



**HAL**  
open science

## Évaluation des capacités des grands modèles de langue à comprendre les dossiers médicaux de patients : Une étude approfondie de l'extraction et la recherche de données des patients

Jesus Lovon, Martin Mouysset, Jo Oleiwan, José G. Moreno, C. Damase-Michel,  
Lynda Tamine

### ► To cite this version:

Jesus Lovon, Martin Mouysset, Jo Oleiwan, José G. Moreno, C. Damase-Michel, et al.. Évaluation des capacités des grands modèles de langue à comprendre les dossiers médicaux de patients : Une étude approfondie de l'extraction et la recherche de données des patients. 20e Conférence en Recherche d'Information et Applications (CORIA) 32ème Conférence sur le Traitement Automatique des Langues Naturelles (TALN) 27ème Rencontre des Étudiants Chercheurs en Informatique pour le Traitement Automatique des Langues (RECITAL) Les 18e Rencontres Jeunes Chercheurs en RI (RJCRI), 2025, Marseille, France. pp.126-127. <hal-05330654>

**HAL Id: hal-05330654**

**<https://inria.hal.science/hal-05330654v1>**

Submitted on 26 Oct 2025

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

# Évaluation des capacités des grands modèles de langue à comprendre les dossiers médicaux de patients : Une étude approfondie de l'extraction et la recherche de données des patients

Jesús Lovón-Melgarejo<sup>1</sup> Martin Mouysset<sup>1</sup> Jo Oleiwan<sup>1</sup> Jose G. Moreno<sup>1</sup>  
Christine-Damase-Michel<sup>2</sup> Lynda Tamine<sup>1</sup>

(1) Université de Toulouse, IRIT UMR 5505, Toulouse, France

(2) Centre Hospitalier Universitaire de Toulouse,  
CERPOP INSERM UMR 1295 - équipe SPHERE,

Faculté de Médecine Université de Toulouse, Toulouse, France

(1) `first.last@irit.fr`, (2) `first.last@univ-tlse3.fr`

## RÉSUMÉ

---

Les dossiers médicaux de patients (DMP) posent des défis uniques, notamment la présence de dépendances contextuelles cachées entre les caractéristiques médicales avec un niveau élevé de dimensionnalité et de disparité des données. Ce papier présente la première étude sur les capacités des grands modèles de langage à comprendre les DMP en vue d'en extraire ou rechercher des données. Nous menons des expérimentations approfondies en utilisant l'ensemble de données MIMICSQL pour explorer l'impact de la structure des *prompts*, des instructions, du contexte et des démonstrations de deux grands modèles de langue, Llama2 et Meditron, sur la performance des tâches d'extraction et recherche d'information. À travers des analyses quantitatives et qualitatives, nos résultats montrent que les méthodes optimales de sélection et de sérialisation des dossiers de patients peuvent améliorer la performance des tâches jusqu'à 26, 79% par rapport aux approches naïves. De même, les scénarios d'apprentissage en contexte avec sélection d'exemples pertinents améliorent la performance d'extraction de données de 5, 95%. Sur la base des résultats de notre étude, nous proposons des lignes directrices destinées à faciliter la conception de modèles basés sur les grands modèles de langue pour supporter la recherche d'information en santé. Les jeux de données et le code sont disponibles via ce lien<sup>1</sup>. Ceci est le résumé de l'article "*Evaluating LLM Abilities to Understand Tabular Electronic Health Records : A Comprehensive Study of Patient Data Extraction and Retrieval*" publié comme papier long à ECIR 2025 (Lovón-Melgarejo *et al.*, 2025).

## ABSTRACT

---

### **Evaluating LLM Abilities to Understand Tabular Electronic Health Records : A Comprehensive Study of Patient Data Extraction and Retrieval**

Electronic Health Record (EHR) tables pose unique challenges among which is the presence of hidden contextual dependencies between medical features with a high level of data dimensionality and sparsity. This study presents the first investigation into the abilities of LLMs to comprehend EHRs for patient data extraction and retrieval. We conduct extensive experiments using the MIMICSQL dataset to explore the impact of the prompt structure, instruction, context, and demonstration, of two backbone LLMs, Llama2 and Meditron, based on task performance. Through quantitative and qualitative analyses, our findings show that optimal feature selection and serialization methods can

---

1. <https://github.com/jeslev/llm-patient-ehr>

enhance task performance by up to 26.79% compared to naive approaches. Similarly, in-context learning setups with relevant example selection improve data extraction performance by 5.95%. Based on our study findings, we propose guidelines that we believe would help the design of LLM-based models to support health search. Datasets and code are available online. This is the summary of the published paper “Evaluating LLM Abilities to Understand Tabular Electronic Health Records : A Comprehensive Study of Patient Data Extraction and Retrieval” at ECIR 2025 (Lovón-Melgarejo *et al.*, 2025).

---

**MOTS-CLÉS** : grands modèles de langue, dossier médical de patient (DMP), données tabulaires, recherche d’information, extraction des données.

**KEYWORDS**: large language models, electronic health record (EHR), tabular data, information retrieval, information extraction .

---

ARTICLE : **Accepté à ECIR 2025.**

---

## Références

LOVÓN-MELGAREJO J., MOUYSSSET M., OLEIWAN J., MORENO J. G., DAMASE-MICHEL C. & TAMINE L. (2025). Evaluating llm abilities to understand tabular electronic health records : A comprehensive study of patient data extraction and retrieval. In *European Conference on Information Retrieval*, p. 153–168 : Springer.