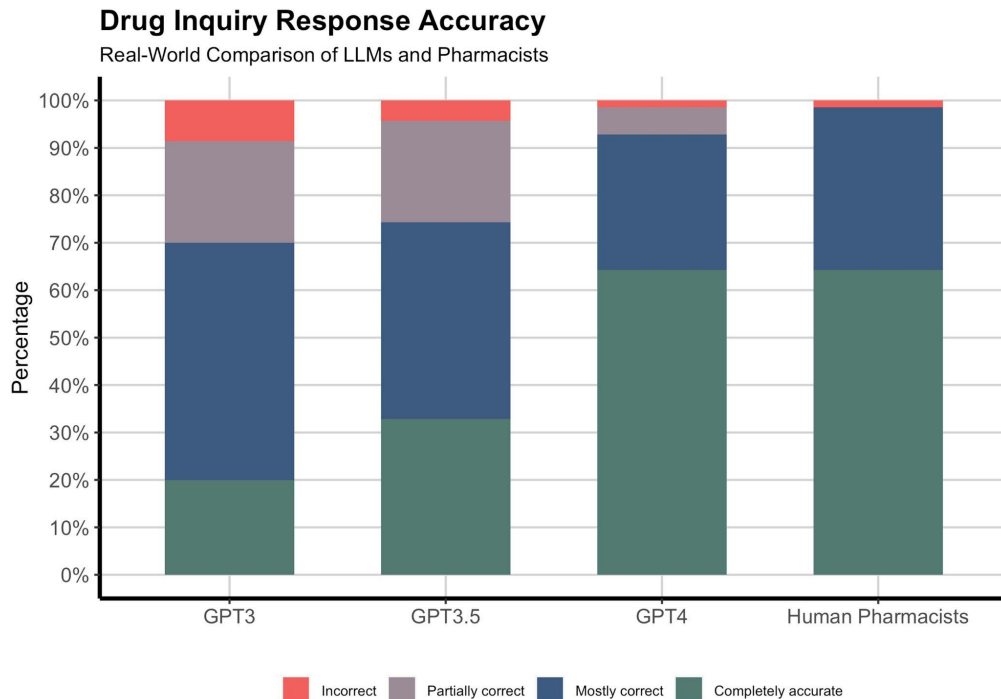


MD GRUPO 2

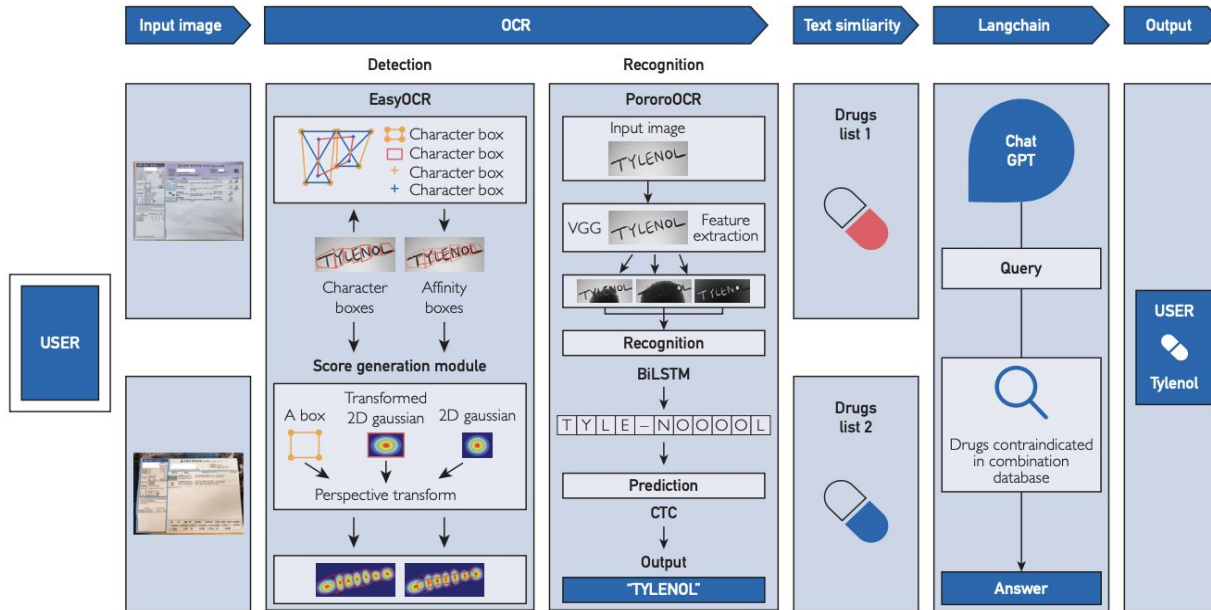
Apresentação Intermédia

Estado da Arte - Large Language Models



- **Perguntas sobre dosagem/administração:**
 - GPT-4: 75% de respostas completamente precisas
 - Farmacêuticos humanos: 62% de respostas completamente precisas
- **Perguntas sobre interações medicamentosas:**
 - GPT-4: 55,55% de respostas completamente precisas
 - Farmacêuticos humanos: 66,66% de respostas completamente precisas
- **Perguntas sobre uso terapêutico:**
 - GPT-4 e farmacêuticos humanos: 66% de respostas completamente precisas

Estado da Arte - Exemplos Acadêmicos

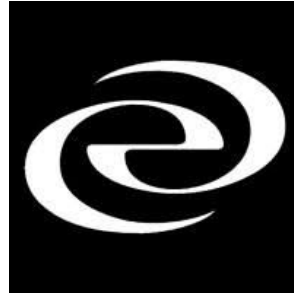


- Modelo usa **Optical Character Recognition** para extração de texto.
- Recorre a uma **LLM** (ChatGPT) para analisar os dados extraídos.

Kim, W. T., Shin, J., Yoo, I. S., Lee, J. W., Jeon, H. J., Yoo, H. S., ... & Kim, Y. J. (2024). Medication Extraction and Drug Interaction Chatbot: Generative Pretrained Transformer-Powered Chatbot for Drug-Drug Interaction. *Mayo Clinic Proceedings: Digital Health*, 2(4), 611-619.

Estado da Arte - Aplicações Comerciais

DocsBot



<https://docsbot.ai/industry/pharmacy>

<https://www.emaginehealth.com/blog/chathcp-ai-healthcare-biopharma/>

<https://ada.com/pt/>

Base de Conhecimento - Pipeline de Processamento

- **Fontes de Dados:**
 - **Fontes Académicas:** PubMed, SemanticScholar, EuropePMC.
 - **Fonte Geral:** Wikipedia.
 - etc¹.
- **Extração, Transformação e Carregamento dos Dados (ETL):**
 1. **Extração:** APIs extraem informação estruturada (*abstracts*, autores, etc.) das fontes selecionadas.
 2. **Transformação:** Utilizamos **spaCy** para NLP, processando e estruturando os dados em formato **JSON**.
 3. **Carregamento:** Os dados estruturados são armazenados numa **base de dados NoSQL** (MongoDB) para consulta e análise.
- **Utilização de (sci)spaCy:**
 - Extração de **entidades** (e.g., fármacos, suplementos, doenças), incluídas nos *abstracts*

¹ Futuramente, poderão ser introduzidas novas fontes (e.g., OMS, DrugBank)

Exemplo de Documento

```
_id: ObjectId('67d74ef2dd5efd1501446351')
title: "Anthocyanins: Modified New Technologies and Challenges."
▼ authors: Array (4)
  0: "Yang Lin"
  1: "Cong Li"
  2: "Lejuan Shi"
  3: "Lixia Wang"
year: 2023
source: "Europe PMC"
abstract: "Anthocyanins are bioactive compounds belonging to the flavonoid class ..."
▼ keywords: Array (4)
  0: "Anthocyanins"
  1: "Molecular modification"
  2: "Physical Modification"
  3: "Application Progress"
doi: "10.3390/foods12071368"
last_updated: "2023-03-23"
▼ spacy_entities: Array (1)
  ▼ 0: Array (2)
    0: "disease prevention"
    1: "DISEASE"
▼ spacy_matched_terms: Object
  ▼ DISEASE: Array (1)
    0: "disease prevention"
  ► SUPPLEMENT: Array (empty)
  ► PHARMACEUTICAL: Array (empty)
  ► MEDICAL_CONCEPT: Array (empty)
```

Desafios Futuros

1. Divergência de Opiniões entre Fontes

Problema: Dois ou mais artigos podem abordar o mesmo tema com conclusões diferentes.

- **Soluções:**

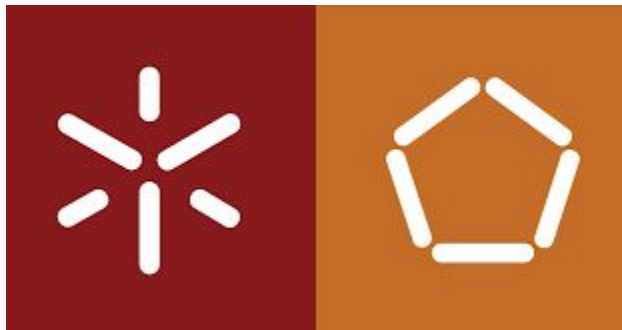
- Apresentar todas as perspectivas ao utilizador, indicando as fontes no chatbot.
- Utilizar bibliotecas como **Consensus** para identificar a posição predominante.

2. Melhoria na Extração e Normalização de Entidades

Problema: Entidades como fármacos podem ter múltiplas designações (ex: **Paracetamol vs. Acetaminofeno**).

- **Soluções:**

- Enriquecimento de dicionários semânticos no **spaCy**.
- Uso de bases de dados de sinónimos médicos



MD GRUPO 2

Apresentação Intermédia