AutoCriteria: A GENERALIZABLE CLINICAL TRIAL ELIGIBILITY CRITERIA EXTRACTION SYSTEM POWERED BY LARGE LANGUAGE MODELS

Surabhi Datta, Kyeryoung Lee, Hunki Paek, Frank J. Manion, Nneka Ofoegbu, Jingcheng Du, Ying Li, Liang-Chin Huang, Jingqi Wang, Bin Lin, Hua Xu, Xiaoyan Wang

MELAX TECH
Now part of Intelligent Medical Objects

MSR126

INTRODUCTION

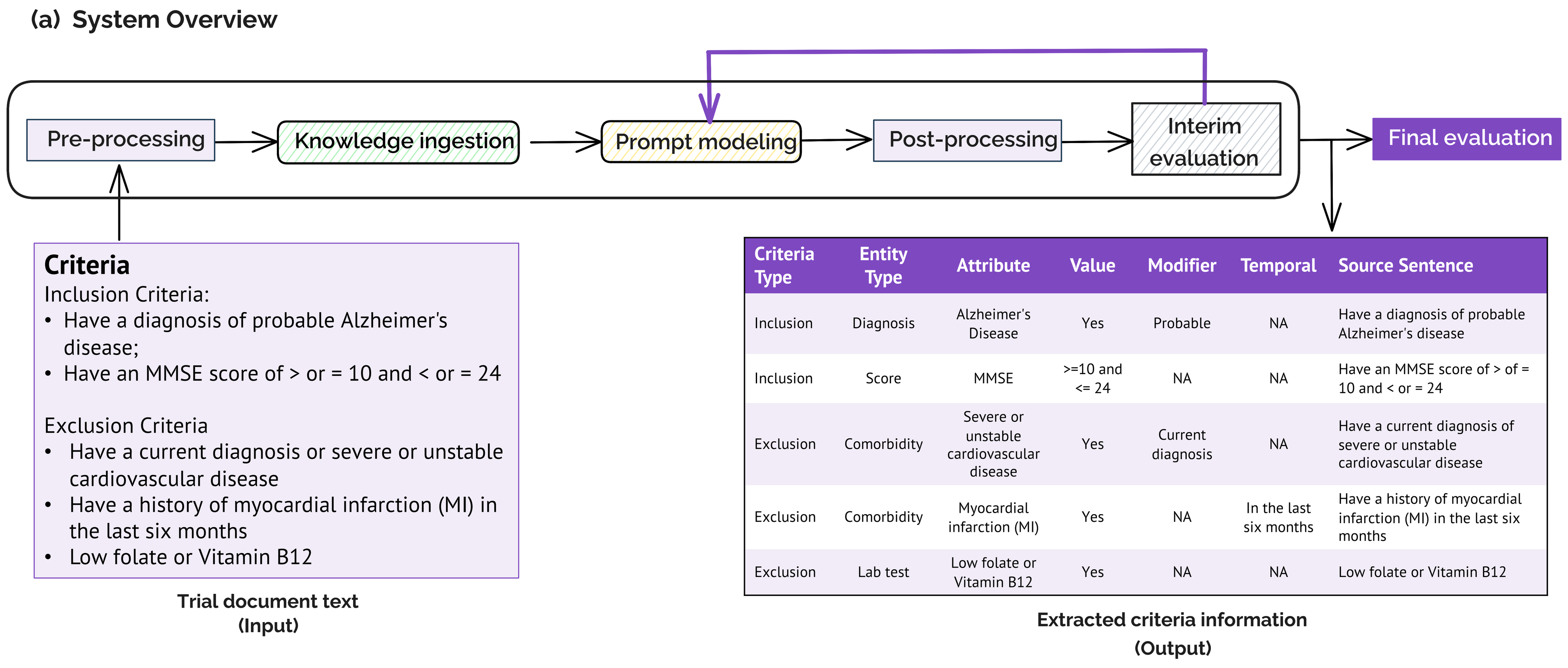
- Eligibility criteria in clinical trials play a pivotal role in patient recruitment and in safety and treatment evaluation, resulting in improved patient care
- Natural language processing (NLP) techniques offer the potential to enhance the efficiency of clinical trial studies by automatically extracting eligibility criteria[1]
- Existing NLP approaches face challenges in capturing fine-grained criteria within a given text and may lack applicability across various disease areas
- Recently, large language models such as ChatGPT and GPT-4 have gained traction in the open NLP domain as well as biomedical and clinical domains[2,3]

OBJECTIVE

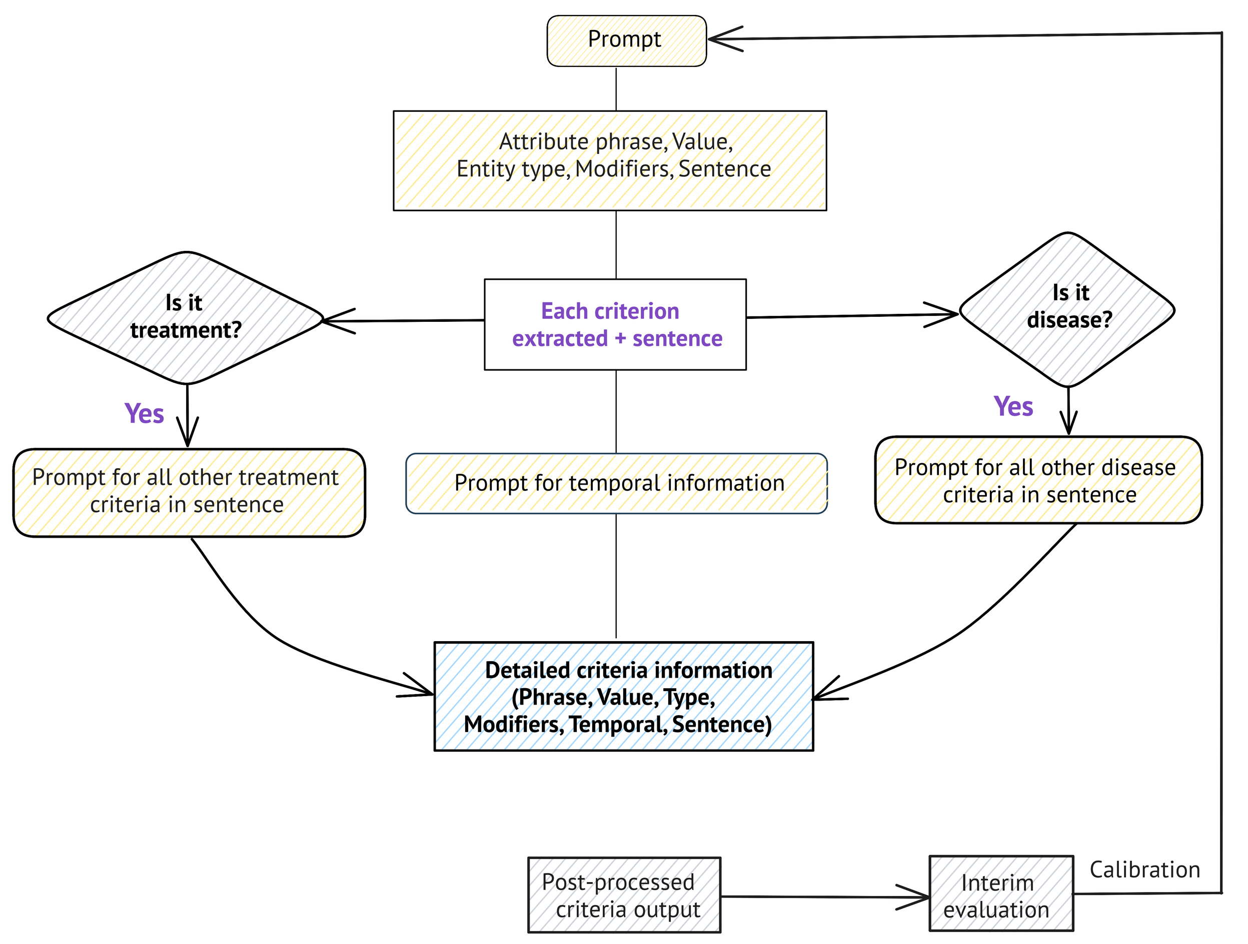
Our aim is to develop a system that automatically extracts eligibility criteria, emphasizes contextual attributes, and can handle diverse diseases utilizing a cutting-edge large language model

AUTOCRITERIA: SYSTEM OVERVIEW AND COMPONENTS

(a) System Overview



(b) Prompt Modeling



Pre-processing

- We split the raw criteria text for each trial document into two parts - Inclusion and Exclusion, we then split each of these parts into smaller chunks and run AutoCriteria on each chunk

Knowledge ingestion

- We identified the ontology of key criteria entities and attributes for each disease with the help of knowledge experts. This knowledge is also leveraged in prompt modeling

Prompt modeling

- We created two separate comprehensive prompts, one for Inclusion and another for Exclusion
- Each prompt consists of three main components – 1) a general instruction, 2) the Inclusion or Exclusion criteria text, and 3) the query that asks about criteria attributes (phrases), their corresponding values, modifier information, entity types, and source sentences

Post-processing

- We processed the GPT-4 responses to handle any inconsistencies in the model output

Interim evaluation

- We evaluated the prompts manually, and iteratively calibrated them using expert feedback for every disease

ELIGIBILITY CRITERIA EXTRACTION RESULTS ON 180 TRIAL DOCUMENTS

Table 1: AutoCriteria results in extracting criteria phrases.

Disease	Precision (%)	Recall (%)	F1
Breast Cancer	87.26	81.17	84.10
Multiple Myeloma	85.53	86.83	86.18
Alzheimer's	94.54	92.38	93.45
NASH	95.08	95.81	95.44
Crohn's	87.21	88.30	87.75
Ulcerative Colitis	91.85	92.99	92.42
SCD	90.46	90.15	90.30
HPAH	87.39	90.23	88.79
HoFH	90.38	88.01	89.18
All	89.62	89.23	89.42

Table 2: AutoCriteria accuracy (%) in extracting contextual details.

Disease	Attribute + Value	Entity type + Attribute + Value	Entity type + Attribute + Value + Temporal	Entity type +Attribute +Value +Temporal + Modifier
Breast Cancer	76.86	74.09	71.51	67.21
Multiple Myeloma	81.68	73.96	68.74	66.54
Alzheimer's	91.05	90.67	88.00	85.14
NASH	91.87	91.61	88.56	86.28
Crohn's	85.10	83.43	81.20	78.83
Ulcerative Colitis	88.01	86.76	86.29	83.64
SCD	87.10	85.91	85.91	83.36
HPAH	90.00	89.42	89.19	86.05
HoFH	86.70	84.83	84.27	82.21
All	85.90	83.37	81.09	78.95

DISCUSSION

- AutoCriteria system not only identifies key criteria but also extracts contextual details including negation/temporal information and classifies entity types
- Our system is highly adaptable to various diseases, including Cancers, Autoimmune diseases, Alzheimer's, NAS, and other rare diseases, highlighting its generalizability
- Strengths of AutoCriteria include:
  - ✓ understands context including exceptions or "if" condition
  - ✓ captures different arm/cohort information
  - ✓ handles the same criteria entity extraction with different values and modifiers
- Limitations of AutoCriteria include:
  - ✓ does not handle sophisticated criteria conditions such as "at least one of these criteria"
  - ✓ sometimes makes mistakes in differentiating the main criteria entities from modifier entities

CONCLUSION

- We have developed a generalizable GPT-based system that can identify granular eligibility criteria information from clinical trial documents across a variety of disease domains without requiring manual annotations
- Our adaptable prompts allow easy customization for new diseases without manual training data, facilitating large-scale criteria analysis
- With enhanced granularity, improved accuracy and generalizability, AutoCriteria has the potential to significantly streamline the clinical trial initiation and conduct process, ultimately reducing time requirements

REFERENCES

[1] Kang, T., Zhang, S., Tang, Y., Hruby, G. W., Rusanov, A., Elhadad, N., & Weng, C. (2017). EliIE: An open-source information extraction system for clinical trial eligibility criteria. Journal of the American Medical Informatics Association, 24(6), 1062-1071.

[2] Singhal, K., Azizi, S., Tu, T., Mahdavi, S. S., Wei, J., Chung, H. W., ... & Natarajan, V. (2023). Large language models encode clinical knowledge. Nature, 620(7972), 172-180.

[3] Kung, T. H., Cheatham, M., Medenilla, A., Sillos, C., De Leon, L., Elepaño, C., ... & Tseng, V. (2023). Performance of ChatGPT on USMLE: Potential for AI-assisted medical education using large language models. PLoS digital health, 2(2), e0000198.

CONTACT

Please contact the first authors via email: Surabhi Datta: sdatta@imohealth.com, Kyeryoung Lee: klee@imohealth.com