Kaggle competition: MCQ Medicine

GoldoRag

Hackathon MistralxAlan

October 13, 2024

Challenge

- Multiple choice questions with possibly multiple answers
- Only 103 questions, 20 in private: High variance expected
- (Harsh) loss: correct only if all answers are correct

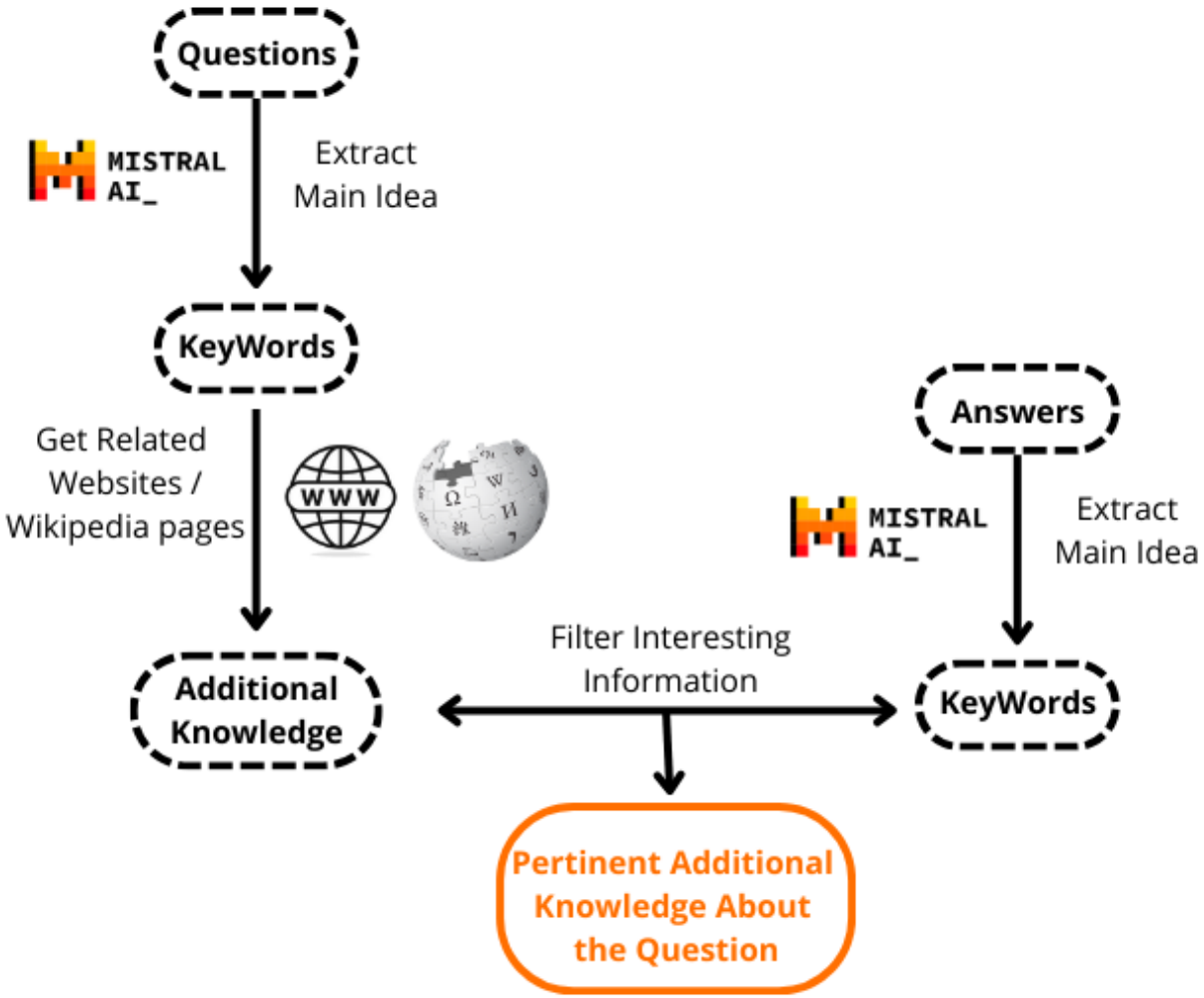
Dataset Analysis

Possible to make categories of questions according to their difficulties using: Blooms Taxonomy (1956)

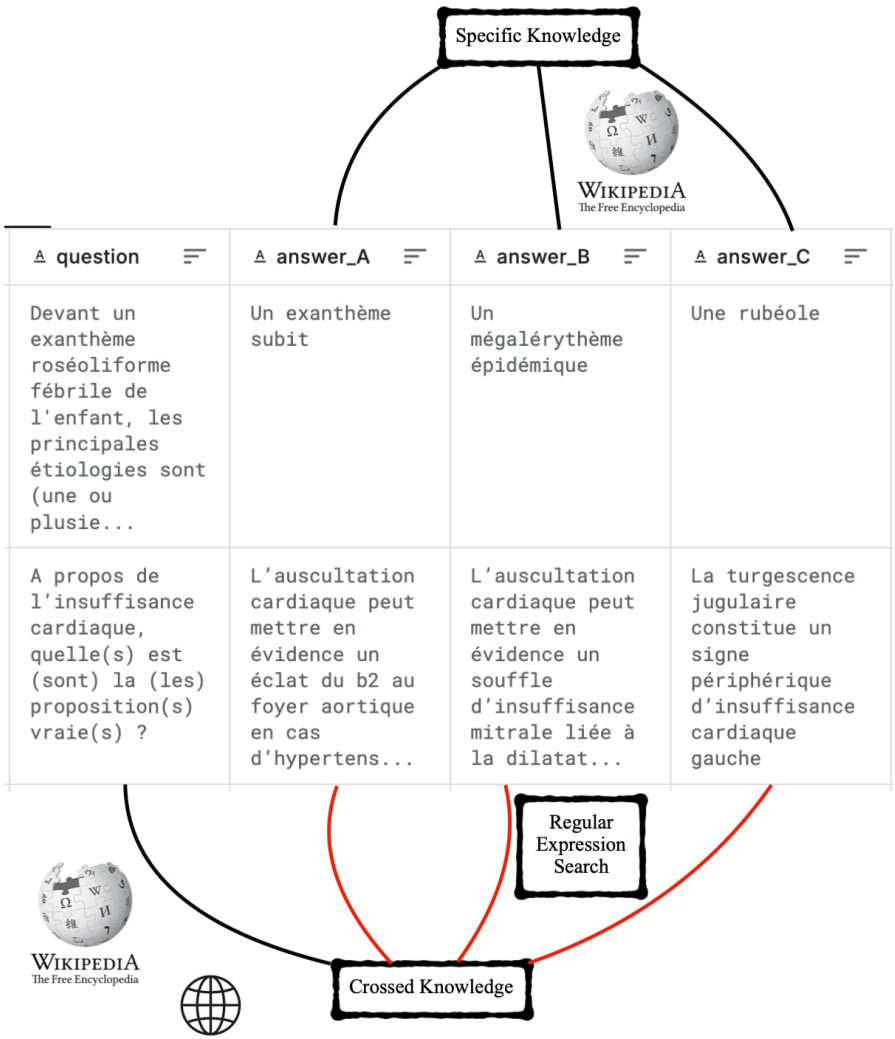
1. **Knowledge**: simple knowledge retrieval
2. Comprehension
3. Application
4. Analysis
5. **Evaluation**: complex scenario
6. Creation

Medical Domain: specific words with only one meaning, easy to search in a text. Simple NLP (regex, lemmatization, frequency of use) can help to easily access needed knowledge.

Idea



Idea



Results

Done:

- Almost no prompt engineering
- Augmented pertinent knowledge from web pages
- Only mistral-large-latest model used
- RAG with MedQA (scenario questions)

Majority voting on several outputs (self-consistency) improves the results but is expensive.

Final Score: 0.41250

Future work

To improve:

- More processing of the knowledge for clean inputs in the model
- Use an adapted fine-tuned and instruction prompt-tuning LLM on MedQA for instance (MedPalm, MedGemini...), for RAG and Question Answering.
- Or build it (Med-PaLM: Singhal, Karan, et al. "Towards expert-level medical question answering with large language models.")
- Build a bigger RAG dataset (open dataset, public domain books)

Thank you for this Hackathon!