

LLM : RAG, CAG, KAG et ontologie

Trois façons d'alimenter un modèle de langage, et une façon d'organiser le savoir

Un LLM sait générer du texte, mais il faut souvent l'aider à accéder à la bonne information, au bon moment, et dans la bonne structure.

01

DOCUMENTS
RETROUVÉS

RAG — Retrieval-Augmented Generation

Un moteur de recherche trouve les passages pertinents dans les documents, puis le modèle répond à partir de ces sources.



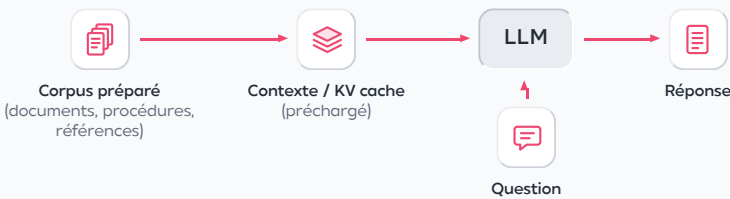
- **Usage** : FAQ, base documentaire, recherche d'information
- **Force** : réponses ancrées sur des sources
- **Limite** : peut rater le bon passage ; dépend de la qualité de la recherche

02

CONTEXTE
PRÉCHARGÉ

CAG — Cache-Augmented Generation

Le contexte utile est préparé ou préchargé à l'avance ; le modèle s'appuie sur ce contexte déjà disponible au lieu de relancer une recherche à chaque fois.



- **Usage** : corpus stable, procédures, documentation peu changeante
- **Force** : rapidité, aucune erreur de recherche, architecture simple
- **Limite** : le corpus doit tenir dans la fenêtre de contexte ; chaque mise à jour impose de tout recharger

03

BASE DE
CONNAISSANCES

KAG — Knowledge-Augmented Generation

Le modèle combine un graphe de connaissances (entités, relations, règles) avec les documents, et décompose la question en étapes de raisonnement.



- **Usage** : données complexes, relations métier, raisonnement multi-étapes
- **Force** : gère mieux les interrelations et la logique métier
- **Limite** : mise en place plus exigeante

04

MODÉLISATION
DU DOMAINE

Ontologie

Une ontologie décrit formellement les concepts d'un domaine, leurs propriétés, leurs relations et leurs règles : c'est une structure de sens, pas un moteur de réponse à elle seule.



- **Usage** : vocabulaire commun, modélisation, cohérence des données
- **Force** : clarifie le sens et normalise les relations
- **Limite** : doit être conçue et maintenue

Comparaison synthétique

Trois méthodes de génération et un modèle de connaissance, complémentaires.

Notion	Source principale	Structure	Cas idéal	Point faible
RAG	documents retrouvés	faible à moyenne	trouver la bonne source vite	peut rater le bon passage
CAG	contexte préchargé	faible à moyenne	corpus stable et réutilisé	limité par la taille du contexte
KAG	base de connaissances / graphe	forte	relations complexes et raisonnement	conception plus lourde
Ontologie	concepts et relations du domaine	très forte	définir le sens commun	ne répond pas seule

Relations et mise en perspective

