Evaluating LLM-based reranking method for medication term normalization

Tavleen Singh, PhD[1], Qiang Wei, PhD[1], Chuck Levecke, BS[1], Jingqi Wang, PhD[1]

[1]IMO Health, Houston, Texas, US

INTRODUCTION

- Medication normalization ensures drug names are consistent and accurate across systems, promoting interoperability and supporting improved patient care.
- In this study, we evaluated the use of large language models (LLMs) to select the most accurate RXCUI for a given medication term through retrieval-augmented generation using IMO Precision Normalize[1].
- IMO Precision Normalize[1] standardizes clinical data from diverse sources and maps it to industry standard, clinically validated terminologies.
- While the correct RXCUI is often included in the top search results, it is not always the first match, requiring manual review by the end user to identify the best-fitting code.

RESULTS

- Table 1 provides the current accuracy of the production version of IMO Precision Normalize1 (using threshold score of 0.5) as baseline and compares it to the LLM-based reranking method used in this study. The correct RXCUI was present in the reranking candidate list for 90% (290/322) of the medication terms.
- The overall accuracy of finding the best candidate improved by 12.4% using GPT-4o model in comparison to the using IMO Precision Normalize[1] only.
- For example, GPT-4o model correctly identified the match for input term: 'AMLODIPINE, NORVASC TAB 5MG' as 'amlodipine 5 mg oral tablet [norvasc]' from the candidate list whereas IMO Precision Normalize[1] returned 'amlodipine 5 mg oral tablet' as its top match which misses the brand name information.

OBJECTIVES

- Identify how LLMs enhance RXCUI selection in medication normalization workflows.
- Understand the role of retrieval-augmented generation and prompt design in boosting accuracy.

METHODS

- Firstly, the three configurations of IMO Precision Normalize[1] were used to maximize the likelihood of finding the correct RXCUI in the top search results by the LLMs.
- These configurations include – 1) direct title matching, 2) parts matching, and 3) parts matching with 'component only' – i.e., with all dosage form, route and strength information stripped out.
- Secondly, for each configuration, we ran IMO Precision Normalize[1] to return the top 20 results – each containing an RXCUI, term type, title and potentially a supplementary list of synonyms or prescribable names, and a normalize score.
- A system prompt consisting of a set of editorial rules was designed for the LLM to select the best choice of RxNorm title from the supplied candidate list for the given input medication term.
- The user prompt provided the input medication term along with its list of candidates generated from previous steps.
- Lastly, we compared the performance of GPT-4o vs. GPT-4o-mini for the given task using a dataset of 436 sample terms (internal IMO dataset): 322 of 436 have RXCUI codes ("Labeled Not Null" in Table 1), and the other 114 terms can't be mapped to RXCUI codes ("Labeled Null" in Table 1).

Method	Overall Accuracy	Labeled Null Accuracy	Labeled Not Null Accuracy	Configuration associated with matching RXCUI when LLM is correct for not null terms		
				Direct title matching	Parts matching	Component only matching
IMO Precision Normalize	274/436 (62.8%)	70/114 (61.4%)	204/322 (63.4%)			
GPT-4o	328/436 (75.2%)	94/114 (82.5%)	234/322 (72.7%)	223/234 (95.3%)	213/234 (91.0%)	111/234 (47.4%)
GPT-4o-mini	277/436 (63.5%)	58/114 (50.9%)	219/322 (68.0%)	207/219 (94.5%)	200/219 (91.3%)	106/219 (48.4%)

DISCUSSION

- This study highlights the potential of using LLMs for medication term normalization.
- By incorporating editorial rules and leveraging multiple search configurations, the LLM-enhanced approach significantly improved selection accuracy by 12.4%, offering a scalable and effective solution for reranking candidate RXCUI codes returned by IMO Precision Normalize[1].
- Future work will focus on enhancing the system prompt for better handling of null terms and exploring few-shot learning examples to support improved performance from lightweight LLMs.

REFERENCES

1) https://www.imohealth.com/imo-precision-normalize/

CONTACT

tsingh@imohealth.com

www.PosterPresentations.com