

LLM-based Information Extraction from Clinical Notes for Downstream Healthcare Applications

Georgios Peikos

Researcher (RTDA)

University of Milano-Bicocca, Milan, Italy

Department of Informatics, Systems, and Communication (DISCo)

Information and Knowledge Representation, Retrieval and Reasoning Lab (IKR3)



Dipartimento di
Informatica, Sistemistica
e Comunicazione



information and knowledge
representation, retrieval, and
reasoning LAB



Today's Talk

PART I

Understanding Clinical Texts and Their Uses

Why NLP is Essential for Unlocking Clinical Data

Overview of NLP Techniques for Clinical Texts

PART II

NLP for Downstream Healthcare Applications

Part I: Clinical Texts & NLP

Clinical Texts – Semi-structured

Lab & Diagnostic Reports

Semi-structured result of lab tests, e.g., blood exams

Often reports quantities, in tabular form

Often comes with comments/observations

Labsmart Software
Sample Letterhead

+91 93183 13723
info@labsmartlis.com
https://www.labsmartlis.com/

Mr. Rajesh Varma
Age / Sex : 26 YRS / M
Referred by : Dr. Aditya Thakur (M.B.B.S)
Reg. no. : 1028
Collected at : 123, MG Road, Pune, Maharashtra – 411001

Registered on : 03/03/2025 08:25 PM
Collected on : 03/03/2025
Received on : 03/03/2025
Reported on : 03/03/2025 08:25 PM

1028

Scan to download

HAEMATOLOGY
COMPLETE BLOOD COUNT (CBC)

TEST	VALUE	UNIT	REFERENCE
HEMOGLOBIN	15	g/dl	13 - 17
TOTAL LEUKOCYTE COUNT	5,100	cumm	4,800 - 10,800
DIFFERENTIAL LEUCOCYTE COUNT			
NEUTROPHILS	79	%	40 - 80
LYMPHOCYTE	L 18	%	20 - 40
EOSINOPHILS	1	%	1 - 6
MONOCYTES	L 1	%	2 - 10
BASOPHILS	1	%	< 2
PLATELET COUNT	3.5	lakhs/cumm	1.5 - 4.1
TOTAL RBC COUNT	5	million/cumm	
HEMATOCRIT VALUE, HCT	42	%	
MEAN CORPUSCULAR VOLUME, MCV	84.0	fL	
MEAN CELL HAEMOGLOBIN, MCH	30.0	Pg	27 - 32
MEAN CELL HAEMOGLOBIN CON, MCHC	H 35.7	%	31.5 - 34.5

Clinical Notes:
A complete blood count (CBC) is used to evaluate overall health and detect a wide range of disorders, including anemia, infection, and leukemia. There have been some reports of WBC and platelet counts being lower in venous blood than in capillary blood samples, although still within these reference ranges.

Possible causes of abnormal parameters:

	High	Low
RBC, Hb, or HCT	Dehydration, polycythemia, shock, chronic hypoxia	Anemia, thalassemia, and other hemoglobinopathies
MCV	Macrocytic anemia, liver disease	Microcytic anemia
WBC	Acute stress, infection, malignancies	Sepsis, marrow hypoplasia
Platelets	Risk of thrombosis	Risk of bleeding

Mr. Sachin Sharma 123
DMLT, Lab Incharge

Dr. A. K. Asthana
MBBS, MD Pathologist

NOT VALID FOR MEDICO LEGAL PURPOSE

Work timings: Monday to Saturday, 10 am to 5 pm

This report was created in the Labsmart software which has inbuilt interpretations. All interpretations are copyright of Labsmart

Clinical Texts – Unstructured

Psychotherapy Progress Note

Clinical Notes

Admission notes

Progress notes

Discharge summaries

Procedure notes

Patient is a **28-year-old male** presenting for a **follow-up psychotherapy session**. He appeared on time, casually dressed, and was cooperative throughout.

He reported feeling "**overwhelmed**" due to increasing **work pressure** and a recent **argument with his partner**. **Sleep** has been disrupted – he reports waking multiple times during the night and feeling tired during the day. **Appetite** has decreased, and he has unintentionally lost **3–4 pounds**.

He **denied suicidal ideation**, but acknowledged **passive thoughts of hopelessness**, stating: "*I just don't want to deal with anything anymore.*" His **effect was constricted**, and **mood** was described as "**flat.**" **Thought process** remained **logical and goal-directed**, though he frequently returned to themes of **guilt** and **inadequacy**.

He described a **panic episode** triggered by the family conflict, involving **chest tightness**, **palpitations**, and **shortness of breath**. He utilized **deep breathing** to partially control the symptoms. He stated this was **more intense than previous panic attacks**, though similar in pattern.

He has been **compliant with sertraline 50 mg daily** and occasionally uses **clonazepam 0.5 mg PRN** for anxiety spikes. He is **not currently using illicit substances**. He **drinks alcohol socially** (1–2 drinks on weekends) and **quit smoking 2 years ago**.





Family history is significant for **major depressive disorder** in his **mother** and **panic disorder** in a **maternal aunt**. He has expressed concern about "ending up like them."

He reported that therapy is **somewhat helpful**, but he continues to struggle with **emotional regulation**, **relationship instability**, and **low self-esteem**. He expressed ambivalence about staying in the relationship, which remains a key stressor.





Plan is to continue **weekly CBT**, focus on **cognitive restructuring**, and introduce **behavioral activation strategies**. He was encouraged to keep a **daily mood and sleep log**. A **safety plan was reviewed** and remains active.

Clinical Texts – Primary & Secondary Uses

Primary Uses (*Direct patient care*)

-  Facilitate Patient Care
-  Diagnosis & Treatment
-  Medication Management
-  Monitoring & Follow-up

Secondary Uses (*Beyond immediate care*)

- Research 
- Clinical Trials 
- Billing & Administration 
- Long Term Education via Data Analysis 

To support secondary uses in an **automated**, **cost-effective**, and **time-efficient** manner, we must rely on **NLP approaches**.

Clinical Notes – Necessity of NLP

Jane M. Reynolds

8765432-1

06/17/2025

HISTORY OF PRESENT MENTAL ILLNESS:

Ms. Reynolds is a 35-year-old female with a history of **generalized anxiety disorder (GAD)** and recurrent major depressive disorder (**MDD**). **She** presents today for routine follow-up. **She reports** increased feelings of worry and restlessness over the **past two weeks**, primarily related to work stress. Sleep has been somewhat disrupted, with difficulty falling asleep 3–4 nights per week. She denies suicidal ideation, hallucinations, or manic symptoms. She reports taking her medications as prescribed and is engaged **in weekly therapy sessions**. Energy and motivation have been slightly lower than usual, but she remains functional in daily activities. Family history is notable for depression in her mother and anxiety in a paternal uncle. This episode seems to have been precipitated by the family conflict, similar to prior instances where interpersonal tension contributed to symptom flare-ups.

CURRENT MEDICATIONS:

Updated in **EMR**. Medications include sertraline 100 mg daily, hydroxyzine 25 mg **PRN** for acute anxiety, and melatonin 3 mg nightly. **Plans to continue current regimen and monitor symptoms.**

ALLERGIES:

No known drug allergies.

SOCIAL HISTORY:

Never smoker. Drinks socially, 1–2 drinks per week. Works as a graphic designer. Lives alone, good family support.

CONSTITUTIONAL: Mild fatigue, no weight changes.

NEUROLOGICAL: No dizziness or headaches.

PSYCHIATRIC: Reports anxiety, occasional restlessness, no suicidal or homicidal ideation.

PHYSICAL EXAMINATION:

VITAL SIGNS: Weight 143 lbs, blood pressure 112/70, pulse 72.

HEENT: **No abnormalities noted.**

CARDIOVASCULAR: RRR, **no murmurs.**

PSYCHIATRIC: Mood mildly anxious, affect congruent. Thought process linear, no delusions or perceptual disturbances. Insight and judgment intact.

Linguistic Challenges

Abbreviations

Negations

Temporal Reasoning

Measurements

Coreferences

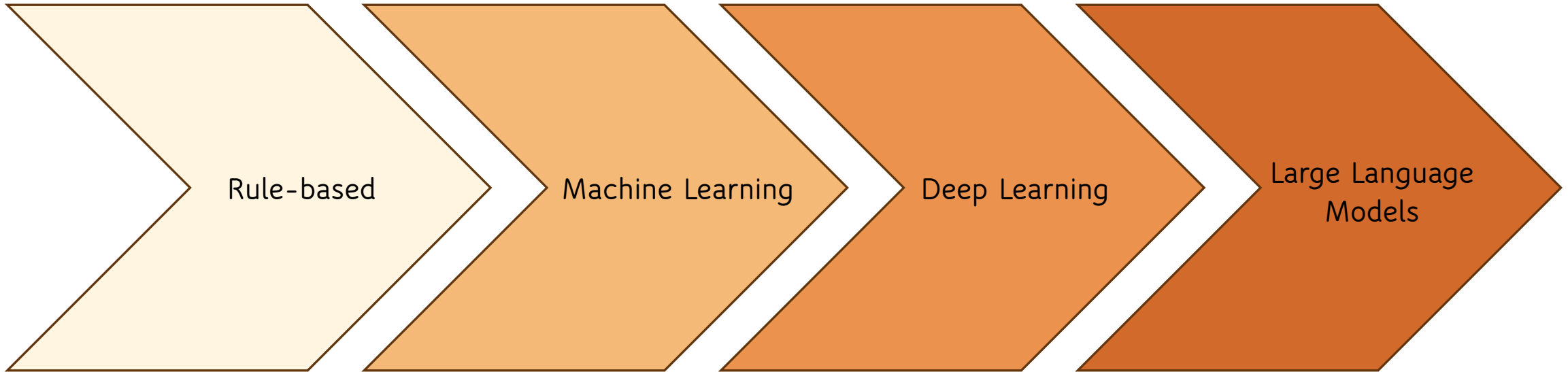
Additional Challenges

Shortness

Misspellings

Personalized Writing

Clinical NLP – NLP Methods



Clinical NLP Methods – Rule-based

✓ Strengths

- High precision in well-defined contexts*
- Transparent and interpretable
- No training
- Easy to deploy for narrow tasks
- Rules can be written by experts
- Deterministic behavior

✗ Weaknesses

- Struggles with ambiguity and complex context
- Sensitive to even small language variability
- Poor scalability to new institutions/departments
- Labor-intensive to maintain and extend
- Rules may override each other
- Limited coverage

Example Tools: ConText, NegEx, MedLEE, cTAKES, QuickUMLS, medspaCy

Simple, transparent, and effective – until the rules break.

Clinical NLP Methods – Machine Learning

✓ Strengths

Less manual programming

Can **generalize** better than rules

Handles **variation** in phrasing

Moderate resource requirements

✗ Weaknesses

Requires feature engineering

Limited **contextual understanding**

Lack of Interpretability

Sensitive to Data Quality

Data annotation can be expensive

Example Tools: SVMs, CRFs, NLTK

Smarter than rules – but hungry for data.

Clinical NLP Methods – Deep Learning

✓ Strengths

Superior performance on many clinical tasks

No need for manual feature engineering

More effective at capturing context and semantics

✗ Weaknesses

Needs larger labeled datasets

Data annotation is more expensive

Less interpretable than rule-based or standard ML

Task specificity; e.g., one model can do well in drug Extraction

Higher computational cost and resources

Example Tools: RNNs, Clinical BERT, BioBERT

Strong and flexible – but hungry for data and less interpretable.

Clinical NLP Methods – LLMs

✓ Strengths

No need for fine-tuning

Few-shot and zero-shot capabilities

General-purpose: one model for many tasks

Domain adaptation with minimal data

Handles long-range dependencies

Handles some of the issues of clinical texts*

✗ Weaknesses

Expensive to train

Black-box behavior

Prone to **hallucination** or factual errors

Is a significant investment for medical organizations

Inherited Biases

Regulatory, ethical, and **privacy** concerns

Example Tools: Med-PaLM, GPT, BioGPT, GatorTron

One model, many tasks, flexible and fluent – yet not always right.

Part II: NLP in Downstream Healthcare Applications

Healthcare Application – Enrollment in Clinical Trials

Task Description

Identify and **match eligible patients** to **appropriate clinical trials** based on clinical information such as admission notes

Importance

Significant number of clinical trials are **delayed** or **closed** due to **insufficient patient enrollment**

Manual matching is **inefficient, time-consuming**, and prone to human error

Solution

Develop better retrieval systems that, given a patient's clinical information, can automatically retrieve eligible clinical trials

Leverage **NLP techniques** to extract patient related information from free-text patient records to enhance retrieval effectiveness

Clinical Trial Enrollment – Approach

User: **Selects** a patient's privately store clinical information

System: Fetches the data from the **Private Database**

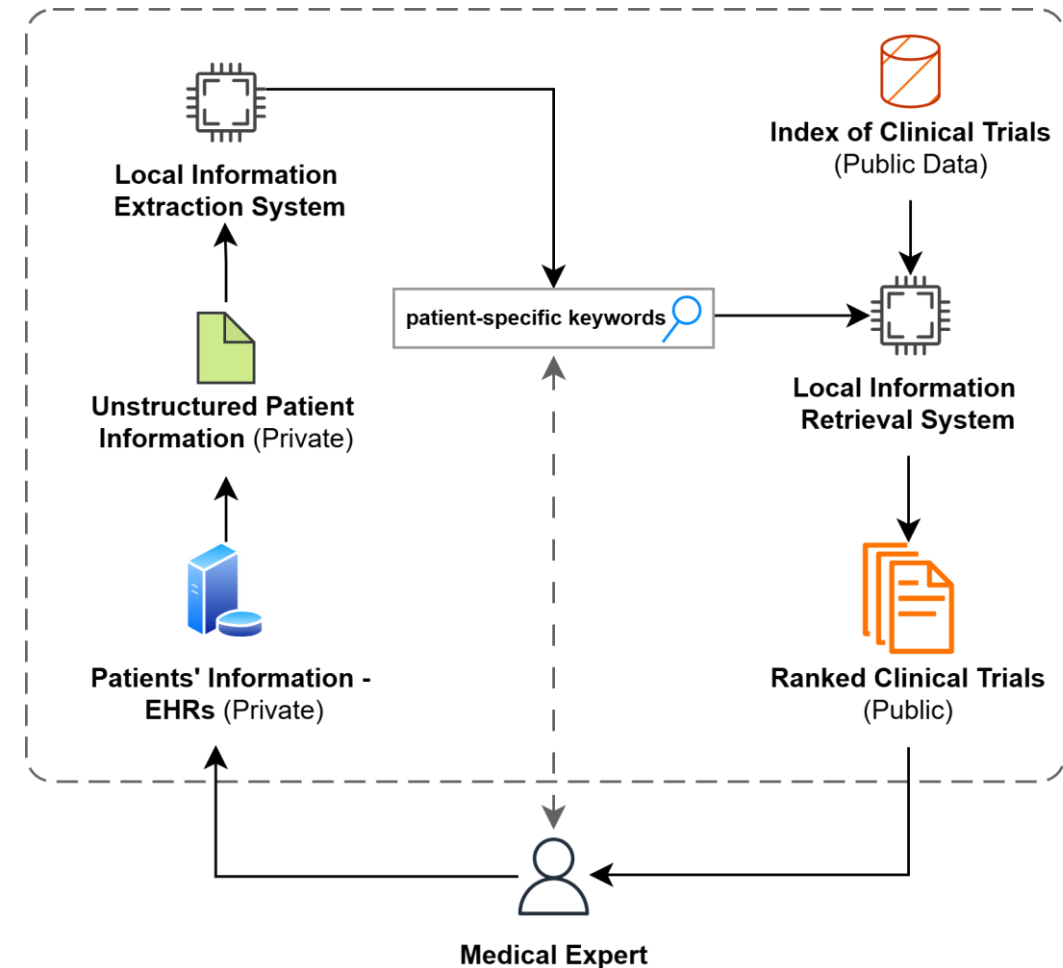
NLP System: **Performs Information Extraction** aiming to create a high-quality query

User: Reviews the generated keywords (removes, adds) [**expert's-oversight**]

IR System: Uses the query to **retrieve clinical trials** in which the patient can participate [**BM25 model**]

User: Reads and reviews the top-ranked clinical trials

An **NLP-based Information Extraction system** can leverage rule-based approaches, machine learning, deep learning architectures, LLMs, or their combinations

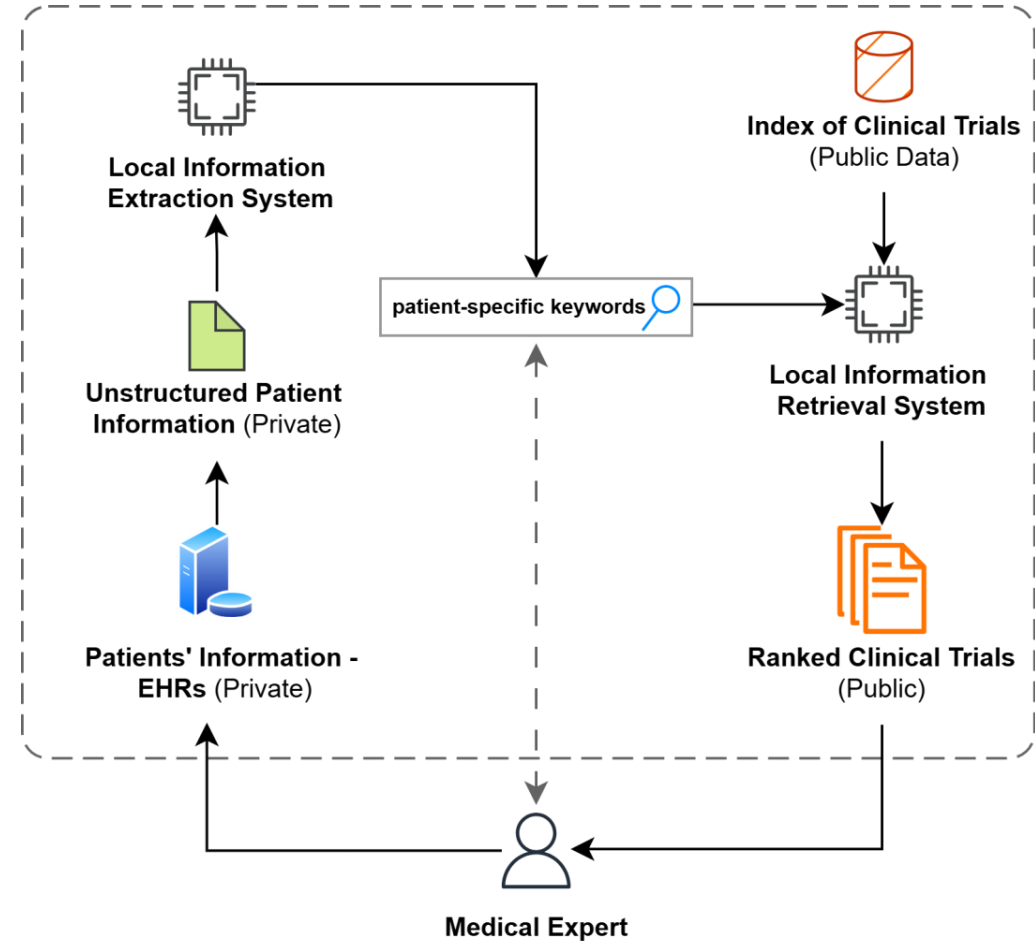


Healthcare Application – Approach

Please identify the patient's medical condition and current treatments, including any alternative names, abbreviations, or synonyms for these terms, as well as any additional criteria that may be important for identifying clinical trials of interest. Respond with a comma-separated list of keywords that will be used for search. Do not elaborate or explain.

Patient's medical note: {Clinical Note}

Patient is a 45-year-old man with a history of anaplastic astrocytoma of the spine complicated by severe lower extremity weakness and urinary retention s/p Foley catheter, high-dose steroids, hypertension, and chronic pain. The tumor is located in the T-L spine, unresectable anaplastic astrocytoma s/p radiation. Complicated by progressive lower extremity weakness and urinary retention. Patient initially presented with RLE [...]



Benchmarking Systems – Collection

Document Collection (corpus): A collection of clinical trial documents taken as a snapshot from ClinicalTrials.gov, a public database of clinical studies maintained by the U.S. National Library of Medicine

Queries: A predefined set of queries that simulate patient admission notes in the emergency department

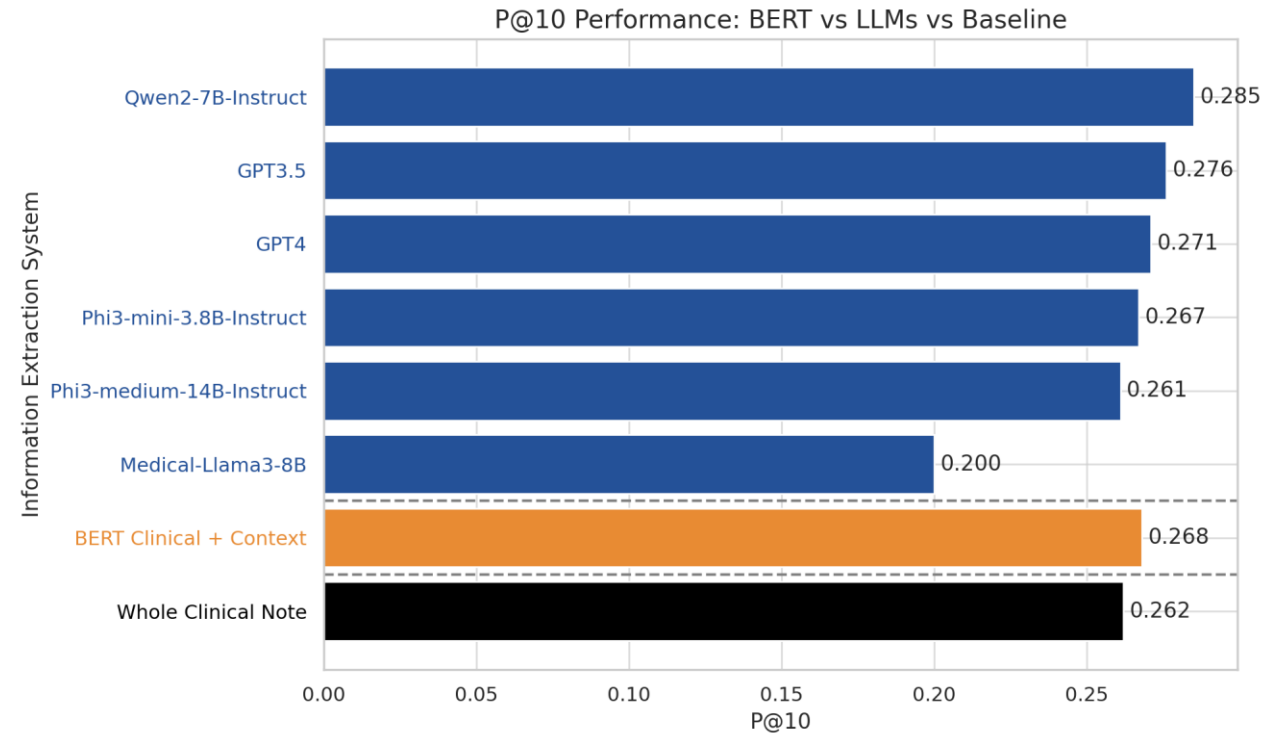
Relevance Judgments (ground truth): A set of expert-labeled query-document pairs indicating whether a clinical trial is Eligible, Excluded, or Not Relevant to the patient case described in the query

Performance Analysis – NLP Approaches

Aim of evaluation: Which NLP approach creates queries that retrieve more eligible clinical trials, resulting in improved enrollment

Systems Evaluated: A combination of a Deep Learning and a Ruled-based approach, Closed-sourced and open-sourced LLMs

[Precision@10 - number of eligible clinical trials in the top 10 retrieved]

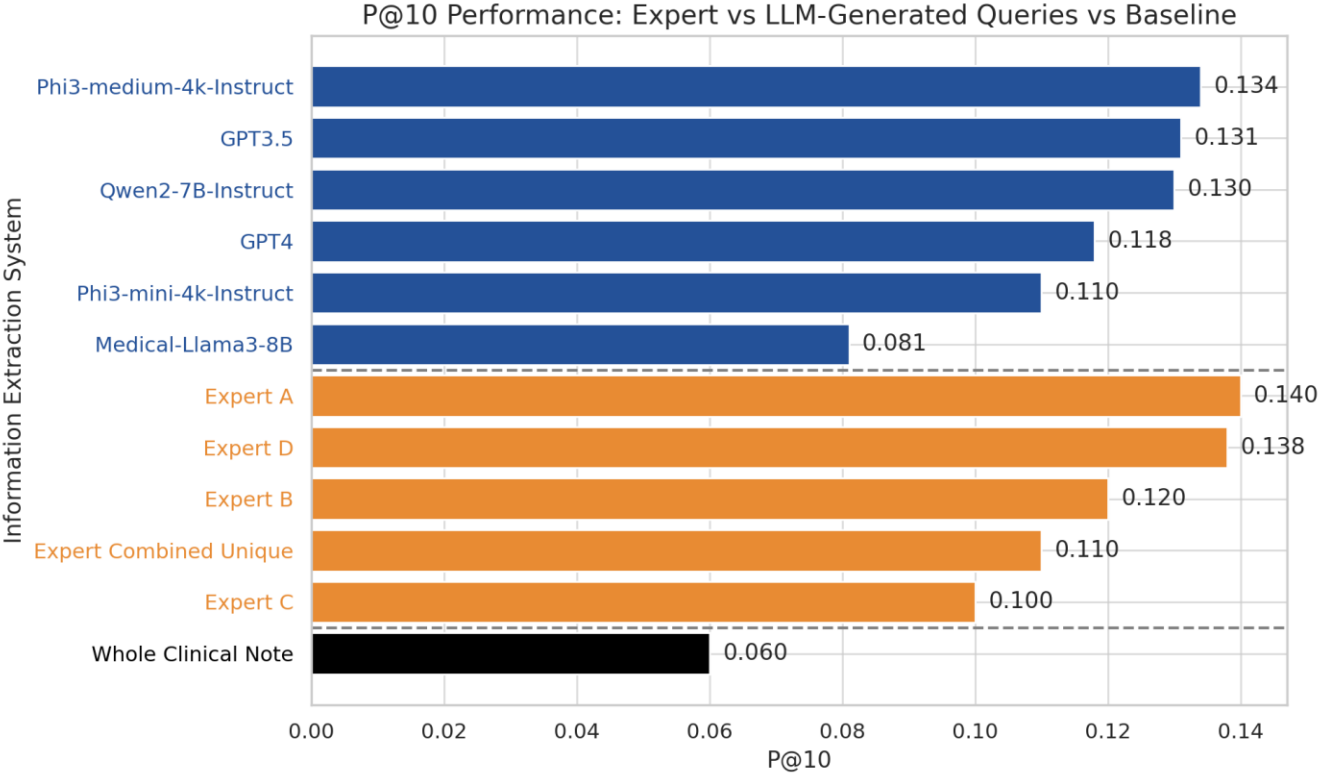


Performance Analysis – LLMs vs Experts

Aim of evaluation: Assess whether LLM-generated queries or expert-generated queries retrieve more eligible clinical trials

Systems Evaluated: Closed-sourced and open-sourced LLMs, expert-generated queries (clinicians)

[Precision@10 - number of eligible clinical trials in the top 10 retrieved]




Concluding Insights – Looking into the future

 Expert-generated queries still lead, but the strong performance of LLMs suggests that **LLM-human synergy**

could be highly effective (e.g., reduce clinical note processing time and query generation)

 Open source LLMs perform **on par with or better** than closed-source LLMs (GPTs)

 Larger or **newer models** (e.g., GPT-4) do **not guarantee better performance**

 Finetuned transformer-based models combined with rule-based approaches are **comparable** to **some LLMs**

 Instruction-tuned models **outperform** domain-specific **non-instruction** models

References & Acknowledgment

Investigating the impact of query representation on medical information retrieval

Peikos, G., Alexander, D., Pasi, G., & de Vries, A. P. (2023, March). Investigating the impact of query representation on medical information retrieval. In European Conference on Information Retrieval (pp. 512-521). Cham: Springer Nature Switzerland.

Utilizing ChatGPT to enhance clinical trial enrollment

Peikos, G., Symeonidis, S., Kasela, P., & Pasi, G. (2023). Utilizing ChatGPT to enhance clinical trial enrollment. arXiv preprint arXiv:2306.02077.

Leveraging Large Language Models for Medical Information Extraction and Query Generation

G. Peikos, P. Kasela and G. Pasi, "Leveraging Large Language Models for Medical Information Extraction and Query Generation," in 2024 IEEE/WIC International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT), Bangkok, Thailand, 2024, pp. 367-372, doi: 10.1109/WI-IAT62293.2024.00058.

This work was funded by the National Plan for NRRP Complementary Investments (PNC, established with the decree-law 6 May 2021, n. 59, converted by law n. 101 of 2021) in the call for the funding of research initiatives for technologies and innovative trajectories in the health and care sectors (Directorial Decree n. 931 of 06-06-2022) - project n. PNC0000003 - AdvANced Technologies for Human-centrEd Medicine (project acronym: ANTHEM). **This work reflects only the authors' views and opinions, neither the Ministry for University and Research nor the European Commission can be considered responsible for them.**

LLM-based Information Extraction from Clinical Notes for Downstream Healthcare Applications

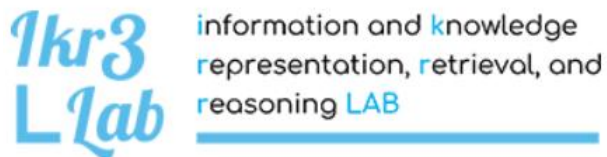
Georgios Peikos

Researcher (RTDA)

University of Milano-Bicocca, Milan, Italy

Department of Informatics, Systems, and Communication (DISCo)

Information and Knowledge Representation, Retrieval and Reasoning Lab (IKR3)



Questions? Contact

✉ georgios.peikos@unimib.it

