comprend des modèles génératifs pré-entraînés et ajustés aux instructions, disponibles en tailles **1B et 3B** (entrée texte/sortie texte). Les modèles **Llama 3.2**, optimisés pour le traitement des instructions, sont spécialement conçus pour des cas d'utilisation en dialogue multilingue, y compris la récupération d'informations agentique et les tâches de résumé.

## Évaluation des Modèles Ollama en Quantification 1B

Afin d'optimiser la quantification, nous avons testé différents modèles Ollama (version 1B) en utilisant le même fichier de test, `test_cat_copy.txt`, qui est intégré dans le script `rag.py`.

- **Transformation du PDF en texte :**

### 3.0.1 Méthodologie

Nous avons identifié trois étapes d'exécution dans le code :

1. **Embedding de la question :** Temps d'exécution mesuré depuis l'envoi de la question jusqu'à l'obtention de son embedding.
2. **Calcul des similarités :** Temps d'exécution pour trouver les similarités entre l'embedding de la question et le texte. La mesure débute après l'embedding de la question et se termine après l'affichage des chunks similaires.
3. **Génération de la réponse :** Temps nécessaire pour obtenir la réponse du chatbot.

Nous testons les trois modèles suivants :

- **`gemma3:1b-it-q4_K_M`**
- **`gemma3:1b-it-fp16`**
- **`gemma3:1b-it-q8_0`**



## 3.1 Questions posées

### Questions avec réponse dans le texte (vérifiables)

**Q1 – Question :** "Depuis combien de temps et dans quel pays remontent les premières preuves de domestication du chat?"

**Réponse attendue :** "Environ 9 000 ans en Égypte ancienne."

**Localisation de la réponse :** Dans la section *Origines et domestication*.

**Q2 – Question :** "Quel rôle jouait le chat dans l'Antiquité?"

**Réponse attendue :** "Il était vénéré, considéré comme un protecteur des récoltes et des maisons, et associé à la déesse Bastet."

**Localisation de la réponse :** Dans la partie décrivant la domestication et les premières relations entre humains et chats.

**Q3 – Question :** "Quelles sont deux caractéristiques physiques mentionnées dans le texte qui permettent au chat de chasser efficacement dans l'obscurité?"

**Réponse attendue :** "Sa tête arrondie avec de grands yeux lui donnant une excellente vision nocturne, et ses vibrisses (moustaches) qui lui permettent de détecter les moindres variations dans l'air."

**Localisation de la réponse :** Dans la section *Morphologie et physiologie*.

### Questions sans réponse explicite dans le texte (pour vérifier que le modèle n'invente pas)

**Q4 – Poids moyen Question :** "Quel est le poids moyen d'un chat domestique adulte?"

**Réponse attendue :** Le texte ne fournit pas cette information.

**But :** Vérifier que le modèle indique qu'il ne dispose pas de cette donnée et ne l'invente pas.

## 3.2 Résultats

### gemma3:1b-it-fp16

**Q1 – Question :** "Depuis combien de temps et dans quel pays remontent les premières preuves de domestication du chat?"

**Résultats :**

Retrieved knowledge :

- (similarity : 0.69) L'avenir des chats dans nos sociétés
- (similarity : 0.68) Les premières preuves de domestication du chat remontent à environ 9 000 ans en Égypte ancienne
- (similarity : 0.66) La santé des chats nécessite une attention régulière

Chatbotresponse:According to the context, the first evidence of cat domestication dates back to approximately 9,000 years in ancient Egypt.

**Temps d'exécution :**

Étape 1 : 1.5056743621826172 s

Étape 2 : 0.00014090538024902344 s

Étape 3 : 0.0017974376678466797 s

**Commentaire :** Le résultat est conforme à l'attendu, la réponse mentionne bien environ 9 000 ans en Égypte ancienne.

**Q2 – Question :** "Quel rôle jouait le chat dans l'Antiquité?"**Résultats :**

Retrieved knowledge :

- (similarity : 0.72) L'avenir des chats dans nos sociétés
- (similarity : 0.70) La santé des chats nécessite une attention régulière
- (similarity : 0.66) Les chats sont réputés pour leur indépendance, mais ils possèdent également un riche répertoire de comportements sociaux et de communications. Contrairement à certaines idées reçues, ils ne sont pas des animaux solitaires par nature

Chatbotresponse:According to the provided text, the cat's role in antiquity was primarily that of an independent animal. However, it also possessed a "rich repertoire of behaviors and communications." It wasn't seen as a solitary creature.

**Temps d'exécution :**

Étape 1 : 1.991600513458252 s

Étape 2 : 0.00010728836059570312 s

Étape 3 : 0.0009253025054931641 s

**Commentaire :** La réponse diffère de l'attendu. Au lieu de mentionner le caractère sacré et

protecteur du chat dans l'Antiquité, le modèle insiste sur son indépendance et son répertoire comportemental.

**Q3 – Question :** "Quelles sont deux caractéristiques physiques mentionnées dans le texte qui permettent au chat de chasser efficacement dans l'obscurité?"

**Résultats :**

Retrieved knowledge :

- (similarity : 0.70) La tête du chat est souvent arrondie, avec de grands yeux qui lui confèrent une excellente vision nocturne, indispensable pour chasser dans l'obscurité.
- (similarity : 0.65) La santé des chats nécessite une attention régulière
- (similarity : 0.65) La communication chez le chat se fait principalement par des moyens non verbaux

Chatbot response: Deux caractéristiques physiques mentionnées dans le texte qui permettent au chat de chasser efficacement dans l'obscurité sont ses grands yeux qui lui confèrent une excellente vision nocturne.

**Temps d'exécution :**

Étape 1 : 1.1878459453582764 s

Étape 2 : 8.821487426757812e-05 s

Étape 3 : 0.0009903907775878906 s

**Commentaire :** La réponse fournie est incomplète puisqu'elle mentionne uniquement les grands yeux et omet les vibrisses, contrairement à ce qui est attendu.

**Q4 – Question :** "Quel est le poids moyen d'un chat domestique adulte?"

**Résultats :**

Retrieved knowledge :

- (similarity : 0.70) Le jeu est également un comportement essentiel chez le chat, tant pour les jeunes que pour les adultes.
- (similarity : 0.69) La santé des chats nécessite une attention régulière.
- (similarity : 0.68) L'avenir des chats dans nos sociétés.

Chatbot response : According to the provided text, the weight of a domestic cat is "également un comportement essentiel chez le chat, tant pour les jeunes que pour les adultes". However, it doesn't provide a specific weight.

**Temps d'exécution :**

Étape 1 : 1.017512559890747 s

Étape 2 : 0.0001201629638671875 s

Étape 3 : 0.0020148754119873047 s

**Commentaire :** La réponse obtenue pour la question sur le poids moyen indique correctement qu'aucune donnée n'est fournie dans le texte. Cependant, le modèle reprend un passage inapproprié, en citant une phrase sur le comportement ludique du chat, ce qui montre une réponse détournée. Cela suggère que le modèle n'a pas fourni une réponse pertinente lorsque l'information est absente.

### 3.3 Questions posées et résultats

**Questions avec réponse dans le texte (vérifiables) :**

**Q1 – Question :** "Depuis combien de temps et dans quel pays remontent les premières preuves de domestication du chat?"

**Réponse attendue :** "Environ 9 000 ans en Égypte ancienne."

**Localisation de la réponse :** Dans la section *Origines et domestication*.

**Résultats (gemma3 :1b-it-fp16) :**

Retrieved knowledge :

- (similarity : 0.69) L'avenir des chats dans nos sociétés
- (similarity : 0.68) Les premières preuves de domestication du chat remontent à environ 9 000 ans en Égypte ancienne
- (similarity : 0.66) La santé des chats nécessite une attention régulière

Chatbot response : According to the context, the first evidence of cat domestication dates back to approximately 9,000 years in ancient Egypt.

**Temps d'exécution :**

Étape 1 : 1.5056743621826172 s

Étape 2 : 0.00014090538024902344 s

Étape 3 : 0.0017974376678466797 s

**Commentaire :** Le résultat est conforme à l'attendu, la réponse mentionne bien environ 9 000 ans en Égypte ancienne.

**Q2 – Question :** "Quel rôle jouait le chat dans l'Antiquité?"

**Réponse attendue :** "Il était vénéré, considéré comme un protecteur des récoltes et des maisons, et associé à la déesse Bastet."

**Localisation de la réponse :** Dans la partie décrivant la domestication et les premières relations entre humains et chats.

**Résultats (gemma3 :1b-it-fp16) :**

Retrieved knowledge :

- (similarity : 0.72) L'avenir des chats dans nos sociétés
- (similarity : 0.70) La santé des chats nécessite une attention régulière
- (similarity : 0.66) Les chats sont réputés pour leur indépendance, mais ils possèdent également un riche répertoire de comportements sociaux et de communications. Contrairement à certaines idées reçues, ils ne sont pas des animaux solitaires par nature

Chatbotresponse:According to the provided text, the cat's role in antiquity was primarily that of an independent animal. However, it also possessed a "rich repertoire of behaviors and communications." It wasn't seen as a solitary creature.

**Temps d'exécution :**

Étape 1 : 1.991600513458252 s

Étape 2 : 0.00010728836059570312 s

Étape 3 : 0.0009253025054931641 s

**Commentaire :** La réponse diffère de l'attendu. Au lieu de mentionner le caractère sacré et protecteur du chat dans l'Antiquité, le modèle insiste sur son indépendance et son répertoire comportemental.

**Q3 – Question :** "Quelles sont deux caractéristiques physiques mentionnées dans le texte qui permettent au chat de chasser efficacement dans l'obscurité?"

**Réponse attendue :** "Sa tête arrondie avec de grands yeux lui conférant une excellente vision nocturne, et ses vibrisses (moustaches) qui lui permettent de détecter les moindres variations dans l'air."

**Localisation de la réponse :** Dans la section *Morphologie et physiologie*.

**Résultats (gemma3 :1b-it-fp16) :**

Retrieved knowledge :

- (similarity : 0.70) La tête du chat est souvent arrondie, avec de grands yeux qui lui confèrent une excellente vision nocturne, indispensable pour chasser dans l'obs-

curité.

- (similarity : 0.65) La santé des chats nécessite une attention régulière
- (similarity : 0.65) La communication chez le chat se fait principalement par des moyens non verbaux

Chatbot response : Deux caractéristiques physiques mentionnées dans le texte qui permettent au chat de chasser efficacement dans l'obscurité sont ses grands yeux qui lui confèrent une excellente vision nocturne.

**Temps d'exécution :**

Étape 1 : 1.1878459453582764 s

Étape 2 : 8.821487426757812e-05 s

Étape 3 : 0.0009903907775878906 s

**Commentaire :** La réponse fournie est incomplète puisqu'elle mentionne uniquement les grands yeux et omet les vibrisses, contrairement à ce qui est attendu.

**Questions sans réponse explicite dans le texte (pour vérifier que le modèle n'invente pas) :**

Q4 –

**Question :** "Quel est le poids moyen d'un chat domestique adulte?"

**Réponse attendue :** Le texte ne fournit pas cette information.

**But :** Vérifier que le modèle indique qu'il n'a pas de donnée ou ne l'invente pas.

**Résultats (gemma3:1b-it-fp16) :**

Retrieved knowledge :

- (similarity : 0.70) Le jeu est également un comportement essentiel chez le chat, tant pour les jeunes que pour les adultes
- (similarity : 0.69) La santé des chats nécessite une attention régulière
- (similarity : 0.68) L'avenir des chats dans nos sociétés

Chatbotresponse:According to the text, the weight of a domestic cat adult is "également un comportement essentiel chez le chat, tant pour les jeunes que pour les adultes". It doesn't provide a specific weight.

**Temps d'exécution :**

Étape 1 : 1.128087043762207 s

Étape 2 : 0.00011324882507324219 s

Étape 3 : 0.0018084049224853516 s

**Commentaire :** La réponse indique que l'information n'est pas fournie, ce qui est correct, mais le modèle se trompe en reprenant un passage inapproprié du texte, montrant une réponse détournée.

## Modèle gemma3 :1b-it-q8\_0

**Q1 – Question :** "Depuis combien de temps et dans quel pays remontent les premières preuves de domestication du chat?"

**Résultats :**

Retrieved knowledge :

- (similarity : 0.69) L'avenir des chats dans nos sociétés
- (similarity : 0.68) Les premières preuves de domestication du chat remontent à environ 9 000 ans en Égypte ancienne
- (similarity : 0.66) La santé des chats nécessite une attention régulière

Chatbotresponse: Selon les informations fournies, les premières preuves de domestication du chat remontent à environ 9 000 ans en Égypte ancienne.

**Temps d'exécution :**

Étape 1 : 1.1267406940460205 s

Étape 2 : 0.00022101402282714844 s

Étape 3 : 0.0029630661010742188 s

**Commentaire :** La réponse correspond exactement à ce qui est attendu.

**Q2 – Question :** "Quel rôle jouait le chat dans l'Antiquité?"

**Résultats :**

Retrieved knowledge :

- (similarity : 0.72) L'avenir des chats dans nos sociétés
- (similarity : 0.70) La santé des chats nécessite une attention régulière
- (similarity : 0.66) Les chats sont réputés pour leur indépendance, mais ils possèdent également un riche répertoire de comportements sociaux et de communications.

Chatbot response: According to the provided context, the cat played a significant role in the ancient world, particularly as a symbol of independence but also as a social animal with a rich repertoire of behaviors.

**Temps d'exécution :**

Étape 1 : 1.6995282173156738 s

Étape 2 : 0.0001251697540283203 s

Étape 3 : 0.0019545555114746094 s

**Commentaire :** La réponse obtenue ne correspond pas à l'attendu car elle ne mentionne pas le caractère sacré du chat ni son association avec Bastet.

**Q3 – Question :** "Quelles sont deux caractéristiques physiques mentionnées dans le texte qui permettent au chat de chasser efficacement dans l'obscurité?"

**Résultats :**

Retrieved knowledge :

- (similarity : 0.70) La tête du chat est souvent arrondie, avec de grands yeux qui lui confèrent une excellente vision nocturne, indispensable pour chasser dans l'obscurité.
- (similarity : 0.65) La santé des chats nécessite une attention régulière
- (similarity : 0.65) La communication chez le chat se fait principalement par des moyens non verbaux

Chatbotresponse: Selon le texte, deux caractéristiques physiques permettent au chat de chasser efficacement dans l'obscurité sont ses grands yeux et son excellente vision nocturne.

**Temps d'exécution :**

Étape 1 : 1.2417068481445312 s

Étape 2 : 6.771087646484375e-05 s

Étape 3 : 0.0008323192596435547 s

**Commentaire :** La réponse est incomplète puisqu'elle omet de mentionner les vibrisses, contrairement à ce qui était attendu.

**Q4 – Question :** "Quel est le poids moyen d'un chat domestique adulte?"

**Résultats :**

Retrieved knowledge :

- (similarity : 0.70) Le jeu est également un comportement essentiel chez le chat, tant pour les jeunes que pour les adultes
- (similarity : 0.69) La santé des chats nécessite une attention régulière
- (similarity : 0.68) L'avenir des chats dans nos sociétés



Chatbotresponse:According to the provided text, the weight of a domestic cat is “également un comportement essentiel chez le chat, tant pour les jeunes que pour les adultes”. However, it doesn’t provide a specific weight.

**Temps d’exécution :**

Étape 2 : 0.0001201629638671875 s

Étape 3 : 0.0020148754119873047 s

**Commentaire :** La réponse indique que l’information n’est pas présente, mais elle reproduit un passage inapproprié, montrant que le modèle n’a pas fourni une réponse pertinente.

## **Modèle gemma3:1b-it-q4\_K\_M**

**Q1 – Question :** "Depuis combien de temps et dans quel pays remontent les premières preuves de domestication du chat?"

**Résultats :**

Retrieved knowledge :

- (similarity : 0.69) L’avenir des chats dans nos sociétés
- (similarity : 0.68) Les premières preuves de domestication du chat remontent à environ 9 000 ans en Égypte ancienne
- (similarity : 0.66) La santé des chats nécessite une attention régulière

Chatbotresponse:According to the provided text, the earliest evidence of cat domestication dates back to approximately 9,000 years ago in ancient Egypt.

**Temps d’exécution :**

Étape 1 : 1.3789432048797607 s

Étape 2 : 0.00022101402282714844 s

Étape 3 : 0.001796722412109375 s

**Commentaire :** La réponse est conforme à l’attendu.

**Q2 – Question :** "Quel rôle jouait le chat dans l’Antiquité?"

**Résultats :**

Retrieved knowledge :

- (similarity : 0.72) L’avenir des chats dans nos sociétés
- (similarity : 0.70) La santé des chats nécessite une attention régulière

- (similarity : 0.66) Les chats sont réputés pour leur indépendance, mais ils possèdent également un riche répertoire de comportements sociaux et de communications.

Chatbotresponse:According to the provided text, the text does not provide information about the role of cats in antiquity.

**Temps d'exécution :**

Étape 1 : 1.1811258792877197 s

Étape 2 : 8.654594421386719e-05 s

Étape 3 : 0.0011930465698242188 s

**Commentaire :** La réponse est incorrecte. Le texte fournit clairement des indications sur le rôle du chat dans l'Antiquité (vénération et protection), mais le modèle ne restitue pas cette information.

**Q3 – Question :** "Quelles sont deux caractéristiques physiques mentionnées dans le texte qui permettent au chat de chasser efficacement dans l'obscurité?"

**Résultats :**

Retrieved knowledge :

- (similarity : 0.70) La tête du chat est souvent arrondie, avec de grands yeux qui lui confèrent une excellente vision nocturne, indispensable pour chasser dans l'obscurité.
- (similarity : 0.65) La santé des chats nécessite une attention régulière.
- (similarity : 0.65) La communication chez le chat se fait principalement par des moyens non verbaux.

Chatbotresponse:Selon le texte, deux caractéristiques physiques permettent au chat de chasser efficacement dans l'obscurité sont ses grands yeux et son excellent vision nocturne.

**Temps d'exécution :**

Étape 1 : 1.2417068481445312 s

Étape 2 : 6.771087646484375e-05 s

Étape 3 : 0.0008323192596435547 s

**Commentaire :** La réponse obtenue pour la question sur le poids moyen indique correctement qu'aucune donnée n'est fournie dans le texte. Cependant, le modèle reprend un passage inapproprié, en citant une phrase sur le comportement ludique du chat, ce qui montre une réponse détournée. Cela suggère que le modèle n'a pas fourni

une réponse pertinente lorsque l'information est absente.

**Q4 – Question :** "Quel est le poids moyen d'un chat domestique adulte?"

**Résultats :**

Retrieved knowledge :

- (similarity : 0.70) Le jeu est également un comportement essentiel chez le chat, tant pour les jeunes que pour les adultes.
- (similarity : 0.69) La santé des chats nécessite une attention régulière.
- (similarity : 0.68) L'avenir des chats dans nos sociétés.

Chatbotresponse:According to the provided text, the weight of a domestic cat is “également un comportement essentiel chez le chat, tant pour les jeunes que pour les adultes”. However, it doesn't provide a specific weight.

**Temps d'exécution :**

Étape 1 : 1.017512559890747 s

Étape 2 : 0.0001201629638671875 s

Étape 3 : 0.0020148754119873047 s

**Commentaire :** La réponse obtenue pour la question sur le poids moyen indique correctement qu'aucune donnée n'est fournie dans le texte. Cependant, le modèle reprend un passage inapproprié, en citant une phrase sur le comportement ludique du chat, ce qui montre une réponse détournée. Cela suggère que le modèle n'a pas fourni une réponse pertinente lorsque l'information est absente.

### 3.4 Discussion

Ces résultats permettent de comparer la précision et la rapidité des différents modèles en posant les mêmes questions sur un texte de référence. Pour les questions vérifiables (Q1 à Q3), nous observons :

- **Q1 :** Les trois modèles fournissent des réponses cohérentes, indiquant correctement qu'il s'agit d'environ 9 000 ans en Égypte ancienne.
- **Q2 :** Les réponses varient. Le modèle gemma3 :1b-it-fp16 restitue des informations erronées, tandis que gemma3 :1b-it-q4\_K\_M indique qu'aucune information n'est fournie, ce qui est en contradiction avec le texte. Cela montre une disparité dans la capacité à extraire des détails contextuels précis.

- **Q3 :** Les modèles reconnaissent l'importance des yeux pour la vision nocturne, mais omettent de mentionner les vibrisses, entraînant des réponses incomplètes.

Pour la question sans réponse explicite (Q4 – Poids moyen), tous les modèles indiquent qu'ils ne disposent pas de l'information, bien que certains reproduisent des passages inappropriés. Cela démontre que, lorsqu'une donnée est absente, le modèle tend parfois à "inventer" une réponse partielle.

Les temps d'exécution mesurés aux trois étapes fournissent également des indications sur la performance de chaque modèle, ce qui est essentiel pour choisir le modèle le plus adapté en termes de rapidité et de précision.