

Diagnosis Certainty and Progression: A Natural Language Processing Approach to Enable Characterisation of the Evolution of Diagnoses in Clinical Notes

Poster number: 14
Contact details:
alfredo.madrid@salud.madrid.org

Alfredo Madrid^{1,3,5}, Caitlin Bullin¹, Lifeng Han³, Judith Andrew^{1,3}, Warren Del-Pinto^{1,3}, Ghada Alfattni^{1,3}, Oswaldo S. Pabón⁴, Ernestina M. Ruiz⁴, Luis Rodríguez⁵, Ana María Aldana², Robyn Hamilton², Karim Webb², Meghna Jani², Goran Nenadic³, William G. Dixon^{1,2}
¹Centre for Epidemiology Versus Arthritis, University of Manchester, Manchester, UK; ²Department of Rheumatology, Salford Royal Hospital, Salford, UK; ³School of Computer Science, The University of Manchester, Manchester, UK; ⁴Universidad Politécnica de Madrid, Madrid, Spain; ⁵Rheumatology Department, and Health Research Institute (IdISSC), Hospital Clínico San Carlos, Madrid, Spain

HeaTAC 2022: the 5th Healthcare Text Analytics Conference

Introduction

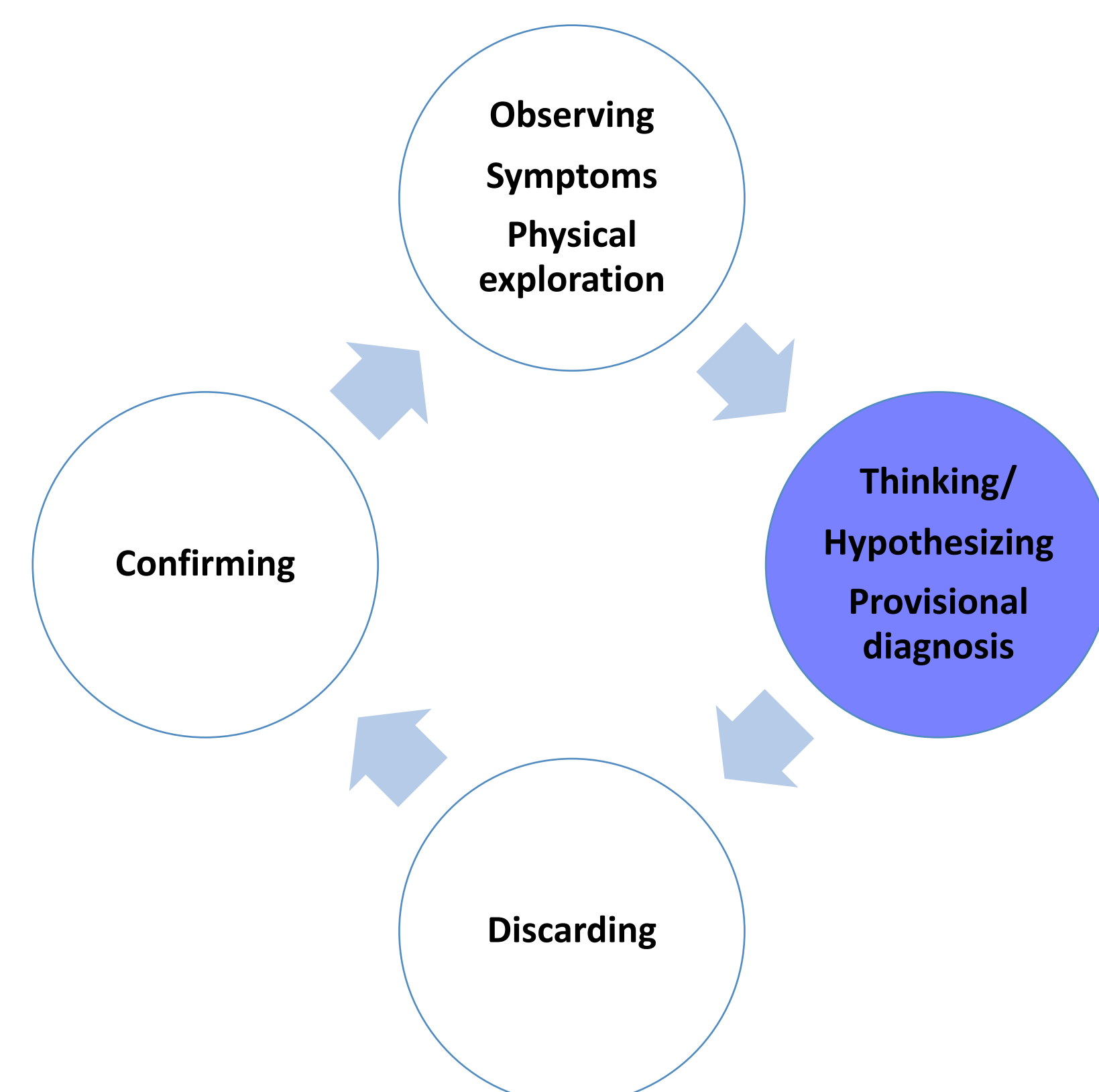
The accurate identification of **diagnoses** in **free clinical narratives** is decisive for characterizing the patients in a medical cohort. Therefore, the **knowledge extraction and information retrieval** tasks must be addressed carefully.

Clinical notes might present **multiple qualifiers** that could change the meaning of a statement: negation, speculation, temporal information, family history and so on.

Sentence	Qualifier	Status
Rheumatoid Arthritis can be ruled out	Negation	Discarded
Rheumatoid Arthritis under investigation	Speculation	Not confirmed
Family history of Rheumatoid Arthritis	Family member	Not confirmed
Under biologic treatment for Rheumatoid Arthritis	Confirmed	Confirmed

Table 1. Clinical notes qualifiers

It is not unusual for caregivers to **preserve uncertainty using broad and ambiguous terms** when they have not full evidence of the disease status of a patient.



The **percentage** of speculative sentences may range between the **11% and the 22.7%**¹

It is estimated that **over 40%** of the data in an **EHR** is stored as free text.

Sample letters	
Letter 1	
DD	
Hospital No. xxxxxxx	
Date/Time of Appt:	02 December 2019 at 09:00
Clinic:	RHEUMATOLOGY
Type of Appt:	New
Rheumatological Diagnosis:	Likely rheumatoid arthritis (mildly seropositive, on biologic since 2012) Previous episodes of episcleritis (x 4)
Non-Rheumatological Diagnosis:	Trigeminal neuralgia Mild Meniere's disease ETOH excess
Medications:	Omeprazole 20mg od Tramadol 50mg prn Paracetamol prn Diclofenac 50mg prn Hydroxychloroquine 200mg od -- to be increased to bd Adalimumab 40mg every 2 weeks Fluoxetine 20mg od Adcal D3 x 1 tablet bd Gaviscon prn
I had the pleasure of meeting Mr. Linacre at the Community Nutrition and Dietetics outpatient clinic on 1 May 2017, referred by Sugra Bibi, Hospital dietician at St Crispin's Hospital, Donaldstown, D05 7TP. He attended the community nutrition and dietetics outpatient clinic for review of feeding. Following a stroke Mr. Linacre acquired swallowing difficulties. During a recent admission to hospital Mr. Linacre was established on PEG tube feeding. The feeding tube insitu is a 15French PEG tube placed 05/04/17. The regimen being: 1000mls Energy Multifibre Feed at 100mls/hours for 10 hours (9am-7pm) with 1400mls water given as divided flushes (e.g. 10x140mls) throughout the day e.g. before and after feed and with medications.	
Yours faithfully,	

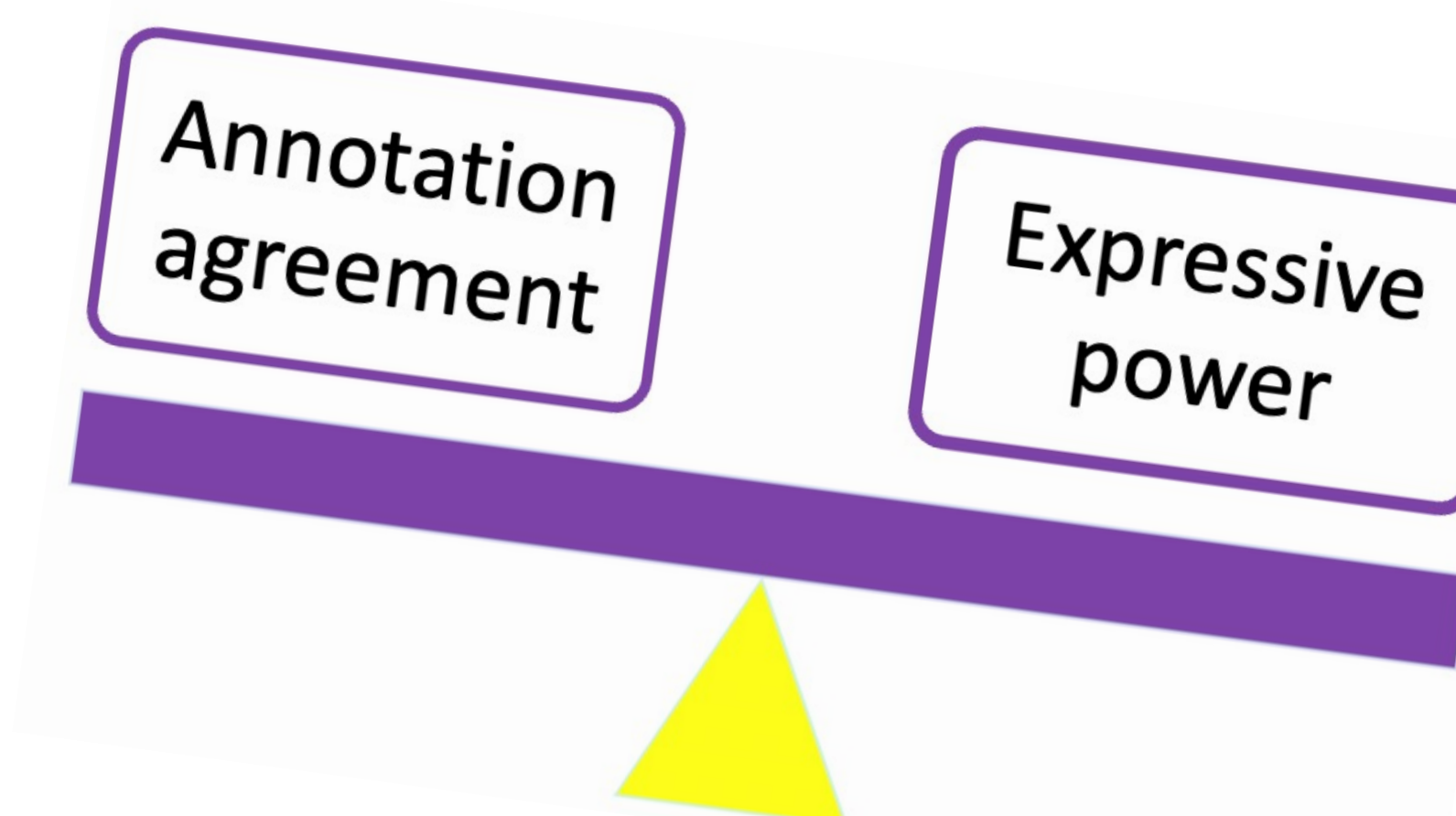
Image 1. Clinical notes with speculation qualifier

Objectives

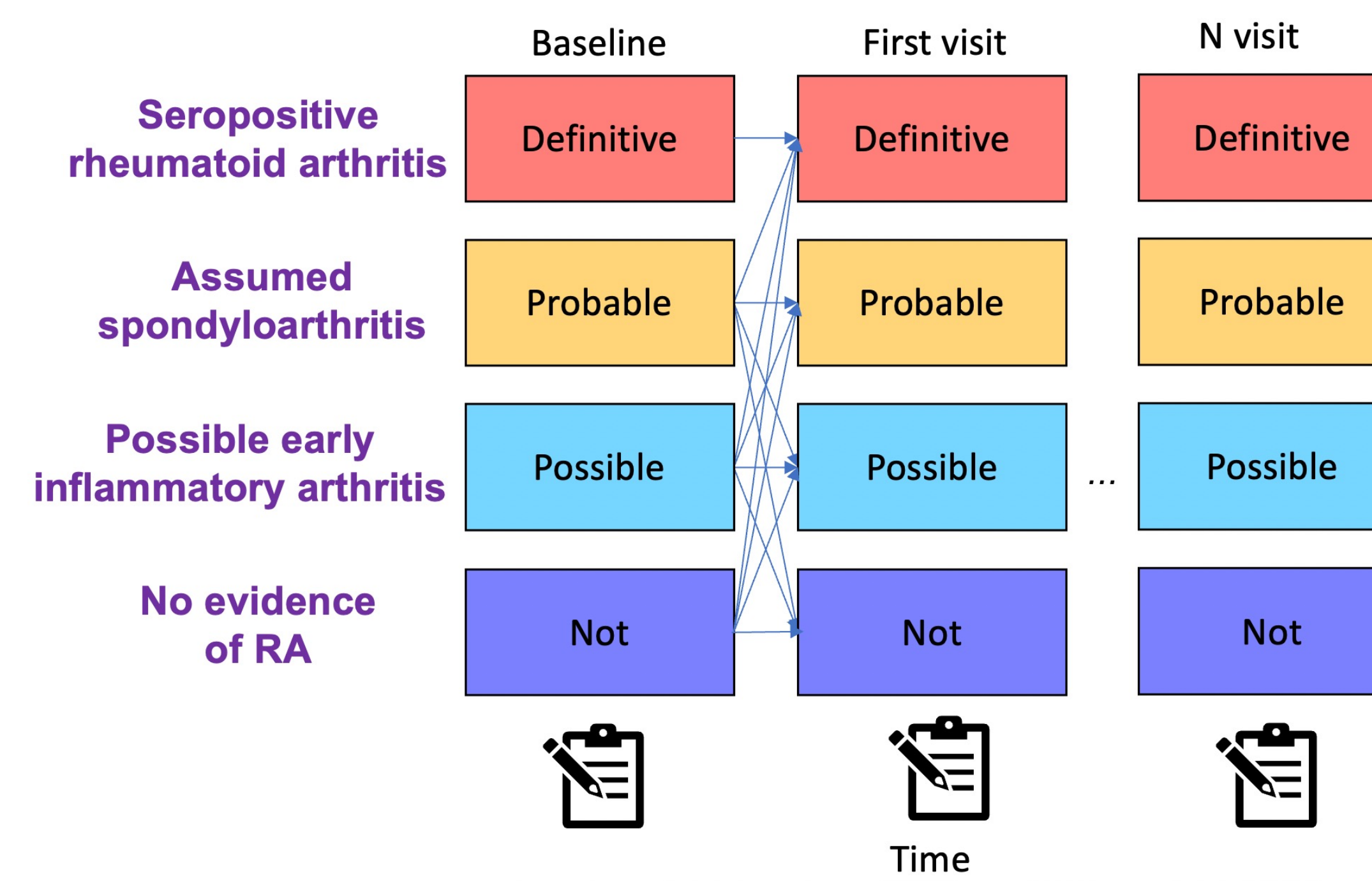
- To appropriately identify the **uncertainty and negation qualifiers** in Rheumatic and Musculoskeletal narratives building a **classification model** that handles four different categories: **Definitive, probable, possible, not**
- To demonstrate the extent to which that influences possible population health research about **incidence/prevalence** of disease and its evolution

Levels of speculation

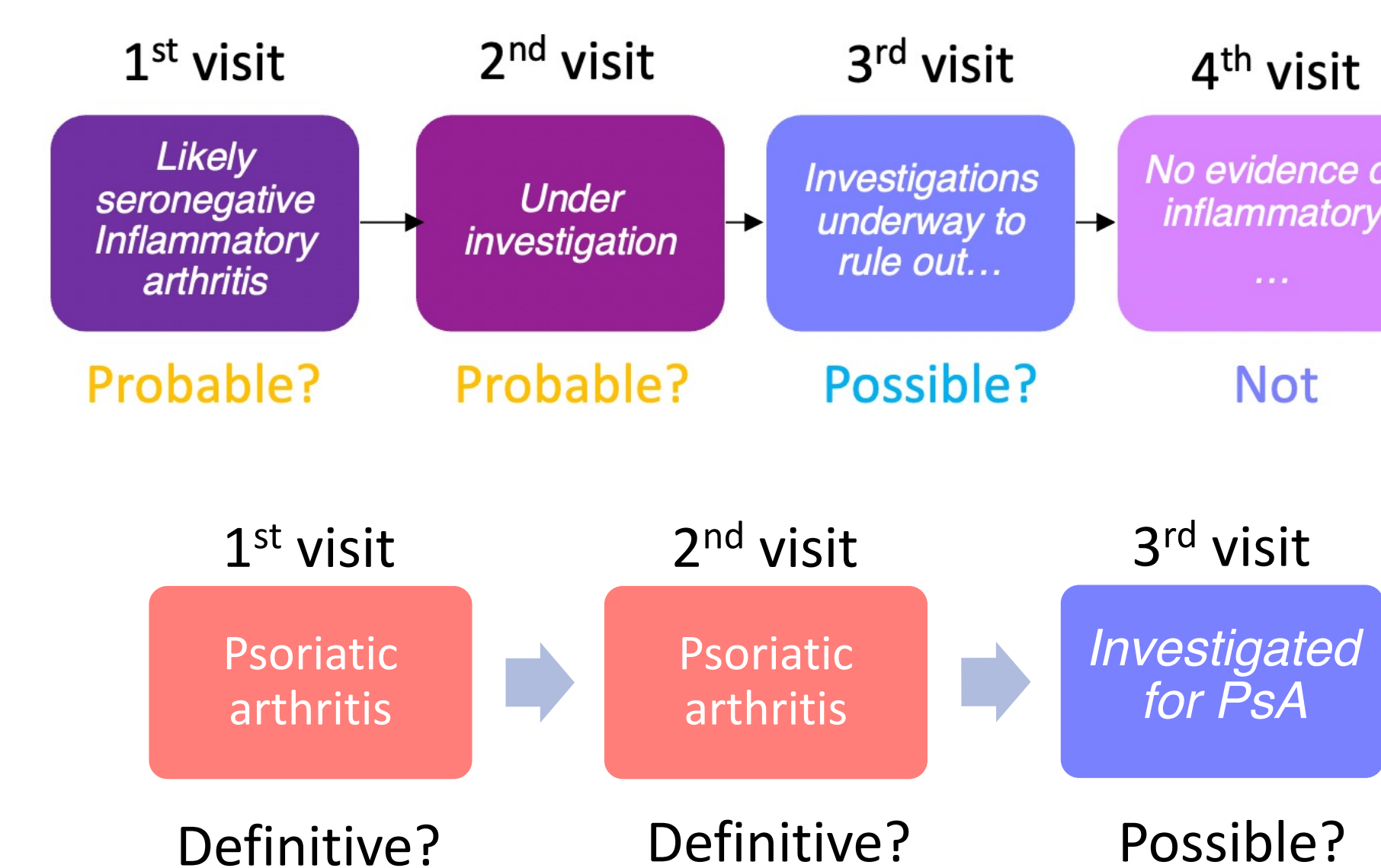
More annotation classes lead to **lower agreement results**. Less granular models **lose expressive power**²



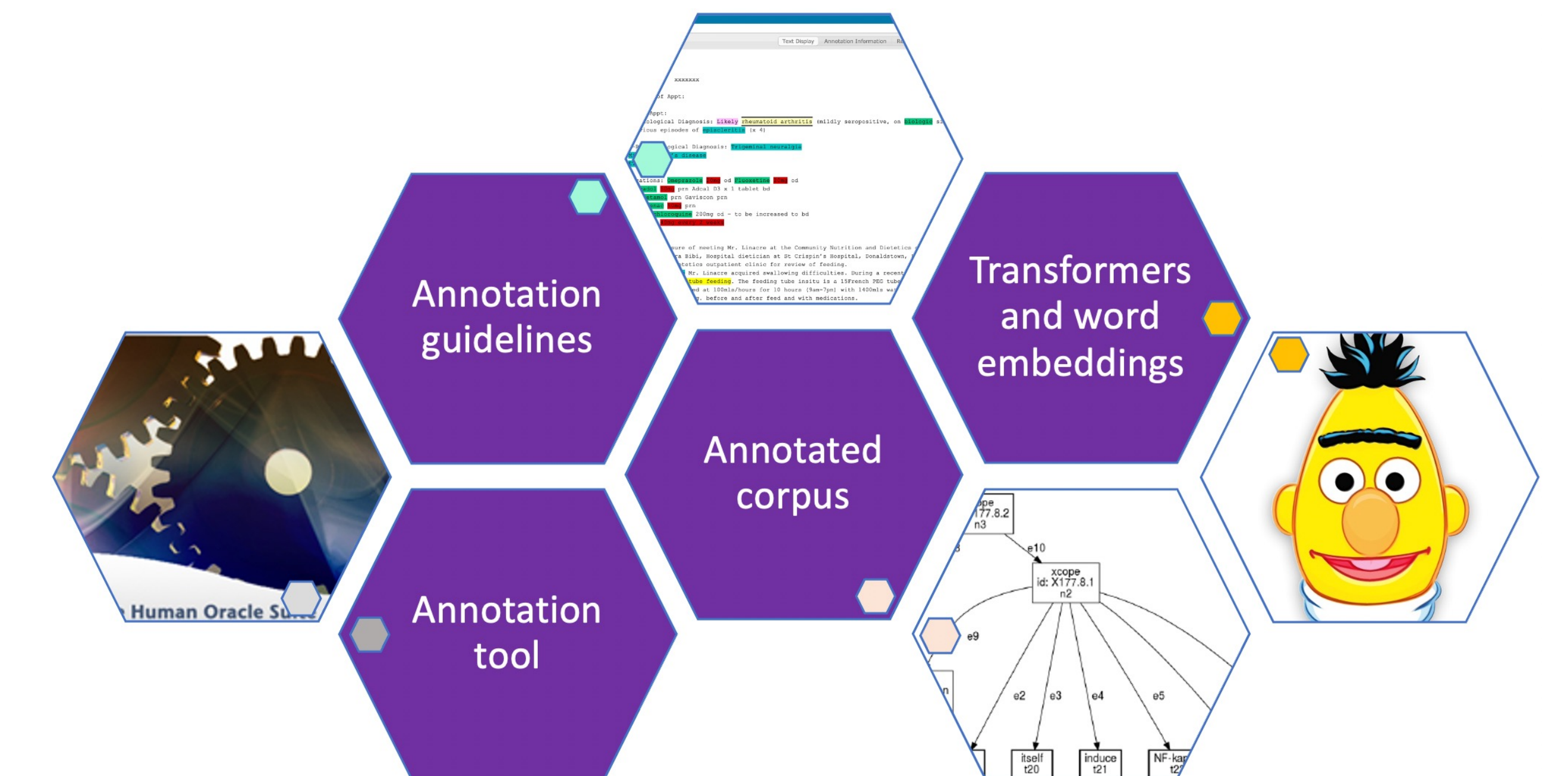
A **binary classifier** is good for identification but don't allow us to characterize the speculation evolution over clinical notes.



Examples



NLP and text mining



Roadmap

Annotation guidelines, and definition of the four different categories: Definitive, Probable, Possible, Not.

Manual annotation of clinical notes by a physician: *eHost, spreadsheet*

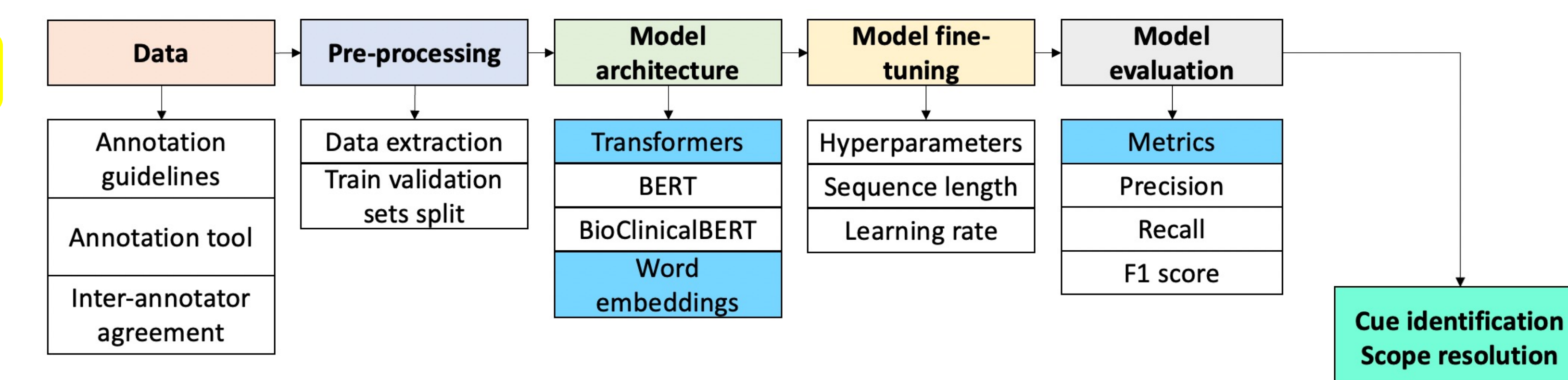
Inter-annotator agreement measure with a subset of clinical notes

Enriching dataset with *Probable* and *Possible* cases

Word embeddings + classification model: Snomed2Vec

Transformers models: BioClinicalBert, MedBert

Cue identification and scope resolution: Precision, recall, F1-Score



References

- ¹Cruz Díaz, N. P., & Maña López, M. J. (2019). Negation and Speculation Detection (Vol. 13). John Benjamins Publishing Company. <https://doi.org/10.1075/nlp.13>
²Velupillai, S. (2012). *Shades of certainty: annotation and classification of swedish medical records* (Doctoral dissertation, Department of Computer and Systems Sciences, Stockholm University).