

Rapport de Benchmark

Évaluation des Stratégies

Étudiant : Arnaud Rambourg

Date : 16/01/2026

Modèle : Zephyr-7B-Beta

1. RÉSULTATS QUANTITATIFS GLOBAUX

Métrique	Stratégie A (LLM Seul)	Stratégie B (RAG)	Stratégie C (Extractif)
Score Mots-Clés	32 / 100	60 / 100	33 / 100
Latence Moyenne	16,66 s	6,68 s	0,13 s
Taux d'Erreurs	0%	0%	0%

2. STRATÉGIE A : LLM SEUL

Comportement : L'IA répond par elle-même sans consulter de base documentaire.

Analyse

- ❌ **Hallucinations** : Invente des procédures locales plausibles mais fausses.
- ❌ **Latence** : Très lente car elle génère de longs textes inutiles.
- ❌ **Précision** : Faible sur les questions administratives.

3. STRATÉGIE B : RAG (GAGNANTE)

Comportement : Recherche dans la FAQ puis génère une réponse basée sur les documents.

Analyse

- ✅ **Fiabilité** : Excellente sur les tarifs et les procédures réelles.
- ✅ **Politesse** : Maintient un ton professionnel.
- ⚠️ **Latence** : Équilibre correct (~6s).

4. STRATÉGIE C : EXTRACTIF

Comportement : Copie-colle le passage le plus probable du document.

Analyse

- 🚀 **Vitesse** : Instantanée.
- ❌ **Limites** : Pas de politesse, réponses trop courtes.
- ❌ **Contexte** : Incapable de reformuler.

5. RECOMMANDATION FINALE

La **Stratégie B (RAG)** est recommandée pour le déploiement.

Points d'attention

- Nettoyer la base FAQ pour éviter les confusions.
- Utiliser un modèle local pour la souveraineté.