Readme

# Telechargement dataset QA :

Script load\_dataset.py

# Telechargement LLM :

Donner les droits au fichier dowload\_llm : dowload\_llm.sh

# Creation rag : script « create\_rag”

* Simule les 2 cas : injection de contextes corrects ou incorrects.
* Compare la réponse du modèle à la réponse attendue.
* Produit un fichier rag\_faithfulness\_results.json contenant toutes les prédictions et leurs évaluations.

# Performance SCR/RCR : script « perf\_scr\_rcr”

* Lire ton fichier de résultats.
* Faire la séparation context\_mode = correct / incorrect.
* Calculer et afficher SCR et RCR.

# Que pourrait-on ajouter pour aller vers une *pipeline complète de publication scientifique* (optionnel) :

| **Étape** | **Utile pour** | **Implémentable** |
| --- | --- | --- |
| 🔄 Répéter plusieurs runs (stochasticity) | Avoir des barres d'erreur | ✅ très simple à ajouter |
| 📊 Grapher les performances | Générer des courbes SCR/RCR | ✅ avec matplotlib |
| 📄 Logger les hyperparamètres | Faciliter la reproductibilité | ✅ très simple |
| 📈 Multi-modèles | Comparer plusieurs modèles Ollama (phi3, mistral, llama3, etc) | ✅ très intéressant pour toi |
| 📂 Intégration pipeline unique | Lancer toute la chaîne d'un coup | ✅ possible avec un simple "main.py" orchestrateur |