Can AI build reliable clinical codelists?

Raqiya Ali

Background:

Codelists ➡ used by researchers when referring to diseases/symptom (ref)(Ox.ac.uk, 2023).

Required in EHR research for consistency and reproducibility of studies

Codelist creation ➡ Time consuming + resource intensive

AIM: To Evaluate whether AI can create accurate codelists

RAG:
Retrieval Augmented Generation

Knowledge base retrieval system + LLM

Methods:

Knowledge base (ICD-10 TABULAR)

↓

Create embedded chunks

↓

User prompt

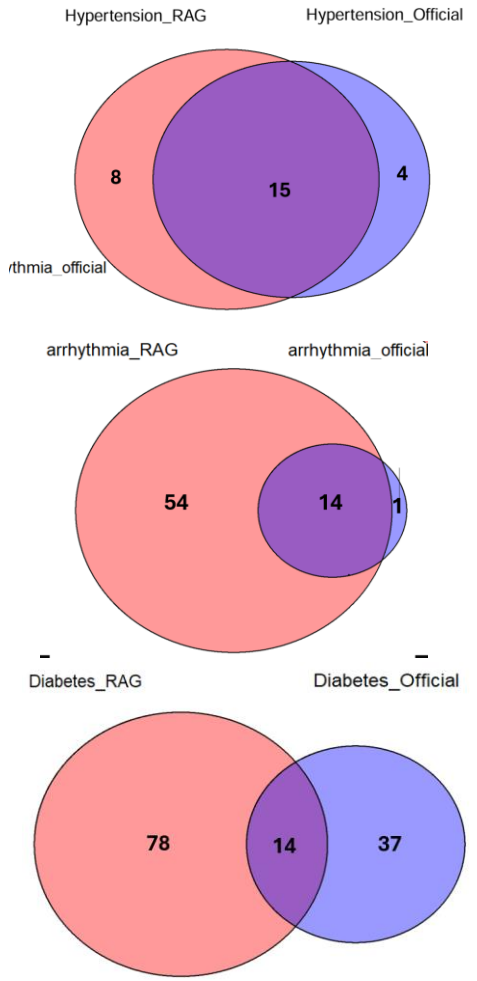
↓

Retrieve relevant chunks

↓

LLM response

Results:



Conclusion :

- AI can capture subset of official codes and extra codes
- Useful for speeding up codelist development
- Clinical expert validation still necessary
- Alignment is disease dependent

Main Takeaway:
more refinement and validation is needed before deployment

Figure legend:
Venn diagram shows overlap between RAG derived codelists (RED) and official codelists (blue)
Shared regions = codes identified by both methods
Non overlapping = unique codes to either RAG or official codelists