TRANSFORMEZ L'ACCÈS À LA CONNAISSANCE LESAFFRE AVEC UN RAG AVANCÉ

JULIE ADALIAN

Data Scientist spécialisée sur les sujets Generative AI - Lesaffre International

MsC Data science - HWU
Ingenieure du numérique - Isep

SOMMAIRE

OUR ACTIVITIES

The infinite potential of microorganisms (yeasts, bacteria…) enables us to position ourselves in the bread making, food taste and pleasure, healthcare and industrial biotechnology markets.

In each of these domains, Lesaffre's ambition is to be **one of the active leaders in the fermentation of microorganisms to better nourish and protect the planet.**

LESAFFRE

EN 2024

3 milliards d'euros de chiffre d'affaires

180 Nos produits sont distribués dans plus de 180 pays

80 sites de production dans plus de **50** pays

88% de part du CA issu de produits vendus ayant été fabriqués par un site certifié GFSI*

63 centres de sciences appliquées

9 laboratoires d'analyse sensorielle

51 Baking Center™

90 partenariats internationaux

700 experts R, D&I

11 000 collaborateurs

96 nationalités

SILVER 2023 ecovadis Sustainability Rating

ENTREPRENDRE ENSEMBLE
POUR MIEUX NOURRIR ET PROTÉGER LA PLANÈTE

** Global Food Safety Initiative*

RAG INTRODUCTION

AVANT LE RAG: L'INDEXATION

L'OBJECTIF EST DE CRÉER UNE BASE DE DONNÉE DE SÉMANTIQUE ISSUE DES DOCUMENTS

Document sur le sujet X

Document pas sur le sujet X

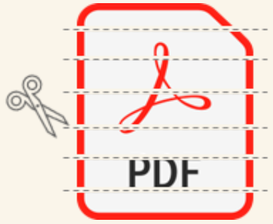
Differents types de documents sont utilisés

AVANT LE RAG: L'INDEXATION

L'OBJECTIF EST DE CRÉER UNE BASE DE DONNÉE DE SÉMANTIQUE ISSUE DES DOCUMENTS

Document sur le sujet X

Document pas sur le sujet X

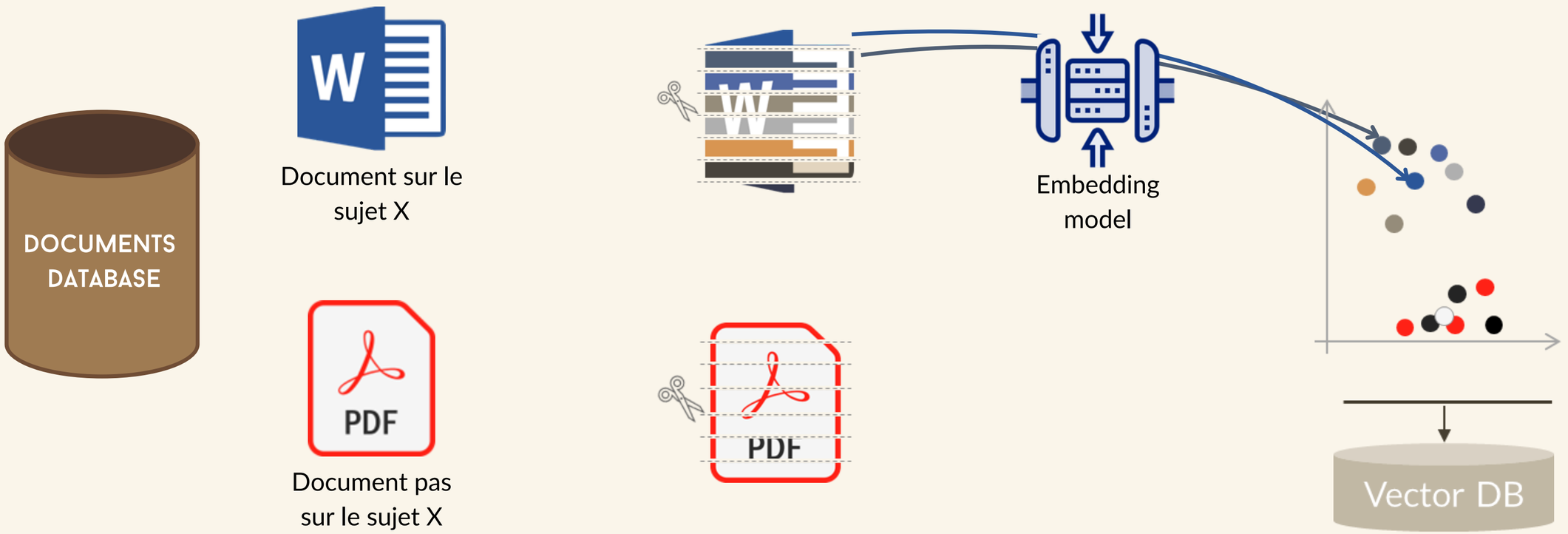


Differents types de documents sont utilisés

Découpés en paragraphes

AVANT LE RAG: L'INDEXATION

L'OBJECTIF EST DE CRÉER UNE BASE DE DONNÉE DE SÉMANTIQUE ISSUE DES DOCUMENTS

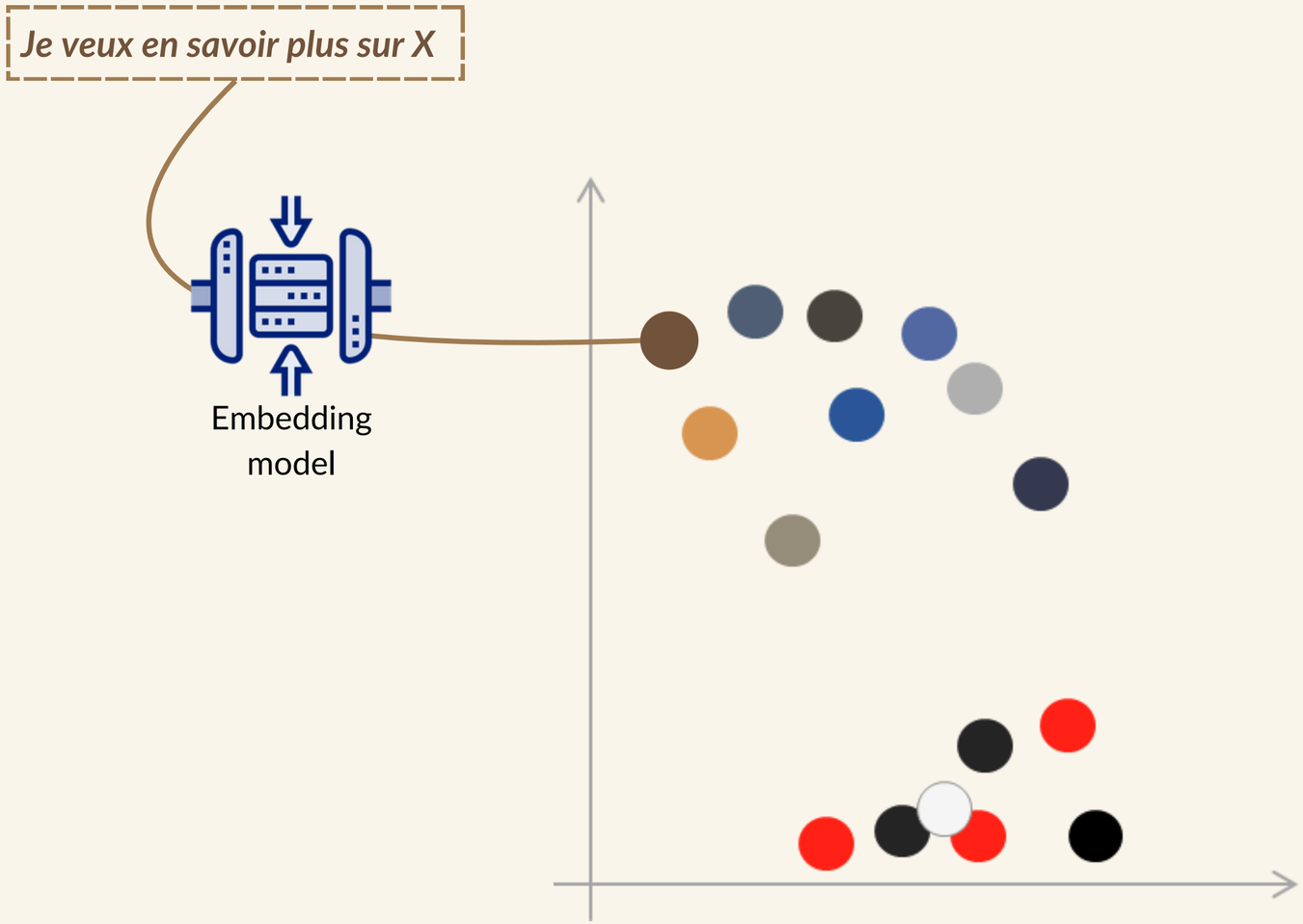


Differents types de documents sont utilisés

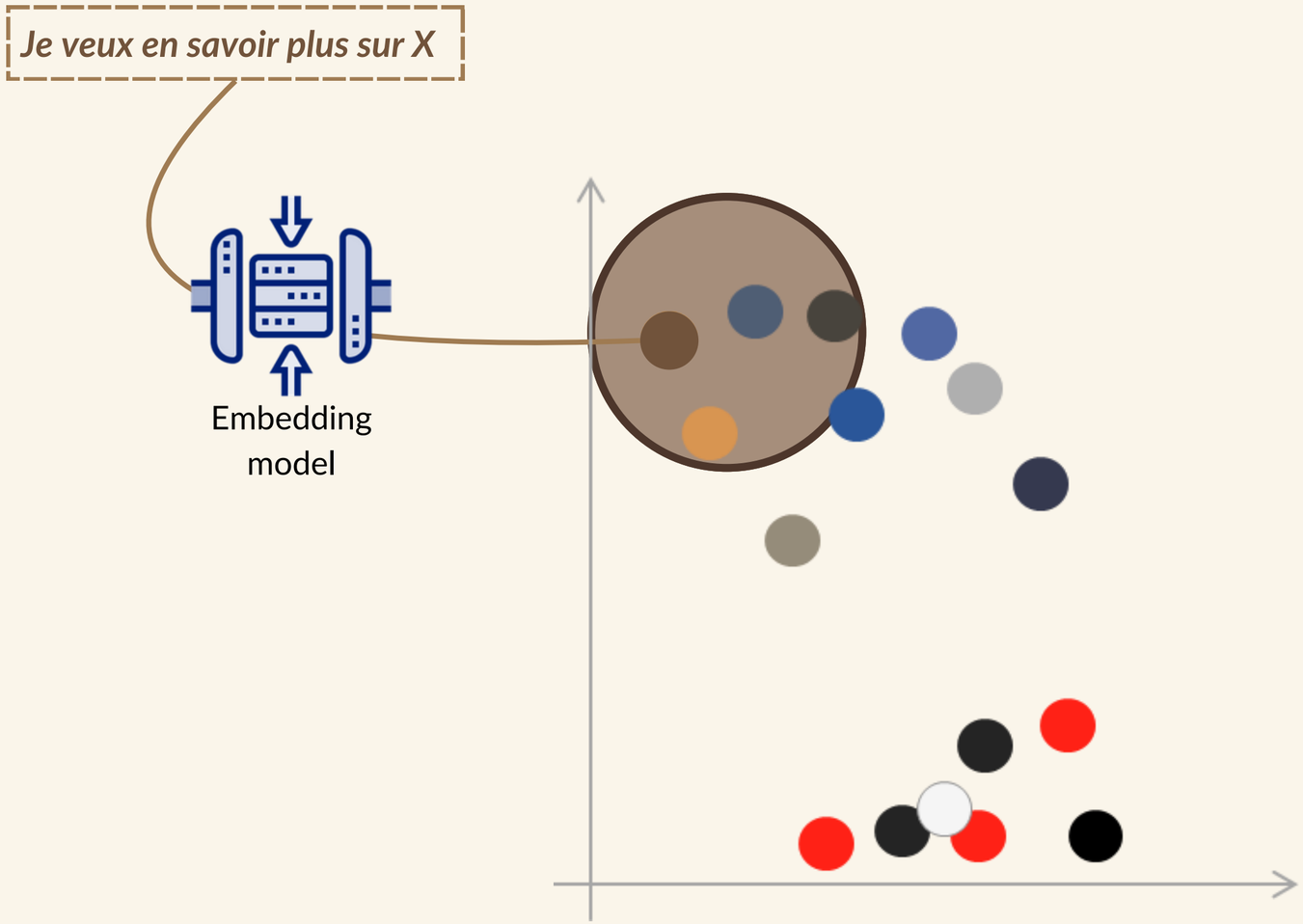
Découpés en paragraphes

Transformés en vecteurs et stockés dans une base de donnée vectorielle

LA BASE DE DONNÉE VECTORIELLE EST REMPLIE, LE RAG PEUT COMMENCER

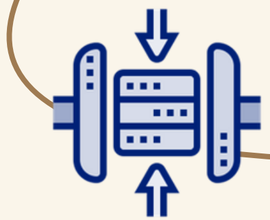


LA BASE DE DONNÉE VECTORIELLE EST REMPLIE, LE RAG PEUT COMMENCER

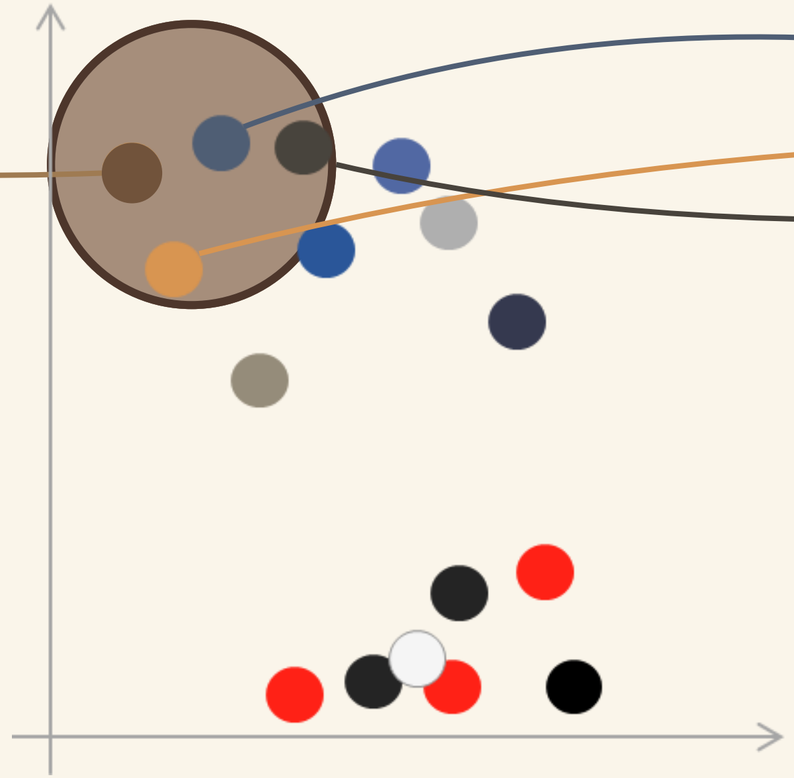


LA BASE DE DONNÉE VECTORIELLE EST REMPLIE, LE RAG PEUT COMMENCER

Je veux en savoir plus sur X



Embedding model



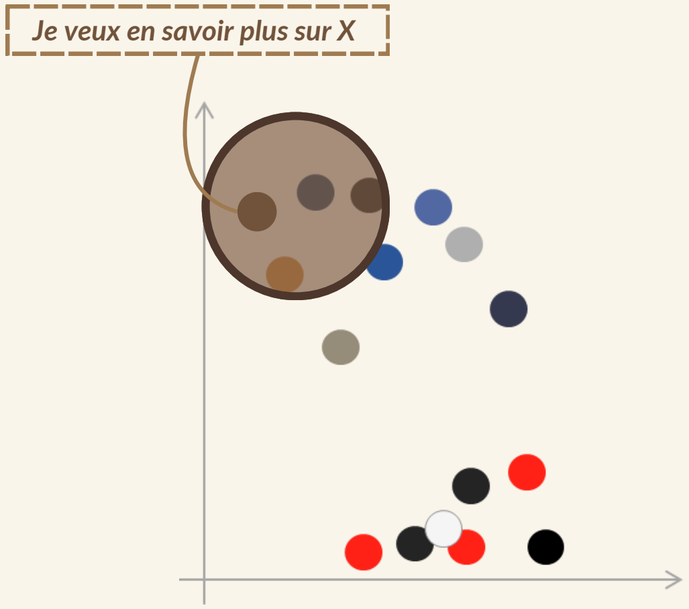
Compte tenu des paragraphes suivants:

Répond à la question:

Je veux en savoir plus sur X

LLM Les documents mentionnent que X concerne ...

RETRIEVAL AUGMENTED GENERATION (RAG) UTILISE LA PUISSANCE DES MODELES DE LANGAGE SUR UNE BASE DE DOCUMENTS SPÉCIFIQUE



Compte tenu des paragraphes suivants:

Répond à la question:

Je veux en savoir plus sur X

LLM Les documents mentionnent que X concerne ...

RETRIEVAL

AUGMENTED

GENERATION

RAG EVALUATION

RAG EVALUATION

- \+ PEUT ETRE FACILEMENT RÉPÉTÉ
- \- EVALUATION PEU FIABLES ET ASSEZ SUPERFICIELLE

ragas DEEPEVAL Giskard

--> PERTINENT POUR ITERER SUR LA PIPELINE DE RAG

- \- COUTEUX & DIFFICILE À IMPLEMENTER
- \+ EVALUATION EN PROFONDEUR

GROUPE D'EXPERTS QUI CONNAIT LES DOCUMENTS

MINIMUM 10 EXPERTS

BIEN BRIEFER LES UTILISATEURS

--> PERTINENT POUR VALIDER LA QUALITÉ DE LA RÉPONSE FINALE

LIMITES DE LA SOLUTION

1. FORMAT DE DOCS
2. CHOIX DE LA BONNE TAILLE DE CONTEXTE
3. POSSIBILITÉ DE FILTRER
4. VOCABULAIRE SPÉCIFIQUE
5. MANQUE DE REPETABILITÉ
6. DIFFERENCES FRANÇAIS / ANGLAIS
7. AVOIR UN SCORE DE PERTINENCE
8. HALLUCINATIONS

IMPLÉMENTATION DES SOLUTIONS

1

FORMAT DE DOCS

OBJECTIF: EXTRACTION DU CONTENU DE CES DIFFÉRENTS DOCUMENTS

.DOC

.DOCX

.PDF

.PPTX

.XLS

.XLSX

1ÈRE ITERATION: UNE FONCTION DE RÉCUPÉRATION PAR TYPE DE FICHIER

2ÈME ITERATION: CONVERSION EN PDF PUIS EXTRACTION DU TEXTE

RETEX:

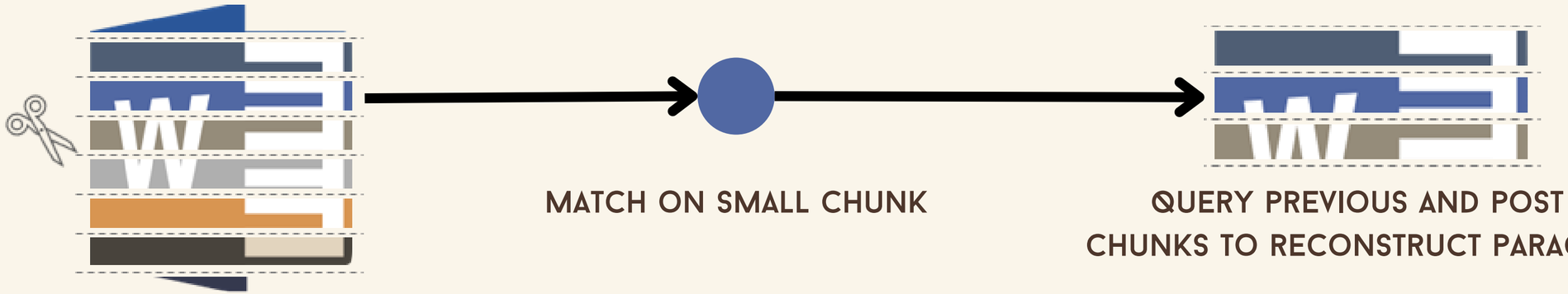
- Certains services proposent d'analyser le contenu des documents pour permettre une récupération plus intelligente et classifiée. Ces fonctions d'analyse peuvent coûter très cher
- Utiliser un LLM avec vision pour extraire le texte peut paraitre etre une bonne idée mais le format du document peut ne pas être correctement récupéré et les hallucinations sont possibles.

ADVANCED RAG

IMPLÉMENTATION DES SOLUTIONS

2 CHOIX DE LA BONNE TAILLE DE CONTEXTE

DILEMME: AVOIR UNE SÉMANTIQUE SUFFISAMMENT PRÉCISE ET DONNER SUFFISAMMENT DE CONTEXTE POUR QUE LE LLM AI SUFFISAMMENT D'ÉLÉMENTS POUR RÉPONDRE
SENTENCE WINDOW RETRIEVAL



IMPLÉMENTATION DES SOLUTIONS

3

POSSIBILITÉ DE FILTRER

SELF-QUERY RETRIEVAL

RETEX:

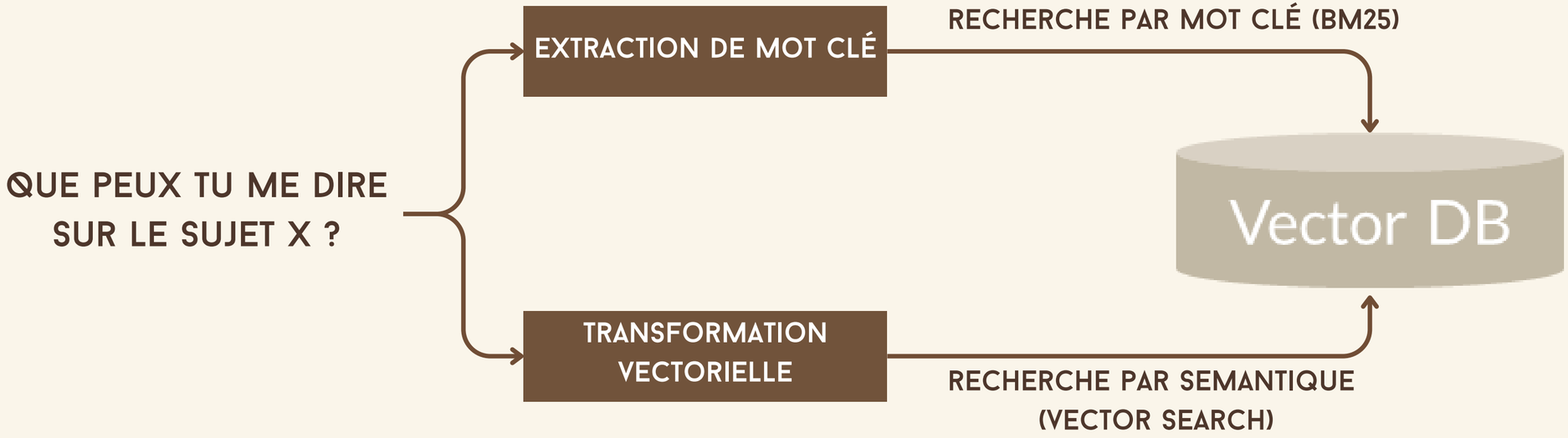
- Quand vous choisissez votre technologie de base de données vectorielle: faites bien attention aux types de requêtes qu'il est possible de faire
- Attention à la définition des types de filtres: il doivent etre suffisamment précis pour que le modèle puisse les comprendre

IMPLÉMENTATION DES SOLUTIONS

4

VOCABULAIRE SPÉCIFIQUE

L'EMBEDDING MODEL EST ENTRAINÉ SUR UN LARGE CORPUS DE DOCUMENTS MAIS COMMENT FAIRE LORSQU'IL S'AGIT DE VOCABULAIRE SPECIFIQUE NICHE OU SPÉCIFIQUES À L'ENTREPRISE ?

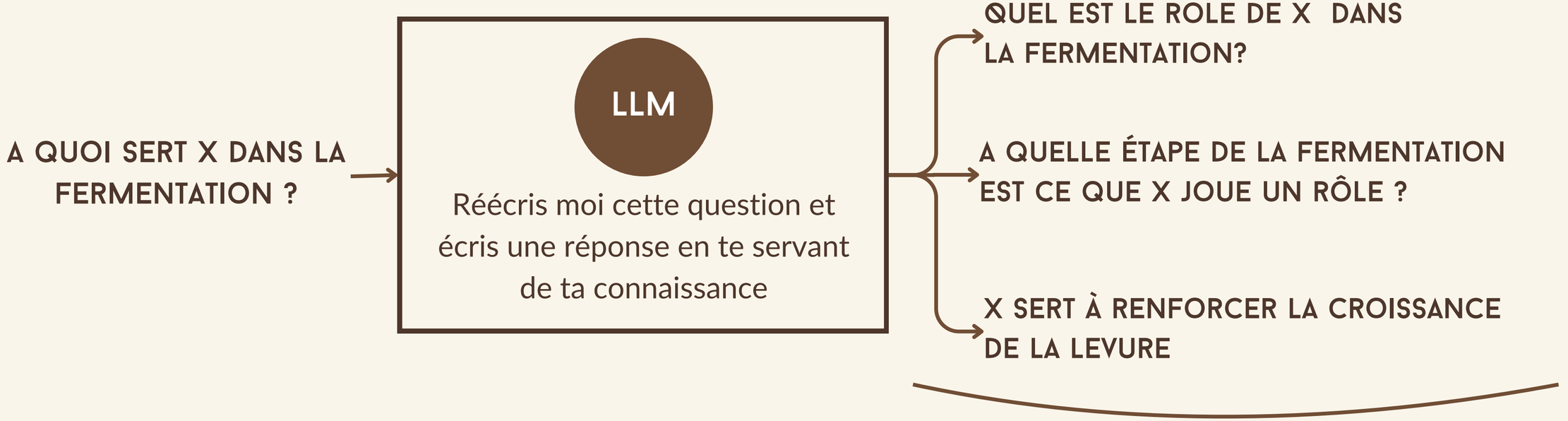


IMPLÉMENTATION DES SOLUTIONS

5

MANQUE DE REPETABILITÉ

SI JE POSE LA QUESTION PLUSIEURS FOIS ET SI JE CHANGE UN MOT J'OBTIENS DES RESULTATS TRÈS DIFFÉRENTS. (QUERY DERIVATIVES + HYDE)



L'INTERROGATION DE LA VECTOR DATABASE AVEC CES DÉRIVÉES QUESTIONS PERMET UNE MEILLEURE COUVERTURE DE LA SÉMANTIQUE

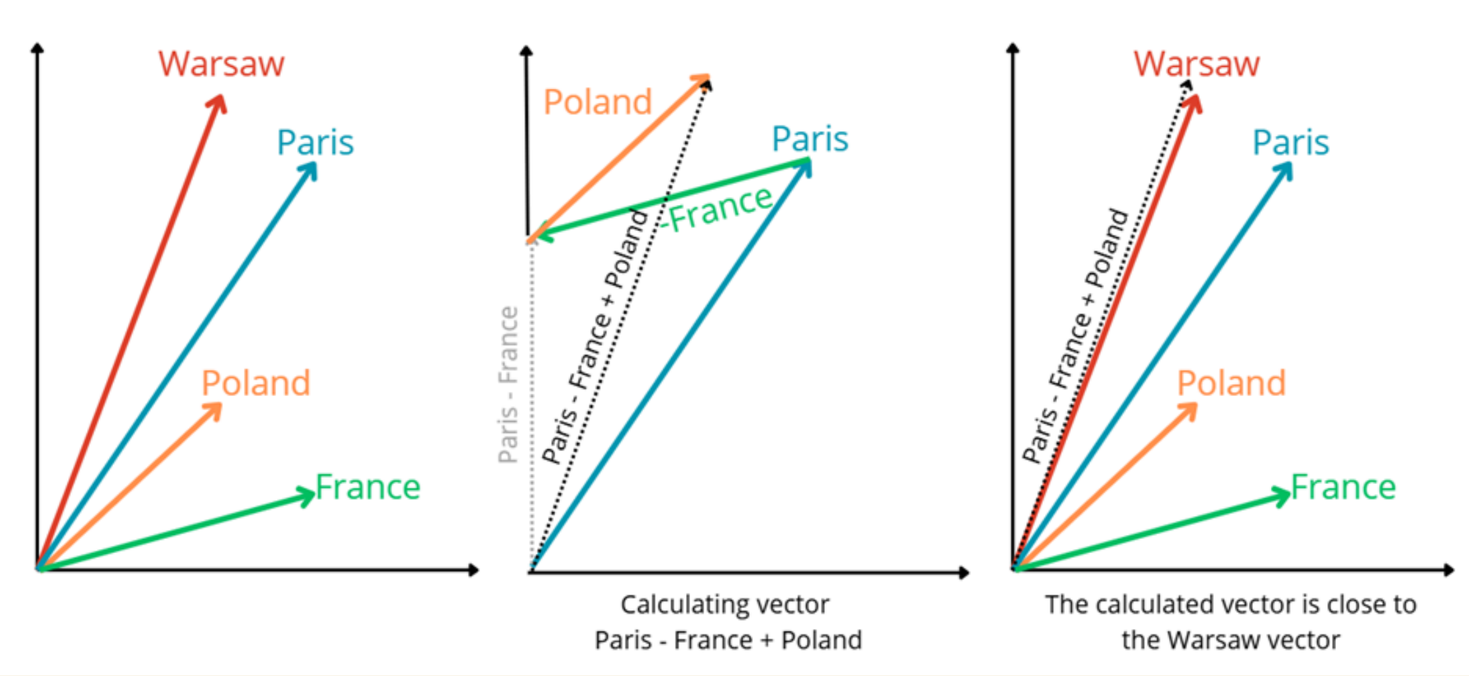
IMPLÉMENTATION DES SOLUTIONS

6

DIFFERENCES FRANÇAIS / ANGLAIS

SI JE POSE LA QUESTION EN DEUX LANGUES DIFFÉRENTES J'OBTIENS DES RÉSULTATS DIFFERENTS

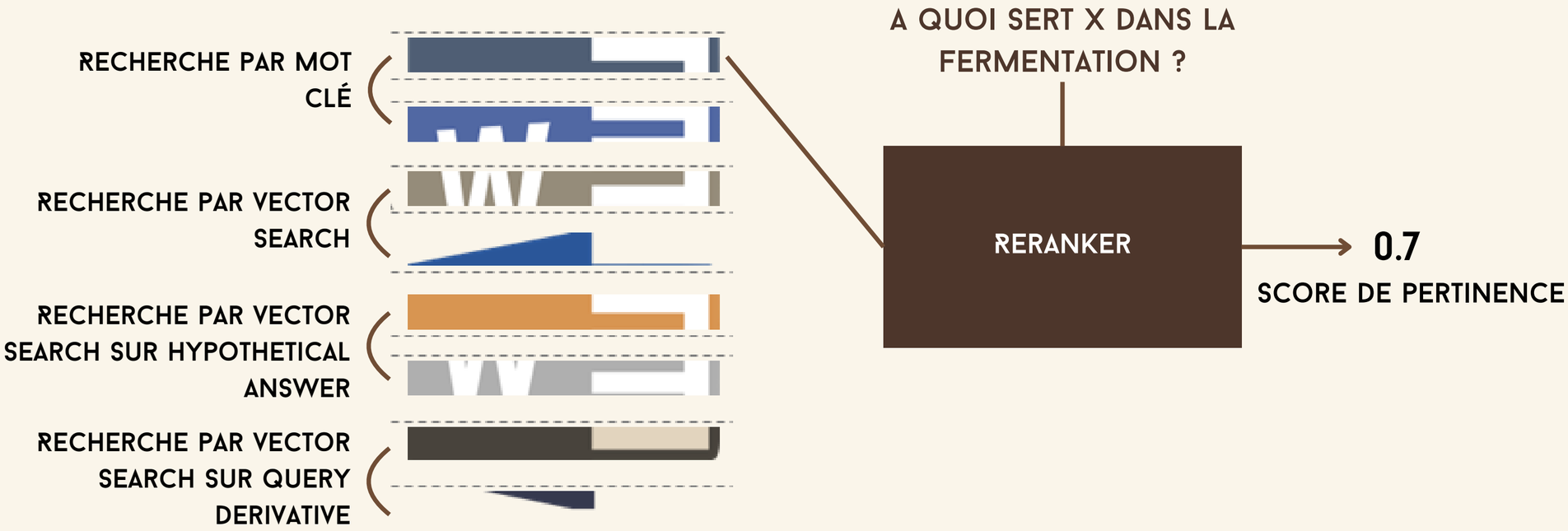
LE MODÈLE D'EMBEDDING QUE VOUS CHOISISSEZ DOIT CORRESPONDRE AU CAS D'USAGE.
IL EXISTE DES MODÈLES D'EMBEDDINGS AVEC DES CAPACITÉS SPÉCIFIQUES: MULTIMODALITÉ, MULTILINGUES



IMPLÉMENTATION DES SOLUTIONS

7 AVOIR UN SCORE DE PERTINENCE

APRÈS L'HYBRID SEARCH ET LES QUERY DERIVATIVES, NOUS AVONS RÉCUPÉRÉ BEAUCOUP DE DOCUMENTS. COMMENT SAVOIR QUELS SONT CEUX QUI SONT LES PLUS PERTINENTS POUR RÉPONDRE À LA QUESTION ?



IMPLÉMENTATION DES SOLUTIONS

8

HALLUCINATIONS

Compte tenu des paragraphes suivants:

Répond à la question:

Je veux en savoir plus sur X

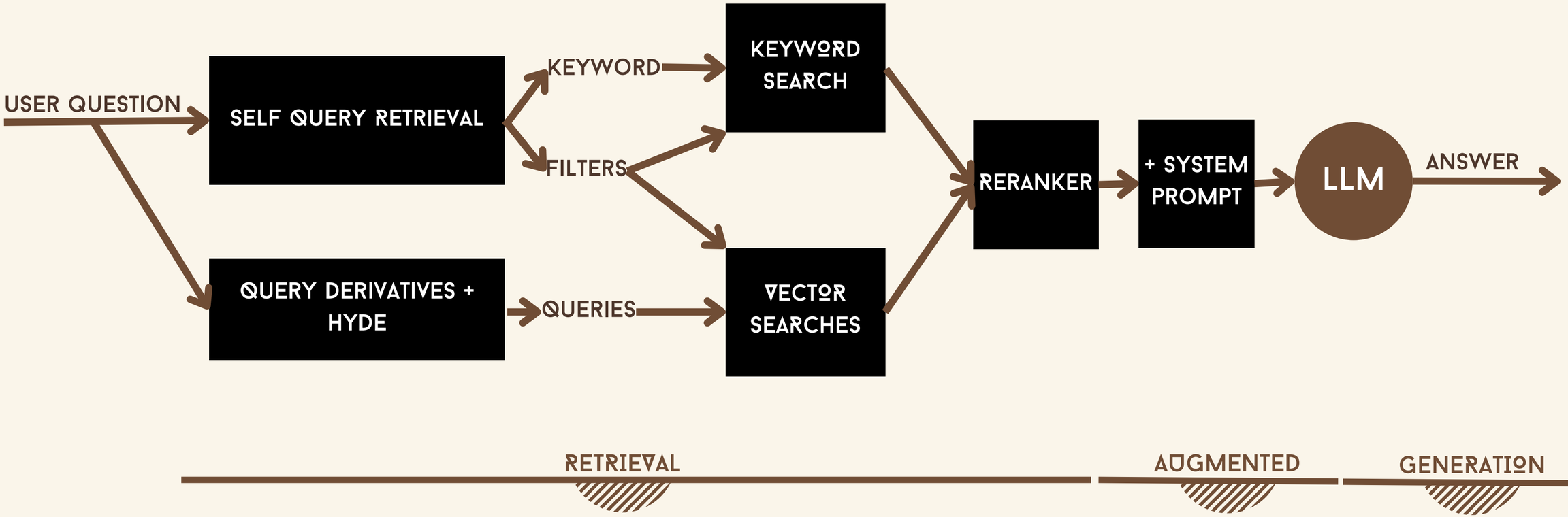
LES HALLUCINATIONS PEUVENT TOUJOURS ARRIVER MAIS ELLES PEUVENT ETRE RÉDUITES EN AJOUTANT UN PROMPT SYSTEM.

IL S'AGIT DU CONTEXTE QUE L'ON DONNE AU MODÈLE AVANT QU'IL NE RÉPONDE.

ON PEUT LUI SPÉCIFIER SON RÔLE, CE QU'IL PEUT ET NE PEUT PAS FAIRE, LUI FOURNIR DES EXEMPLES, LUI INDIQUER COMMENT AJOUTER DES SOURCES ETC.

UNE AUTRE SOLUTION SERAIT D'AJOUTER UN AUTRE APPEL AU MODÈLE OU UNE ÉTAPE DE REASONNING

RÉCAPITULATIF RAG PIPELINE



QUESTIONS ?