Instruction Based Zero-shot

#Task
#Instruction: I1, I2...
#Clinical note: NOTE TEXT

Prompt

{Phenotype present: Yes/No
Clinical Reason: abc def }

JSON Output

Prompt for each ocular condition
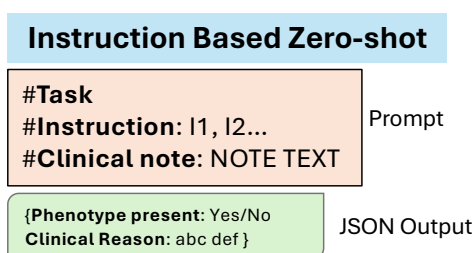Glaucoma
Diabetic Retinopathy
Age-related Macular Degeneration

Randomly selected patients (age:40-80 years) who received eye care between 2012-21 with minimum 2 eye care visits.
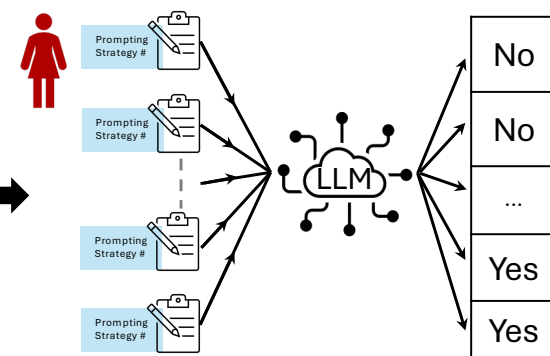
Scan EHR to label each patient for presence or absence of ocular condition independently

Stanford University
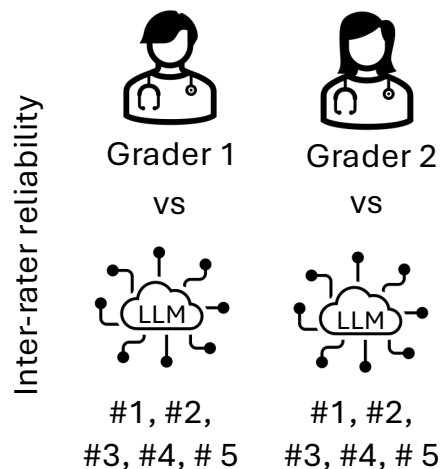University of Michigan

Grader 1   Grader 2

Prompting Strategy #

LLM

No
No
...
Yes
Yes

YES

Evaluation for each patient

Individual Model Performance

Inter-rater reliability

Grader 1
vs
LLM
#1, #2, #3, #4, #5

Grader 2
vs
LLM
#1, #2, #3, #4, #5

Consensus-Based Approach Performance

Number of LLMs (N)

LLM
#1
#2
#3
#4
#5

MAJORITY VOTING

Final Annotations

N ≥ 3   N ≥ 4   N ≥ 5

No	No	No
No	No	No
...	...	...
Yes	Yes	No

Cases where Majority voting did not agree
Grader 1

| No | No | No |
| | Yes | Yes |

Inter-rater reliability

Final Annotations

N ≥ 3   N ≥ 4   N ≥ 5
vs   vs   vs
Grader 2 Grader 2 Grader 2