

# RETHINKING REAL- WORLD DATA WITH LARGE LANGUAGE MODELS

LUMC-NLP-LAB

JULIUS HEEMELAAR

AIOS CARDIOLOGIE

POSTDOC



What is the  
real gain?






# Reducing the administrative burden of clinical and research activities

## Risico toepassingsgebieden van GAI

| Toepassing                               | Risico Niveau             | Potentie Niveau |
|--|---------------------------|-----------------|
| Automatisering van administratieve taken | Laag tot gemiddeld risico | Hoog            |

## Perspective on the use of AI



| Role in advice | Data quality & data access & data combination |                            |                                       |                           |                                    | Integration & interpretation of data: personal diagnosis, prediction and advice |                       |                                      |                                | Engaging & user aware advice |  |
|----------------|---|----------------------------|---------------------------------------|---------------------------|------------------------------------|---|-----------------------|--------------------------------------|--------------------------------|------------------------------|--|
| Challenge      | Unstructured data                             | Privacy & legislation      | Burden of data combination            | Burden of data collection | Interoperability & standardization | Black box models  | Knowledge access      | (Health) inequity                    | Continuous data interpretation | Engagement                   | Lack of agency   |
| Solution       | Large language models (generative A.I.)       | Federated learning & S-MPC | Automated data standardization (LLMs) | Speech to data (LLMs)     | Automated FAIRification            | Explainable and trustworthy AI  | Large language models | Bias aware AI/ assessability of A.I. | n=1 statistics                 | AI for Engagement            | Personalized advice models<br><br>LLM's -> comprehensive advices/ language |

# The Scalable Future of Healthcare Research with LLMs

**Scalability Challenge:** EHR-based research is overwhelmed by ever-growing volumes of free text data.

**Unlocking Insights:** Crucial information is buried in these texts, creating barriers to efficiency.

**Solution:** Large language models (LLMs) automate text analysis, unlocking unprecedented scalability.

## **Impactful Use Cases:**

- Faster and more accurate clinical trial eligibility screening

- Retrospective real-world disease incidence and outcome insights

**Transformative Potential:** LLMs are poised to revolutionize healthcare research



“student behind a computer  
underneath a lamp surrounded  
by stacks of paper, digital art”



**Julius × DALL·E**

Human & AI



# Quick math

Recent preparations for retrospective study on breast cancer

Pilot 300 patients

Question: how many follow up documents?

72.000 notes (~ 240 notes per person)

5000 patients → 1.2 million notes

18-20.000 patients (full cohort)

**4-5 million notes!**

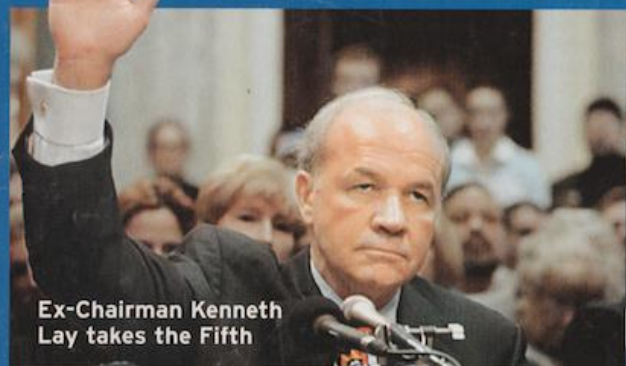


DID YOU LOSE MONEY? CHECK OUR LIST OF 500 MUTUAL FUNDS TO FIND OUT

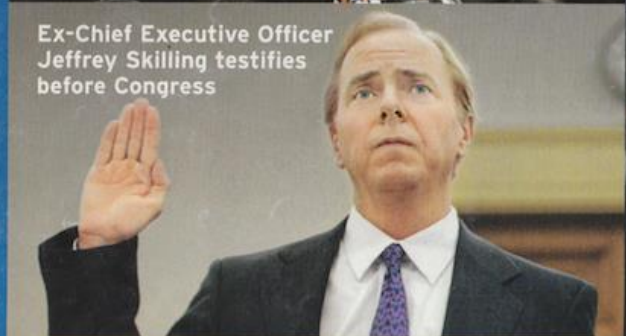
# ENRON EXPOSED!

How millions of innocent Americans  
lost their life savings

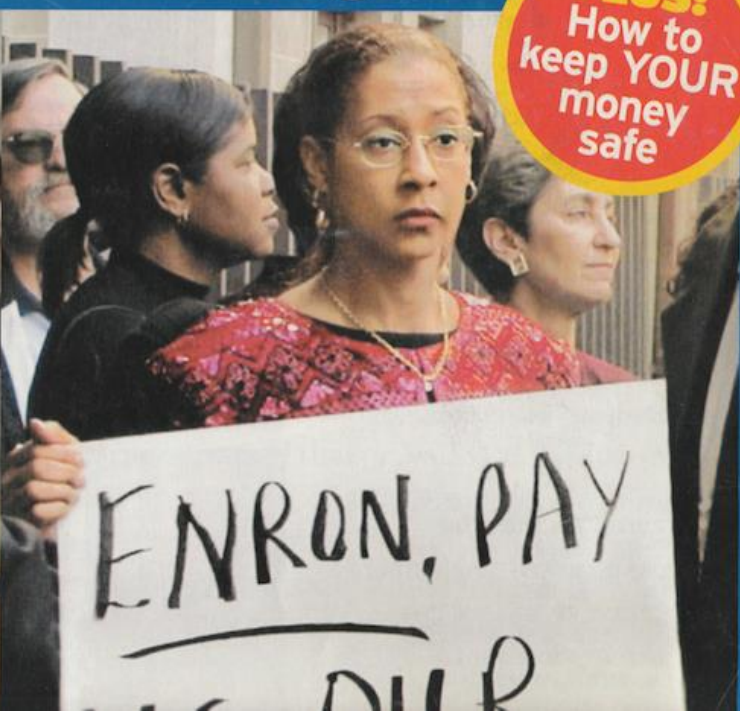
**PLUS:**  
How to  
keep YOUR  
money  
safe



Ex-Chairman Kenneth Lay takes the Fifth



Ex-Chief Executive Officer Jeffrey Skilling testifies before Congress



*“ When dozens of FBI agents descended on Enron's headquarters, they carted away hard drives and hundreds of boxes of documents. The task force had an onerous task: making sense and a criminal case out of approximately ten million documents. ”*



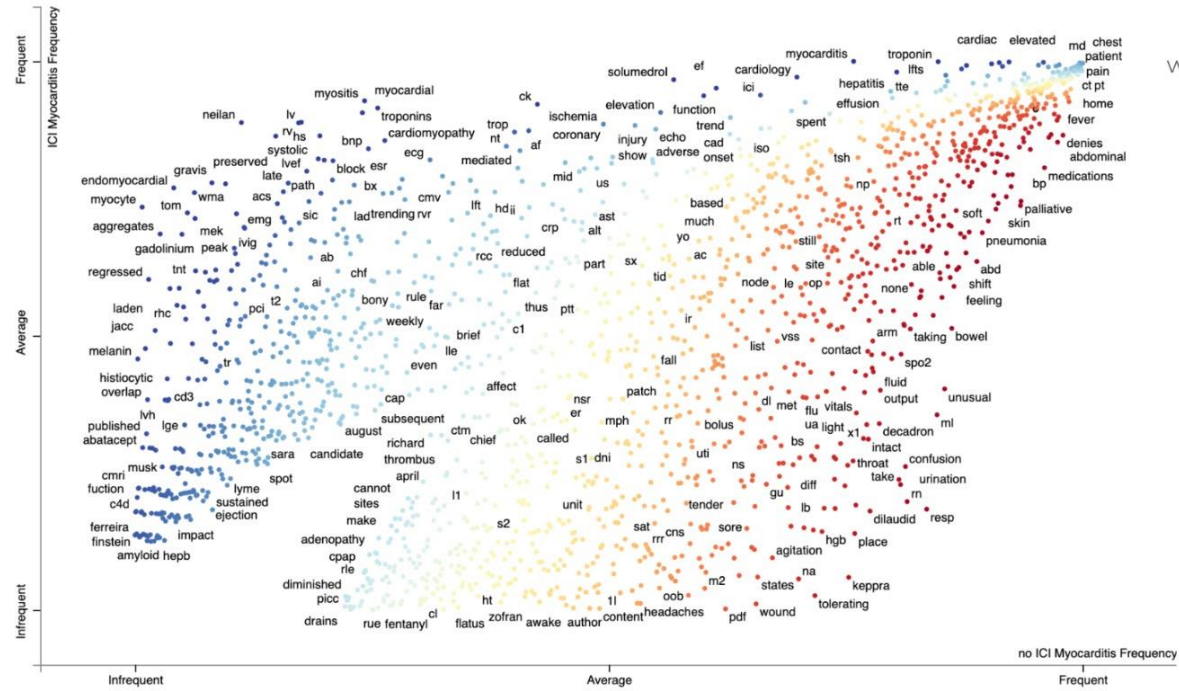
"The Tour de France is lovely"

"Mathieu van de Poel wint de Ronde van Vlaanderen"

"I cycle to work"

Embedding model

|      |      |      |      |
|------|------|------|------|
| 0.83 | 0.52 | 0.01 | ...  |
| 0.95 | 0.36 | 0.37 | ...  |
| 0.28 | 0.54 | 0.83 | .... |



Words associated with ICI Myocarditis:

- myocarditis
- myositis
- solumedrol
- neilan
- troponins
- rv
- troponin
- lv
- myocardial
- cellcept
- cardiology
- ck
- ef
- endomyocardial

Embedding



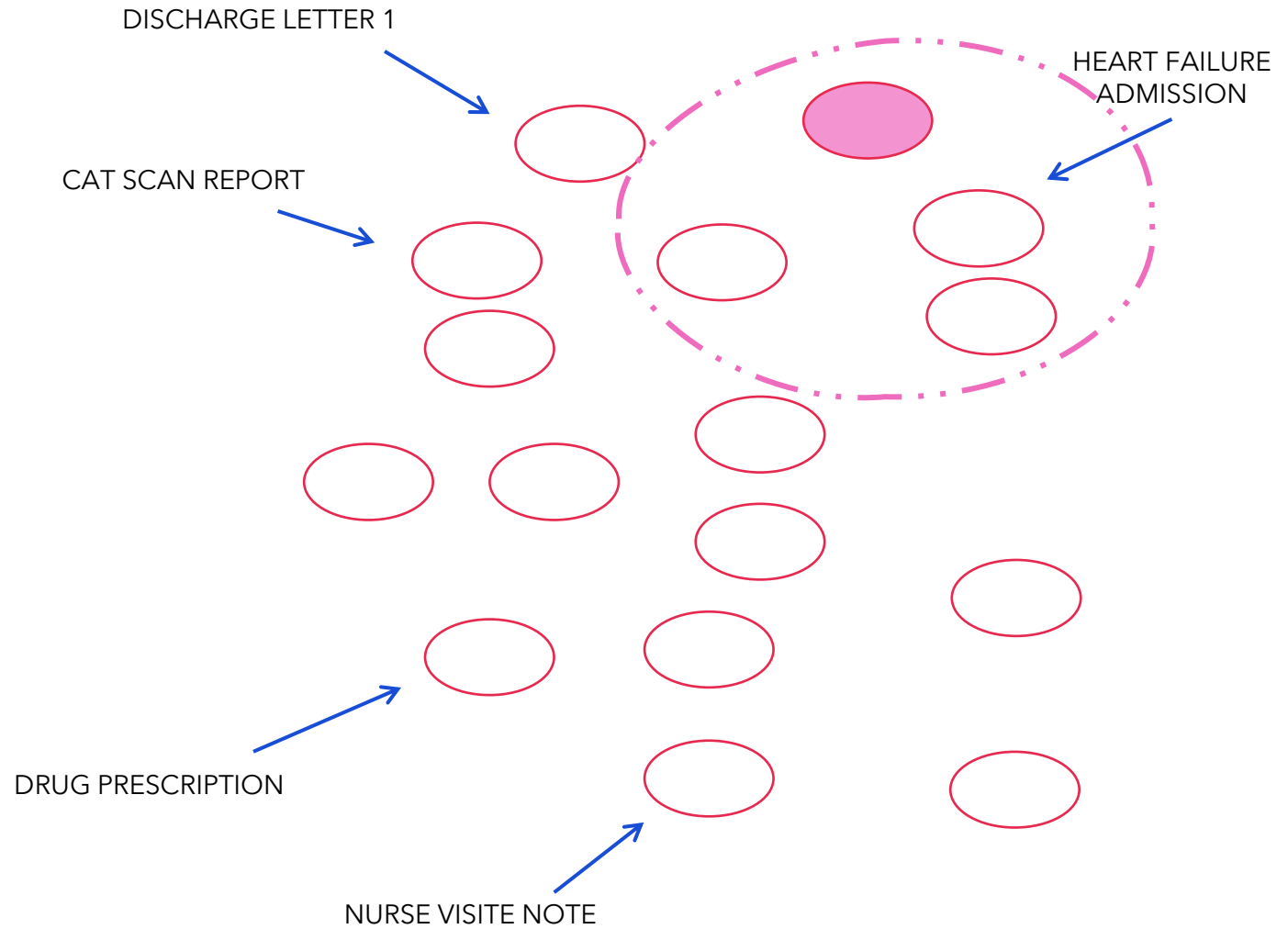
What if..

**Context**

Complete record of patient X

**Query**

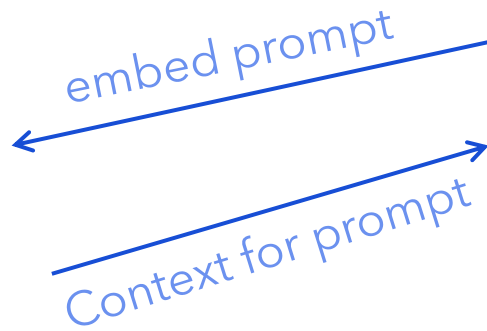
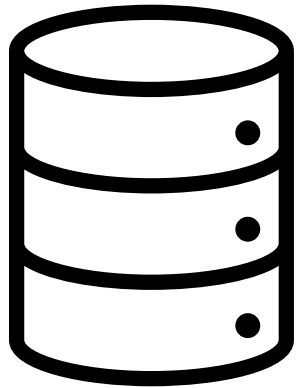
*"Does this patient have heart failure?"*



Complete record of patient X

retrieval

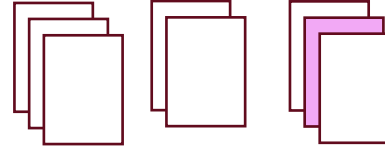
Vector store with complete record of patient X



prompt

Query = "does this patient have heart failure?"










+context =



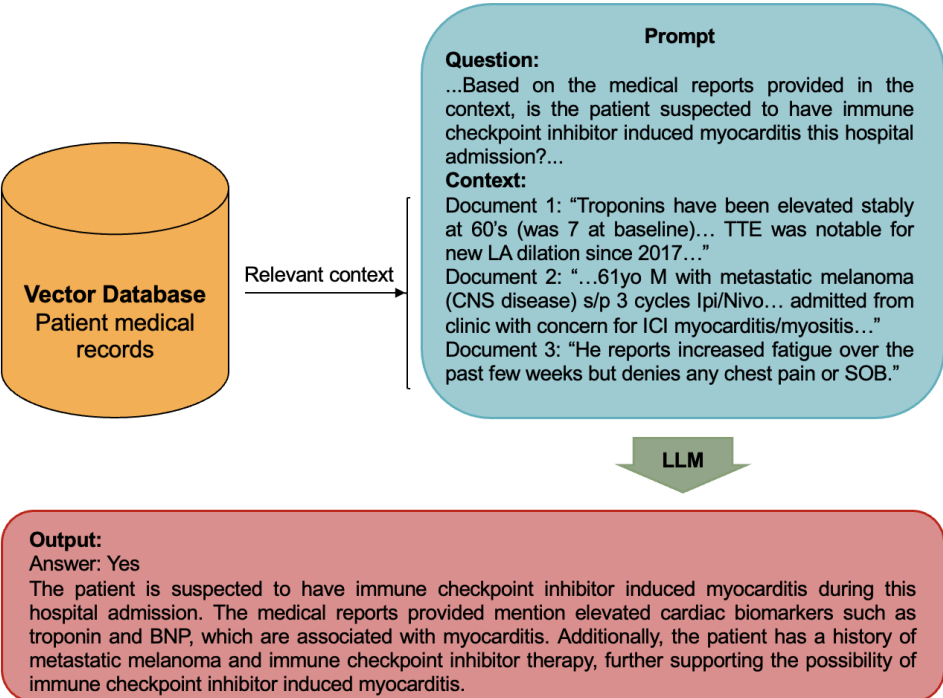
LLM

Retrieval augmentation generation (RAG) - 'chatting' with the records

# Enhancing Precision in Detecting Severe Immune-Related Adverse Events: Comparative Analysis of Large Language Models and International Classification of Disease Codes in Patient Records

Virginia H. Sun, MD<sup>1,2</sup> ; Julius C. Heemelaar, MD<sup>1,2,3</sup> ; Ibrahim Hadzic, MSc<sup>1,4,5,6</sup> ; Vineet K. Raghu, PhD<sup>1,2</sup> ; Chia-Yun Wu, MD<sup>7,8</sup>; Leyre Zubiri, MD, PhD<sup>1,7</sup> ; Azin Ghamari, MD<sup>1,2</sup>; Nicole R. LeBoeuf, MD, MPH<sup>1,9,10</sup> ; Osama Abu-Shawer, MD, MS<sup>11</sup> ; Kenneth L. Kehl, MD, MPH<sup>1,12,13</sup> ; Shilpa Grover, MD, MPH<sup>1,14</sup> ; Prabhsimranjot Singh, MD<sup>1,13</sup>; Giselle A. Suero-Abreu, MD, PhD, MSc<sup>1,2,15</sup> ; Jessica Wu, BA<sup>1,2</sup> ; Ayo S. Falade, MD, MBA, APGD<sup>16</sup> ; Kelley Grealish, MSN, NP<sup>7</sup>; Molly F. Thomas, MD, PhD<sup>17,18,19</sup>; Nora Hathaway, MSN, NP<sup>7</sup> ; Benjamin D. Medoff, MD<sup>1,20</sup> ; Hannah K. Gilman, BS<sup>1,2</sup>; Alexandra-Chloe Villani, PhD<sup>1,21,22</sup> ; Jor Sam Ho, MPH<sup>1,2</sup>; Meghan J. Mooradian, MD<sup>1,7</sup> ; Meghan E. Sise, MD<sup>1,23</sup> ; Daniel A. Zlotoff, MD, PhD<sup>1,15</sup> ; Steven M. Blum, MD<sup>1,7,21,22</sup> ; Michael Dougan, MD, PhD<sup>1,24</sup> ; Ryan J. Sullivan, MD, PhD<sup>1,15</sup> ; Tomas G. Ntley, MD, MPH<sup>1,15</sup> ; and Kerry L. Reynolds, MD<sup>7</sup> 


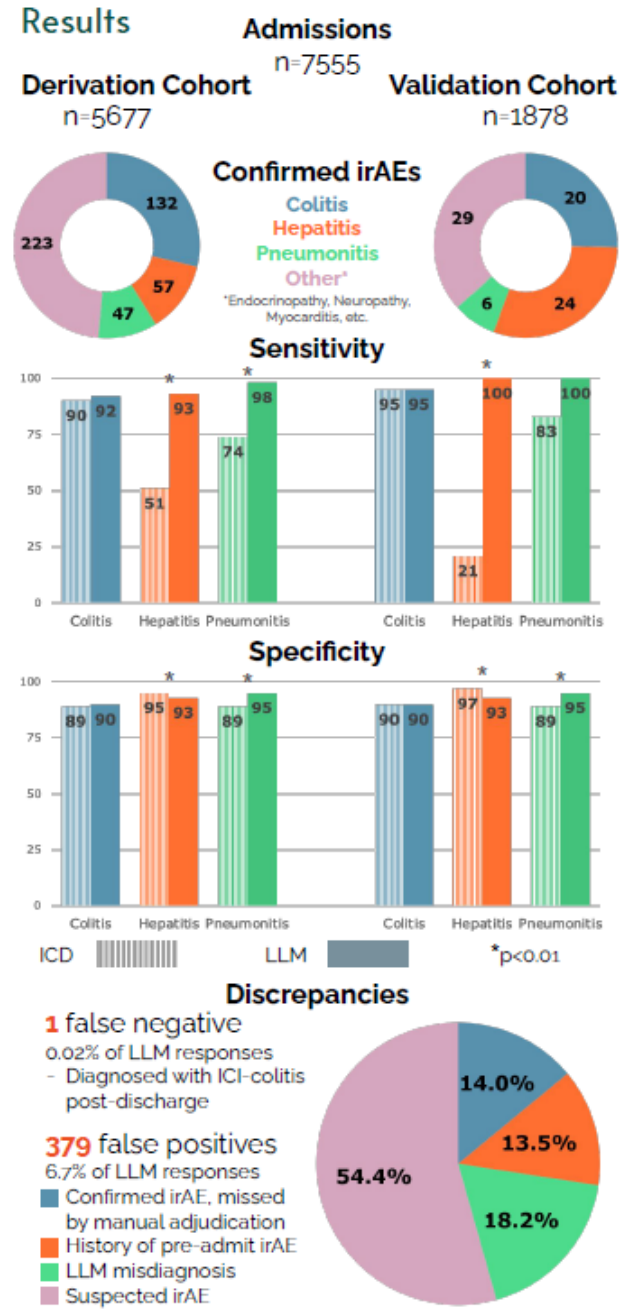
## APPLIED TO DETECTING COMPLEX ADVERSE EVENTS OF IMMUNOTHERAPY



Open-source **large language models** are an accessible and generalizable tool that can retrospectively detect **immune-related adverse events** among patients on immune checkpoint inhibitor therapy, outperforming ICD codes in accuracy and manual adjudication in efficiency.

|  |  |   |
|--|--|---|
| <b>94.2</b><br>Sensitivity<br>compared to 71.8% with ICD-codes | <b>92.5</b><br>Specificity<br>compared to 91.1% with ICD-codes | <b>9.42</b><br>sec/chart<br>versus ~15 min with manual chart review |
|--|--|---|

Copies of this slide deck obtained through Quick Response (QR) Code are for personal use only and may not be reproduced without permission from ASCO® and the author of these slides



In preprint

## Matching Patients to Clinical Trials with Large Language Models

Qiao Jin, M.D.<sup>1</sup>, Zifeng Wang, M.S.<sup>2</sup>, Charalampos S. Floudas, M.D., D.MSc., M.S.<sup>3</sup>,  
Fangyuan Chen, B.S.<sup>4</sup>, Changlin Gong, M.D.<sup>5</sup>, Dara Bracken-Clarke, M.D.<sup>3</sup>, Elisabetta  
Xue, M.D.<sup>3</sup>, Yifan Yang, B.S.<sup>1,6</sup>, Jimeng Sun, Ph.D.<sup>2</sup>, Zhiyong Lu, Ph.D.<sup>1</sup>



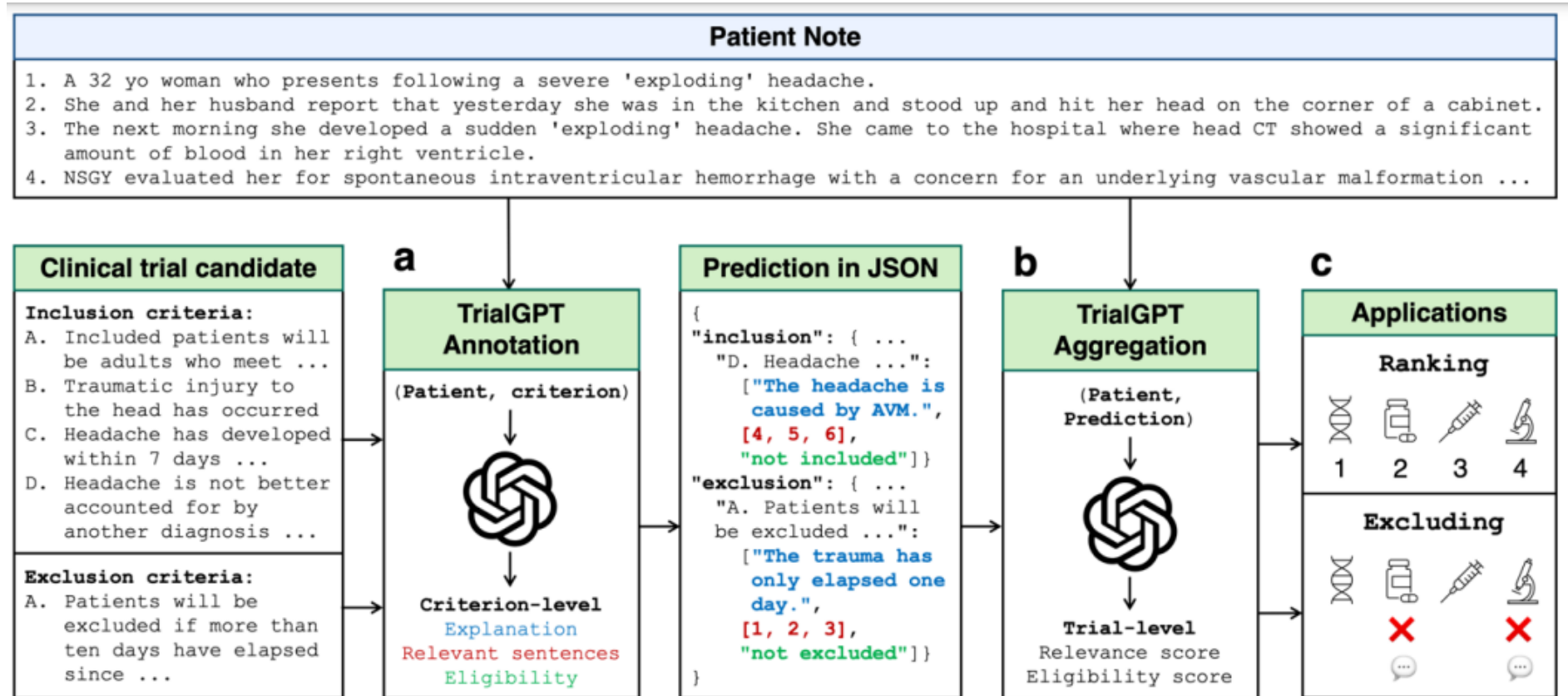
National Institutes  
of Health



**“criterion-level  
accuracy of 87.3%,  
close to the expert  
performance  
(88.7%-90.0%)”**






**“reduced screening  
time by 42.6%”**

LLMs can aid in  
clinical trial  
screening





# A Call for Artificial Intelligence Implementation Science Centers to Evaluate Clinical Effectiveness

Christopher A. Longhurst , M.D., M.S.,<sup>1</sup> Karandeep Singh , M.D., M.M.Sc.,<sup>1</sup> Aneesh Chopra , M.P.P.,<sup>2</sup> Ashish Atreja , M.D., M.P.H.,<sup>3</sup> and John S. Brownstein , Ph.D.<sup>4,5</sup>

Received: March 1, 2024; Revised: May 7, 2024; Accepted: May 16, 2024; Published: July 10, 2024

## Key Challenge in AI Implementation in Healthcare:

### Surplus of AI Models, Minimal Practical Impact:

While there is a large number of AI models available, their implementation in healthcare has been marginal, with little to no effect on patient-centered outcomes

### Relevant to the Netherlands in the LLM space

It remains unclear how well large language models perform within Dutch healthcare settings, particularly when dealing with medical jargon.

# Get involved with LUMC-NLP-LAB!

LUMC-NLP-LAB is open for collaborations!

**Implement** NLP methods, including large language models, in Dutch healthcare.

**Develop and validate** LLM applications for medical research and clinical innovation.

**Collaboration** within departments and with external partners.

Emphasize **implementation science** and **education** in all initiatives.



**Julius Heemelaar**  
Cardioloog i.o. en postdoc  
onderzoeker @ LUMC

j.c.heemelaar@lumc.nl  
X: @CardioOnco\_JHE



**Marieke van Buchem**  
Innovatiemanager @  
CAIRElab

m.m.van\_buchem@lumc.nl