

The challenges of industrializing Natural Language Processing for health data

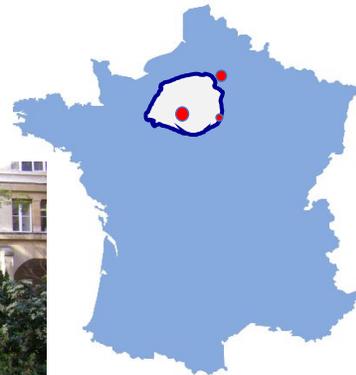
Xavier Tannier
xavier.tannier@sorbonne-universite.fr

Montréal, April 8, 2025





Research in Computer Science for Health



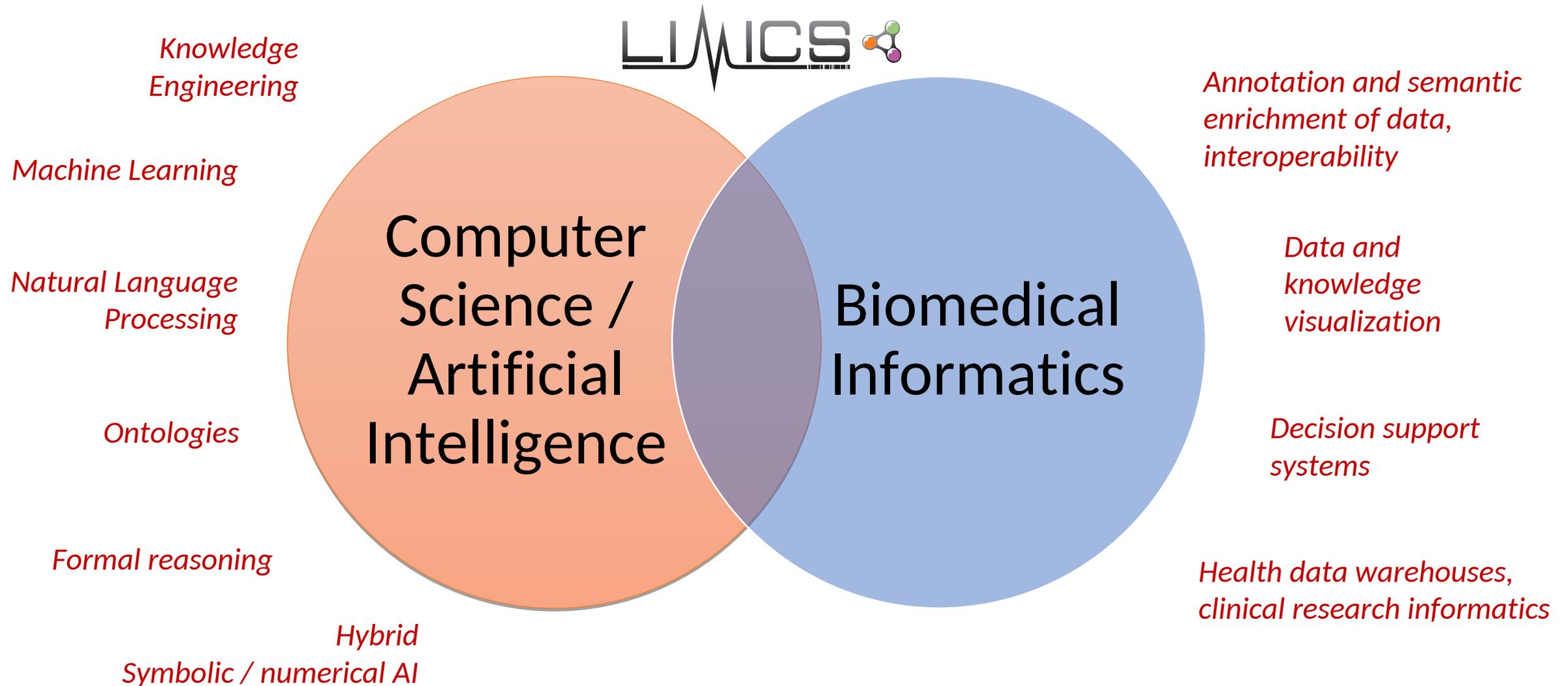
Sites AP-HP

- Hôpital Tenon
- Hôpital Avicenne
- Hôpital Charles Foix
- Hôpital La Pitié Salpêtrière
- DSN AP-HP

And our own site in Hôpital Trousseau



Disciplinary fields of research



Themes

M₃P

**Drug representation:
pharmaceutical knowledge,
prescription, pharmacovigilance**

Catherine DUCLOS
Marie-Christine JAULENT



**Clinical data warehouses &
cohorts /
Secondary use of healthcare data**

EDIRC

Christel DANIEL

**Decision support for care,
management, public health and
healthcare research**

ADES

Emmanuelle KEMPF
Brigitte SÉROUSSI

**Research in informatics and
health informatics**

RIIS

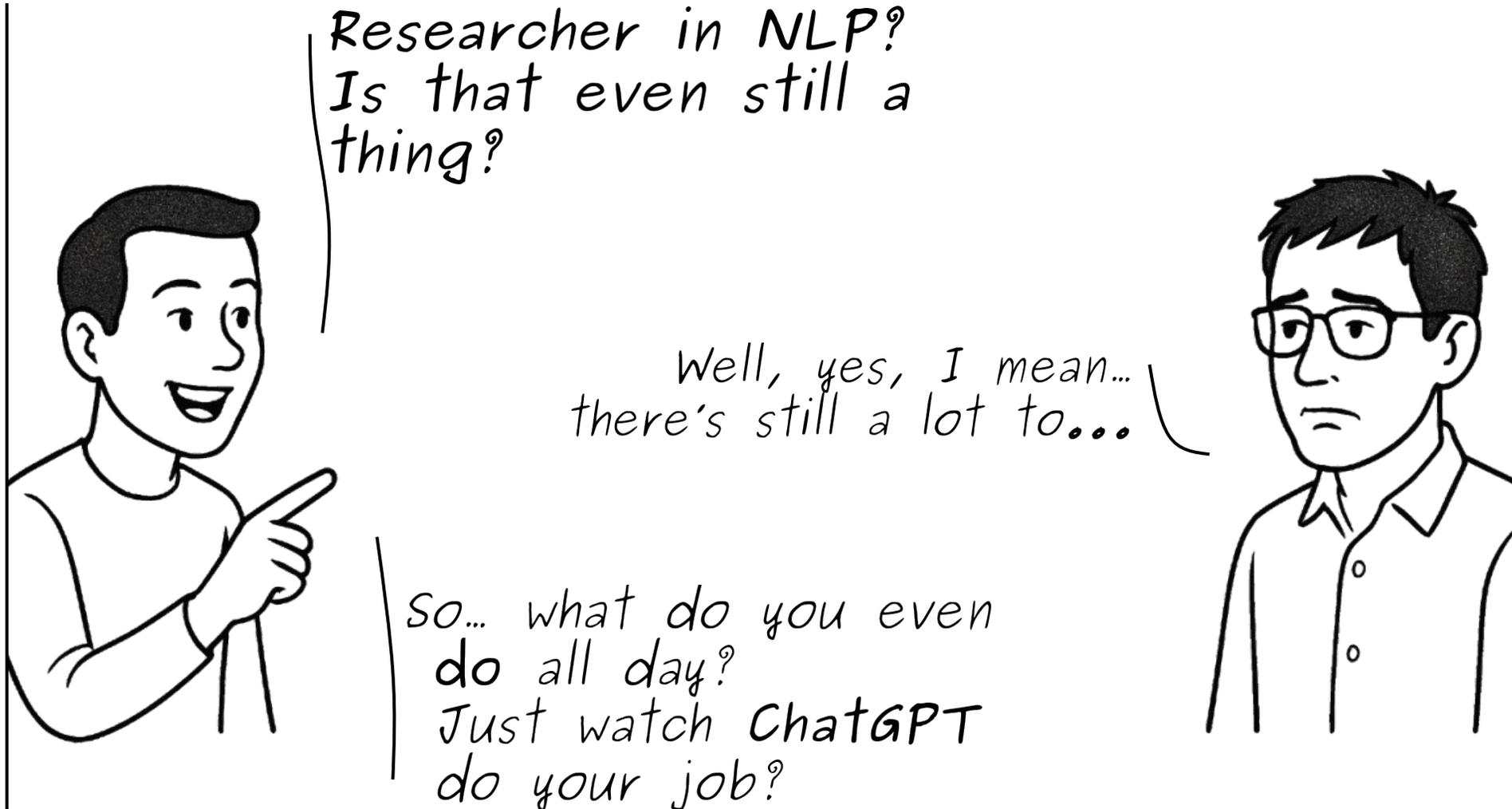
Jean-Baptiste LAMY
Chan Le Duc

Who am I?

Xavier Tannier

- Professor in Computer Science at Sorbonne Université
- Researcher at Limics
- Natural Language Processing, Information extraction, machine learning... applied to health data

Being an NLP researcher in 2025



Being a *clinical* NLP researcher in 2025



Hi, could you build me ASAP a chatbot extracting data about complications of [REDACTED], in surgery for [REDACTED], from patients' reports?

} That's only 176 variables to extract

} I have no example to show you because it's pretty rare

} But ChatGPT will do the job, right?

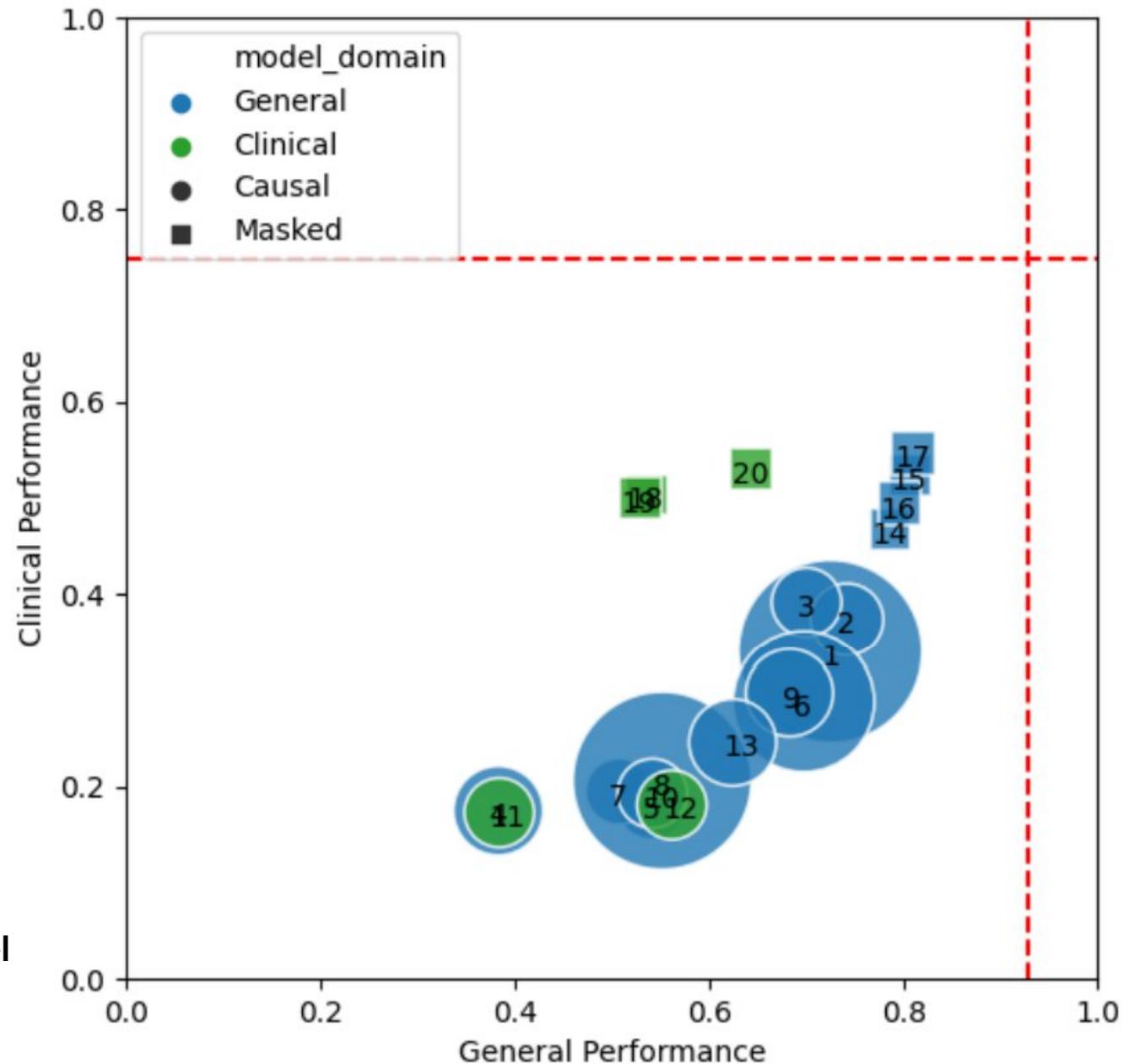


... And yet (1/2)

In few-shot clinical entity recognition: masked language models outperform generative model prompting

(Naguib et al., 2024)

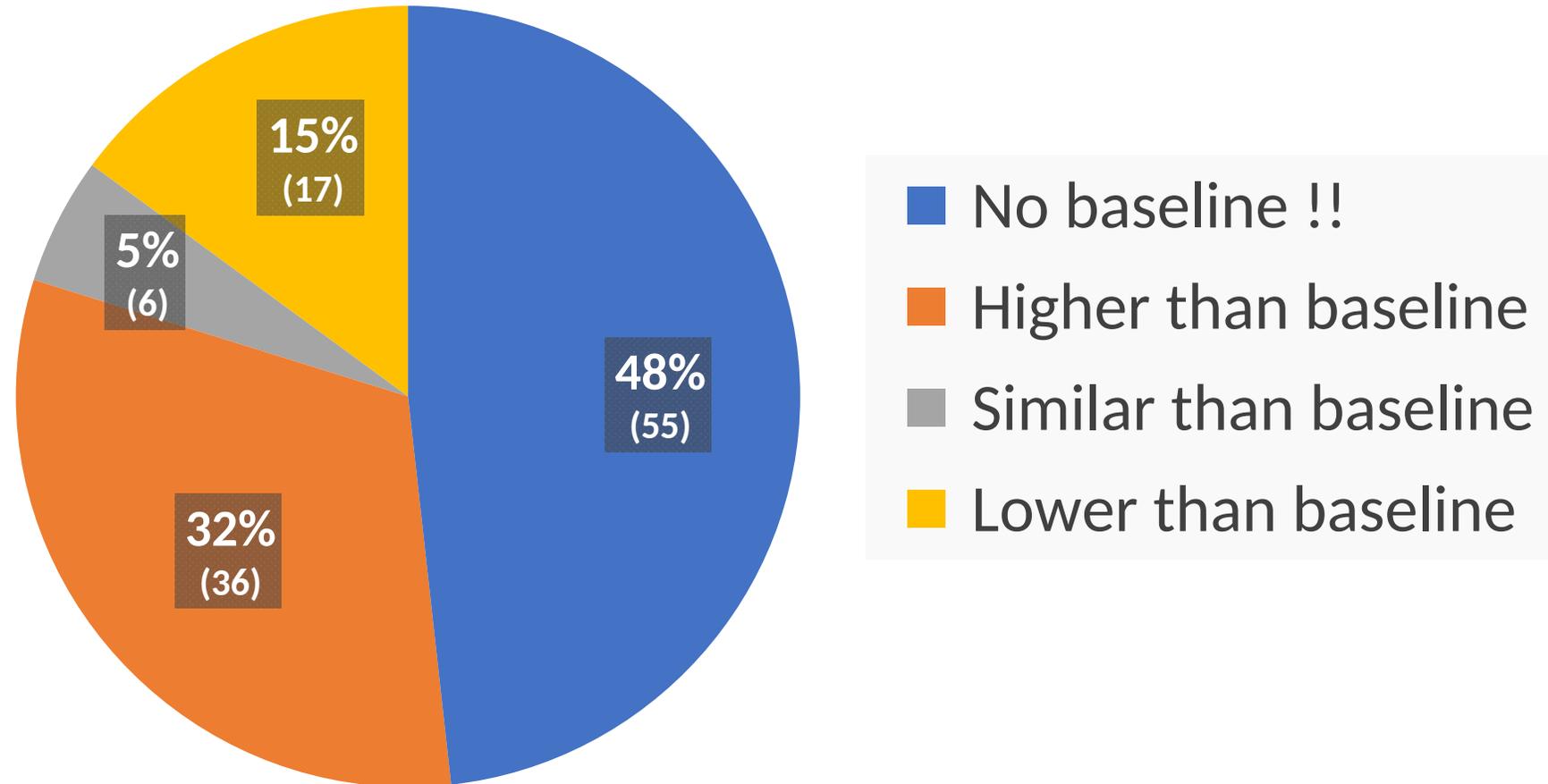
Marco Naguib, Xavier Tannier, Aurélie Névéol.
Few-shot clinical entity recognition in English, French and Spanish: masked language models outperform generative model prompting.
in *Findings of EMNLP 2024*.



... And yet (2/2)

Comparison of LLM prompt engineering versus other approaches, through a systematic literature review

(Zaghir et al., 2024)



Natural Language Processing

&

Healthcare data

Why Natural Language Processing?

ASSISTANCE HÔPITAUX PUBLIQUE DE PARIS HÔTEL - DIEU



HÔTEL-DIEU
1, place du Parvis Notre-Dame
75181 PARIS Cedex 04

Univ'ité Paris V
René Descartes

POLE IMAGERIE ET
EXPLORATIONS
FONCTIONNELLES

SERVICE DE
MEDECINE NUCLEAIRE
Secrétariat: 01 42 34 82 41
Télécopie: 01 42 34 83 13
Installation 40 BS 441 L2BT

Chef de Service:
Pr N (JL. MORETTI)

Responsable:
Dr N. CAILLAT-VIGNERON
Maître de Conférence
Professeur Hospitalier
Biophysique, Médecine Nucléaire
Hématologie
01 42 34 82 41
nailce.caillatvigneron@hd.aphp.fr

Dr J.F. TOUSSAINT
Maître de Conférence
Professeur Hospitalier
Physiologie, Cardiologie
01 42 34 82 40
jean - francois.toussaint@hd.aphp.fr

Dr B. BOURGEOIS
Assistant Hospitalier-attaché
01 42 34 82 42
benoit.bourgeois@hd.aphp.fr

Dr E. BLOCH
Dr T. KONDO *Attaches
de Cardiologie*

Radiopharmacie:
G. H. DEEB
01 42 34 82 42

Cadre Médico-Technique:
M. G. GARCIA
01 42 34 85 97
michel.garcia@hd.aphp.fr

Nom :
Prénom :
Date Naissance : 27/10/1955
NIP : 2407093801

Demandeur : Médecin de Ville 75009
Docteur : EVRRARD-CTR MED EUROPE
N° de Demande : 7070473
Date d'examen : 05/12/2007

EV
SCINTIGRAPHIE AU GALLIUM 67

CLINIQUE:
Myofasciite à macrophages, dans un contexte d'aggravation de l'impotence fonctionnelle.
Première scintigraphie ; biopsie montrant deux îlots caractérisés par des lymphocytes sur le deltoïde.

TECHNIQUE: Examen réalisé 48 heures après injection d'un traceur.
Images par plans centrés sur gamma-caméra: Symbia Siem...

RESULTAT:
Les clichés réalisés sur la ceinture scapulaire montrent une fixation focale à l'épaule gauche, plus ou moins homogène, à la moitié supérieure des deux bras.
Sur la ceinture pelvienne, on note une activité musculaire focale à l'ischion fessier et des muscles des cuisses de façon très prédominante.
Les articulations des hanches sont également relativement normales.
Sur la partie inférieure des cuisses et sur les mollets, on note une fixation focale, bien délimitée, hétérogène au maximum pour les deux chevilles, de même que sur les deux chevilles. Sur la partie inférieure des cuisses et sur les mollets, on note une fixation focale, bien délimitée, hétérogène au maximum pour les deux chevilles. Sur la partie inférieure des cuisses et sur les mollets, on note une fixation focale, bien délimitée, hétérogène au maximum pour les deux chevilles. Sur la partie inférieure des cuisses et sur les mollets, on note une fixation focale, bien délimitée, hétérogène au maximum pour les deux chevilles.

CONCLUSION:
Fixation musculaire hétérogène et bien caractéristique, indifférenciée, aux chevilles, les genoux, les hanches et les épaules.
Il est bien difficile de faire la part entre la responsabilité de la fixation focale des entéropathies qui sont constantes chez de nombreux patients et celle de la fixation focale musculaire.
Merci de me tenir au courant, cordialement.

Docteur

05/12/2007 14:00
GALLIUM 67
GA 67 138,51
MBq
GA 67 N°Lot:96282
Lot: Trousse

Centre Hospitalier Régional Universitaire
7, rue Pasteur
67001 STRASBOURG Cedex
Tél. 03 88 12 34 78

COMPTE RENDU OPERATOIRE En date du : 15/06/2010

E... Edouard Né le 15/05/1955
Opérateurs : Didier D..., Fabien F...
Assistant : Jeanne J...
Anesthésiste : Gérard G... Anesthésie : anesthésie générale

Cholécysectomie - Cholangiographie - Cœloscopie

Indication
Patient de cinquante-cinq ans, adressé par le Dr A... pour la prise en charge d'une lithiase vésiculaire symptomatique connue de longue date.

Il a comme antécédent une appendicectomie.
Son poids est de 105 kg pour 1,70 m.
Il n'a pas de traitement en cours.

Il est donneur de sang de manière régulière : a priori il n'y a pas d'anémie.
Le patient signale des douleurs abdominales diffuses qui remontent à 2005.

Depuis un an, il décrit plus particulièrement des douleurs d'apparition rapide, avec pléiopathies disparates dans les deux heures qui suivent, au niveau de l'hypochondre droit.
La plainte principale concerne ces douleurs dont l'intensité a été progressivement croissante jusqu'à ce jour.

Echographie des voies biliaires (01/06/2010) : calcification hyperéchogène mobile, décrit un calcul de 2 x 2 cm avec cône d'ombre postérieur.

Indication de cholécystectomie par cœloscopie plutôt que laparotomie, vu l'absence de complications. Information modalités, bénéfices, risques.

Intervention

Création du pneumopéritoine par méthode « ouverte ».
Mise en place de quatre trocarts (optique à l'ombilic, épigastrique, latéral droit, opérateur à gauche de la ligne médiane).
Cœloscopie montrant une vésicule tendue et volumineuse, aux parois œdématisées recouvertes d'importantes adhérences de l'épiploon, de libération difficile.

[E... Edouard, CRO 15/06/2010, 1/2]

Top sujets				
Page	1	2	3	4
Page	Précédente			Suivante
Sujets de la catégorie Diabète	Dern. page	Auteur du sujet	Nombre de Rép.	Lu
Le Parc Astérix fête ses 30 ans, venez passer un bon moment en famille Jouez et tentez de remporter 4 entrées pour le parc !				
À lire avant de poster : La modération sur Doctissimo		Jacoline	0	626
Diabète, insuline, activités physiques : comment gérer après 50 ans ?	37	14 690	13/11/2015 à 20:03	Cyril
Topics utiles.		James	0	49 5
"le jeu à la con" bis repetita	1629	carangevine	81 405	837 5
on poste sa glycémie du matin!	1214	Cyril	60 677	886 0
bbeloop un système qui a du plomb dans le	5	Cyril	203	2 99
CEST ?	13	Crodlieundee	615	3 23
angement de traitement car changement diagnostic = hypo et glycémie basse		Cris29	2	25
limes 640g		Mimichoco	3	43
WS PERMIS DE CONDUIRE	2	ledom35	78	838
ur se dire bonsoir, faire un coucou ou tres nouvelles	108	loi17dc	5 395	30 5
bbète et perte de poids : ???:		lem50zg	41	21 2
ssible DT1 ?		toune54	5	73
yennes lecteurs de glycémie	99	petitsoleil33	4 941	91 5
potages: nos difficultés autour du bête..		Crodlieundee	12	163



Relation Extraction from News Articles (RENA): A Tool for Epidemic Surveillance

Jaeff Hong², Duong Dung¹, Danielle Hutchinson¹, Zubair Akhtar¹, Rosalie Chen¹, Rebecca Dawson¹, Aditya Joshi², Samsung Lim³, C Raina MacIntyre¹ and Deepti Gurdasani^{1,4,5},

¹Kirby Institute, University of New South Wales, Sydney, Australia

²School of Computer Science & Engineering, University of New South Wales, Sydney, Australia

³School of Civil and Environmental Engineering, University of New South Wales, Sydney, Australia

⁴William Harvey Research Institute, Queen Mary University of London, London, UK

⁵School of Medicine, University of Western Australia, WA, Australia

jaeff.hong@gmail.com, {duong.dung, danielle.hutchinson, zubair.akhtar, rosalie.chen, rebecca.dawson, aditya.joshi, s.lim, r.macintyre, d.gurdasani}@unsw.edu.au

Abstract

Relation Extraction from News Articles (RENA) is a browser-based tool designed to extract key entities and their semantic relationships in English language news articles related to infectious diseases. Constructed using the React

News Articles (RENA), uses decoder-only foundation models (also known as 'large language models', *i.e.*, LLMs) to extract semantic relations in infectious disease-related news articles in the English language¹ at the document level. When an epidemiologist enters a news article, RENNA extracts a list of entities and relations present in the article. This streamlines an automated process for epidemiologists and researchers who need a method to acquire a large congregate of structured relations from their selected article. By making use of RENNA, they can be aided in their research without having to read a large set of news articles or delve into manual data curation.

Whilst many previous relation extraction (RE) tasks have focused on the sentence level, it is evident that many relations exist between different sentences, presenting another challenge in extracting relations at the document level (Xu, Chen, and Zhao 2021). We assume a simplistic definition of a relation: a relation connects exactly two entities. In the context of epidemic intelligence, the sentence 'A patient died due to COVID-19 today' results in the relations {death number: '1', relation: 'death of', infectious disease: 'COVID-19'}, {death number: '1', relation: 'occurred on', event date: 'today'} and {infectious disease: 'COVID-19', relation: 'occurred on', event date: 'today'}. In this specific case, if two relations are identified, the third can be inferred.

RENA can be used by epidemiologists, public health officials and teachers or students of public health to extract information from news articles of interest. While the utility of RENNA is for infectious disease-related epidemic intelligence, it can potentially be used for news articles across domains and application areas, including journalists/news publishers who want to verify and investigate information across multiple sources.

Architecture

Figure 1 shows the architecture of RENNA. In the model training phase, we generated a set of 300 synthetic articles, annotated with relevant entities and relationships

ALIGNING LARGE LANGUAGE MODELS FOR CLINICAL TASKS

Supun Manathunga
University of Peradeniya
NeuraSense Research

Isuru Hettigoda
NeuraSense Research

ABSTRACT

Large Language Models (LLMs) have demonstrated remarkable adaptability, showcasing their capacity to excel in tasks for which they were not explicitly trained. However, despite their impressive natural language processing (NLP) capabilities, effective alignment of LLMs remains a crucial challenge when deploying them for specific clinical applications. The ability to generate responses with factually accurate content and to engage in non-trivial reasoning steps are crucial for the LLMs to be eligible for applications in clinical medicine. Employing a combination of techniques including instruction-tuning and in-prompt strategies like few-shot and chain-of-thought prompting has significantly enhanced the performance of LLMs. Our proposed alignment strategy for medical question-answering, known as 'expand-guess-refine', offers a parameter and data-efficient solution. A preliminary analysis of this method demonstrated outstanding performance, achieving a score of 70.63% on a subset of questions sourced from the USMLE dataset.

Is - Clinical Applications - Alignment Strategy - Medical Question-Answering

Artificial Intelligence (AI) research was mainly focusing on specific tasks like mastering the game of Go, solving complex mathematical problems, and generating human-like text. The advancement of the deep learning techniques, particularly the transformer models, has laid the groundwork for Large Language Models (LLMs), which exhibit capabilities in tasks they weren't explicitly trained for, a phenomenon observed as these models scale [4]. The development of these expansive LLMs may be bringing us closer to the general intelligence [5], [6].

Text corpora containing medical knowledge and this knowledge becomes ingrained in the task-agnostic nature of LLMs, they find utility across a spectrum of clinical medicine, it is imperative for these models to aptly grasp the nuances of clinical medicine, and engage in reasoned analysis with a certain level of discernment. Mechanisms for content moderation, guarding against harmful content, and ensuring the model's alignment with

LLM capabilities exhibited by LLMs, they need to be aligned before deploying for clinical applications. This paper introduces a novel alignment strategy for LLMs, designed to align LLMs with clinical tasks.

NLP tasks

9 blood cultures the 26/6/15 to
Staphylococcus aureus methicilline sensitive.

BACTERIA RESISTANCE

VASCULAR DISEASE

Patient with arterial ulcers of the lower limbs,
followed at St Joseph (Dr Wyliana) with skin
graft in October 2015

PROCEDURE

ANATOMY MEDICAL DEVICE

Contraception with intrauterine devices

Patient with no evident sign
of cranial traumatism [fact:neg]
[C0018674] (Craniocerebral Trauma)

Patient with history of
bariatric surgery
[fact:patient history]
[C1456587] (Bariatric Surgery)

Suspicious masses are observed in the right breast.
The exam is rated ACR 2 on both right and left.

The diagram illustrates the localization of terms in the text. Three curved arrows labeled 'localization' connect the following pairs of terms: 'right breast' (from the first sentence) to 'right' (from the second sentence), 'right breast' to 'left' (from the second sentence), and 'right and left' (from the second sentence) to 'right' (from the second sentence).

NLP tasks

MAMMOGRAPHY:

There is a **1.8 cm round mass** with a circumscribed margin in the **left breast** in the **anterior depth central** to the nipple. There also is a **1.4 cm oval mass** with an obscured margin in the **left breast** in the **anterior depth of the inferior region**.

LOCALIZATION

ASSESSMENT: **BI-RADS Category 3**

SCORE	LATERALITY	LESION	LATERALITY	LOCALIZATION	SHAPE
BI-RADS 3	LEFT	mass	LEFT	QSI	1.8
BI-RADS 1	RIGHT	mass	LEFT	QI	1.4

NLP tasks

ASSISTANCE HÔPITALIÈRE
PUBLIQUE DE PARIS
HÔTEL - DIEU

NOM: M. M. Prénom: M. M. Date de Naissance: 20/01/1958 N° de Demande: 120422
Date d'admission: 08/12/2007 Date d'expiration: 07/12/2007
Service: SCINTIGRAPHIE AU GALIUM 67

DIAGNOSTIC: Myocardie à ischémie, dans un contexte d'aggravation récente des douleurs et de l'aggravation des troubles. Pratiques radiologiques: hépatite massive dans l'abdomen associé à des lymphes pleins ou à des adénopathies.

TECHNIQUE: Examen réalisé 48 heures après l'injection de 18.3 MmCi de Citrate de Gallium 67. Images par plan antérieur sur gamma caméra: Système Scintex.

RESULTAT: Les clichés réalisés sur la caméra scintillaire montrent une bonne délimitation de l'artériation des splénes avec une activité affaiblie en raison de l'aggravation des troubles. Sur les autres parties, on voit une activité normale à un niveau homogène en raison des splénes. Les splénes sont donc considérés comme des splénes de type normal. Les artériations des splénes sont également radiologiquement normales. Sur la partie hépatique, on voit des splénes massifs homogènes plus ou moins actifs, sans adénopathies, hépatome ou tumeur primitive de part et d'autre des splénes. Il ne s'agit pas d'un processus localisé de type tumeur, on peut dire qu'il s'agit d'un processus hépatique de type normal. Sur les autres parties, il existe une activité hépatique normale.

CORRECTION: Examen réalisé 48 heures après l'injection de 18.3 MmCi de Citrate de Gallium 67. Images par plan antérieur sur gamma caméra: Système Scintex.

Reason for admission = low-kinetic fracture?

Family history?

Endogenous or exogenous Cushing's syndrome ?
(complications due to glucocorticoid excess)

ASSISTANCE HÔPITALIÈRE
PUBLIQUE DE PARIS
HÔTEL - DIEU

NOM: M. M. Prénom: M. M. Date de Naissance: 20/01/1958 N° de Demande: 120422
Date d'admission: 08/12/2007 Date d'expiration: 07/12/2007
Service: SCINTIGRAPHIE AU GALIUM 67

DIAGNOSTIC: Myocardie à ischémie, dans un contexte d'aggravation récente des douleurs et de l'aggravation des troubles. Pratiques radiologiques: hépatite massive dans l'abdomen associé à des lymphes pleins ou à des adénopathies.

TECHNIQUE: Examen réalisé 48 heures après l'injection de 18.3 MmCi de Citrate de Gallium 67. Images par plan antérieur sur gamma caméra: Système Scintex.

RESULTAT: Les clichés réalisés sur la caméra scintillaire montrent une bonne délimitation de l'artériation des splénes avec une activité affaiblie en raison de l'aggravation des troubles. Sur les autres parties, on voit une activité normale à un niveau homogène en raison des splénes. Les splénes sont donc considérés comme des splénes de type normal. Les artériations des splénes sont également radiologiquement normales. Sur la partie hépatique, on voit des splénes massifs homogènes plus ou moins actifs, sans adénopathies, hépatome ou tumeur primitive de part et d'autre des splénes. Il ne s'agit pas d'un processus localisé de type tumeur, on peut dire qu'il s'agit d'un processus hépatique de type normal. Sur les autres parties, il existe une activité hépatique normale.

CORRECTION: Examen réalisé 48 heures après l'injection de 18.3 MmCi de Citrate de Gallium 67. Images par plan antérieur sur gamma caméra: Système Scintex.

Diabetic patient?

Smoker?

Able to make his/her own decision?

Objectives

Selecting patients
matching criteria

Pseudonymising patient
records

Selecting
similar patients

Structuring
data

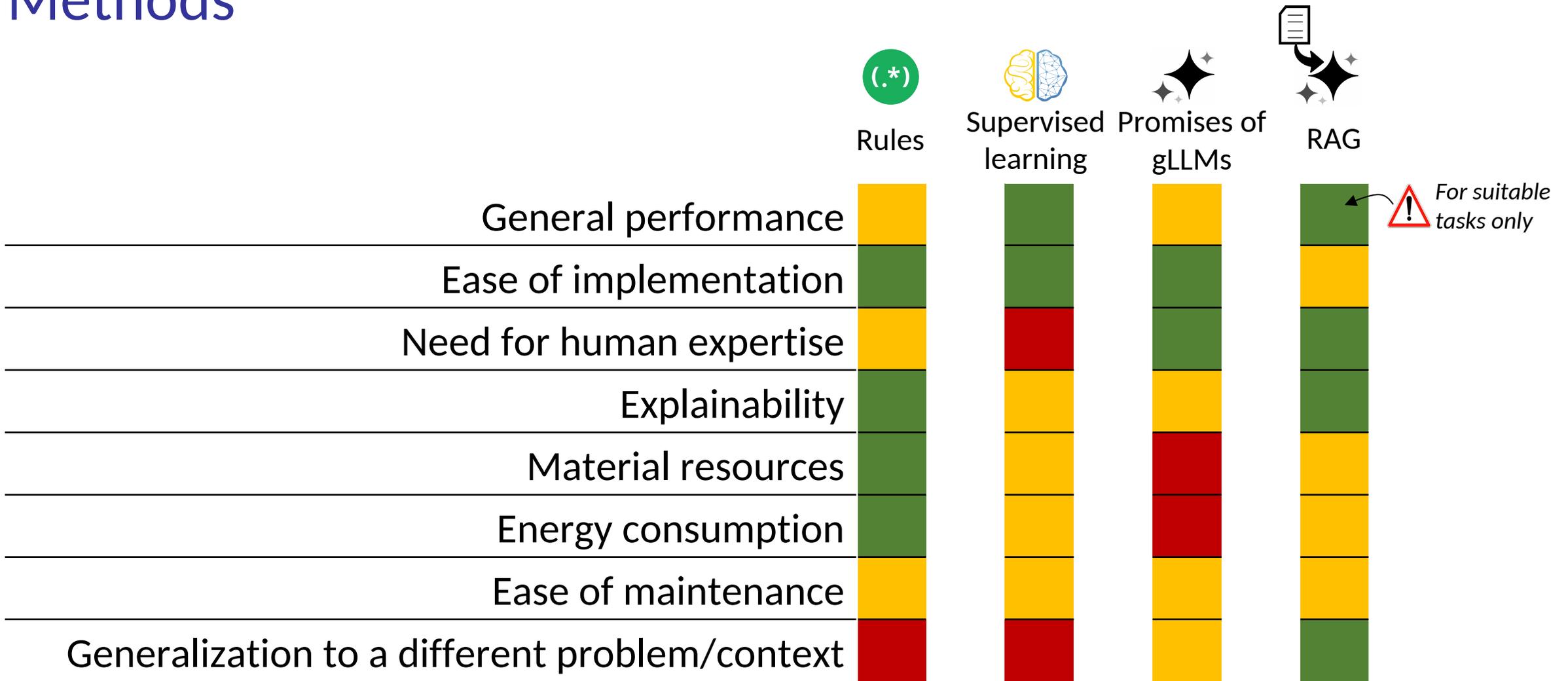
Summarizing and visualizing
patient's history

Indexing &
Retrieving

Pre-filling
forms

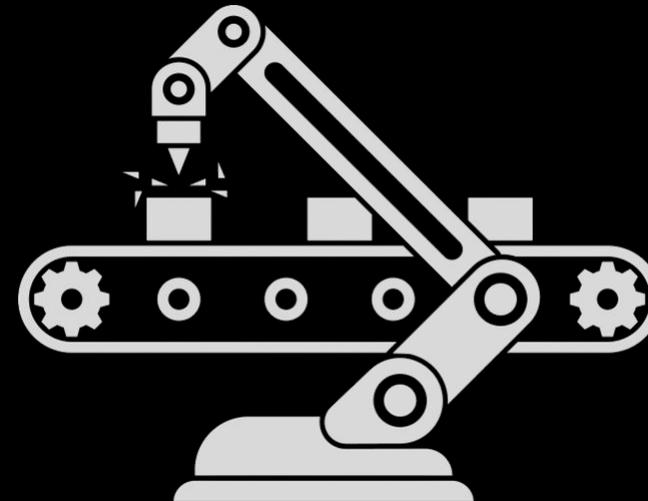
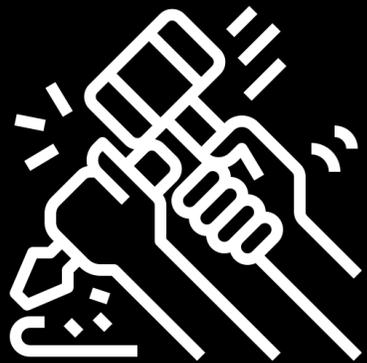
...

Methods



[see backup slides for more details...](#)

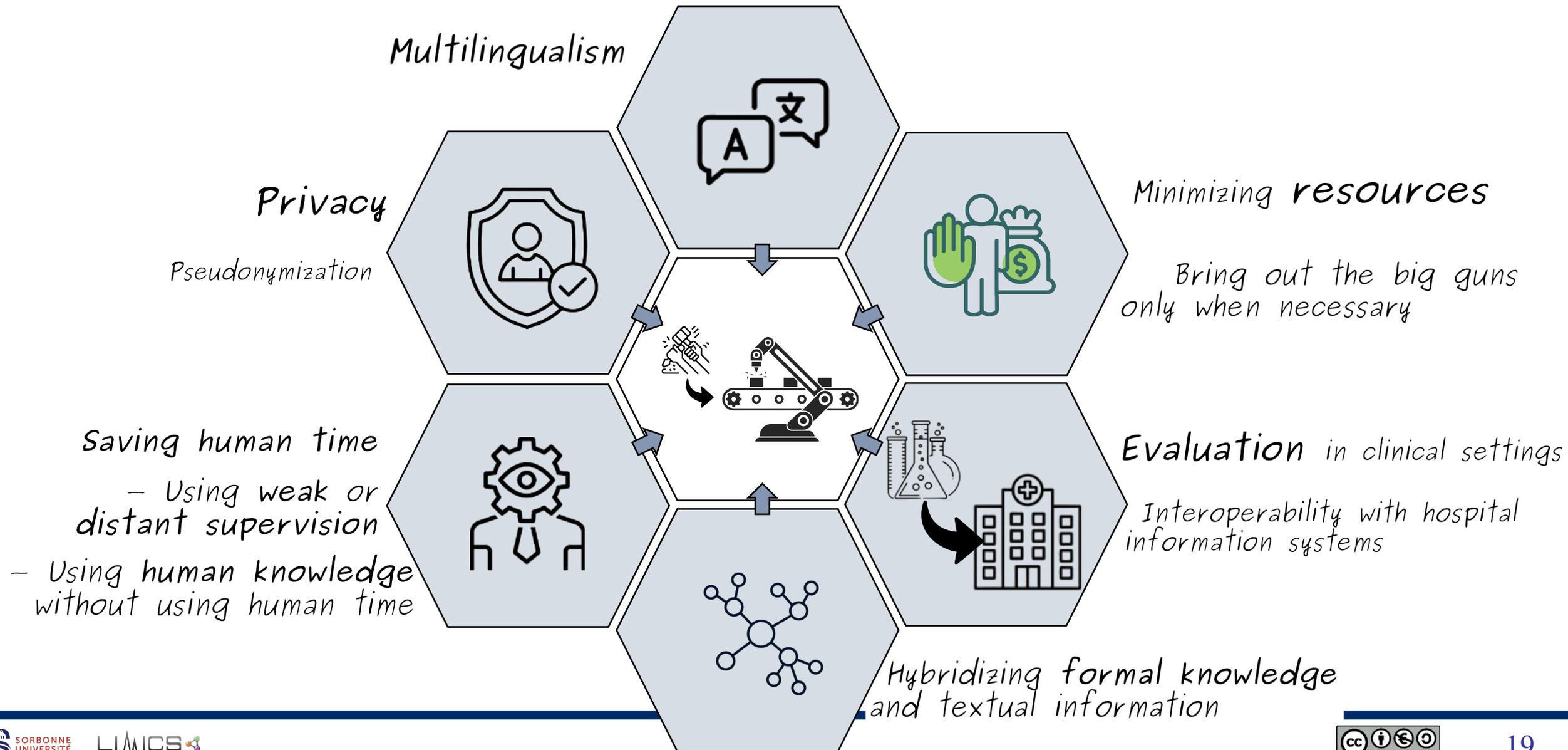
The challenges of industrializing NLP for health data



Disclaimer



The challenges of industrializing NLP for health data



Example 1: Pseudonymization

Privacy

Pseudonymization

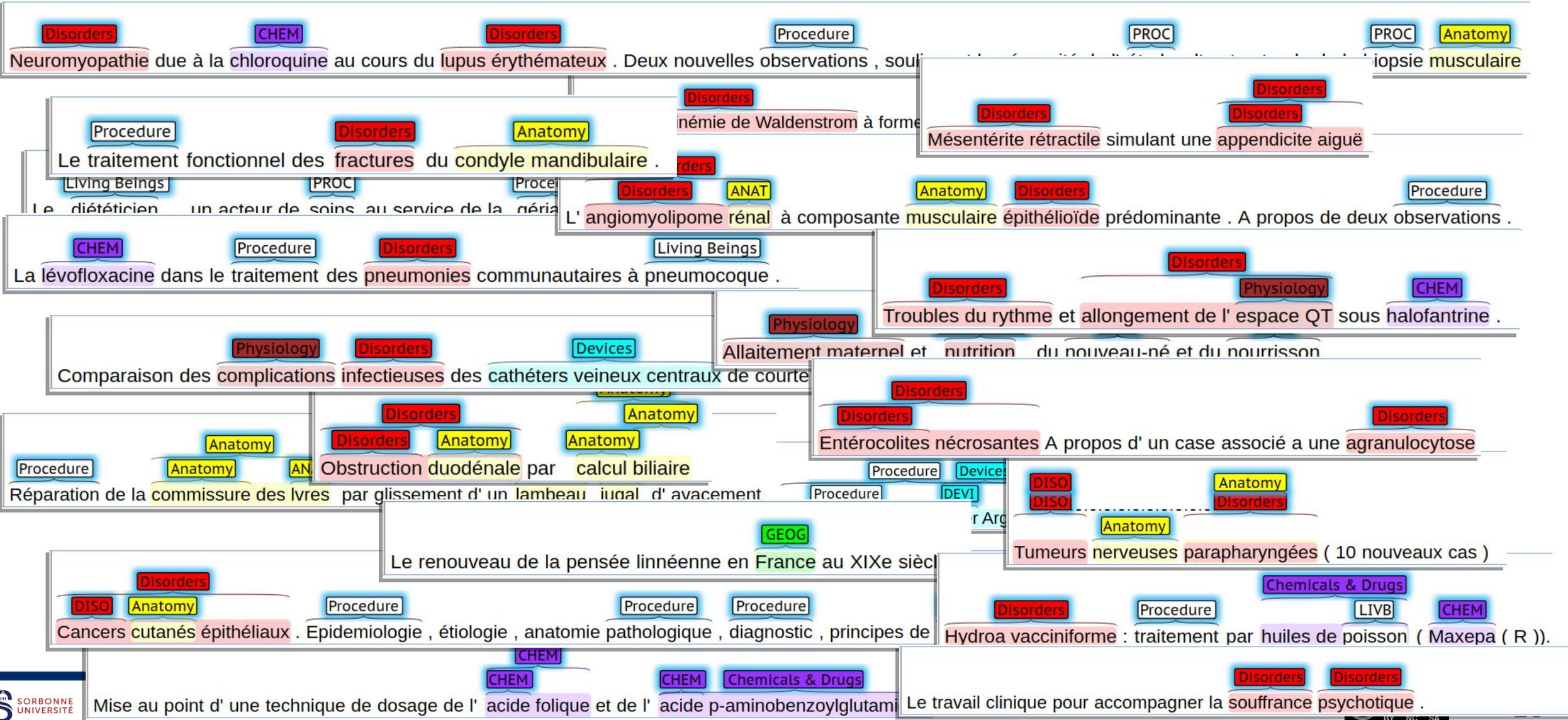
with

ASSISTANCE PUBLIQUE
HÔPITAUX DE PARIS
Entrepôt de Données de Santé



Supervised learning systems

Névél, A., Grouin, C., Leixa, J., Rosset, S., & Zweigenbaum, P. (2014).
The Quaero French Medical Corpus: A Ressource for Medical Entity Recognition and Normalization.



Supervised learning systems

Névéol, A., Grouin, C., Leixa, J., Rosset, S., & Zweigenbaum, P. (2014).
The Quaero French Medical Corpus: A Ressource for Medical Entity Recognition and Normalization.

Neuromyopathie due à la chloroquine au cours du **lupus érythémateux** . Deux nouvelles observations , soulignant la nécessité de l' étude ultrastructurale de la biopsie musculaire

Hémangiomes choroïdiens circonscrits traités par photocoagulation

Composition musculaire **épithélioïde** prédominante . A propos de deux observations .

Le traitement fonctionnel

Le travail clinique pour

Comparaison des complications infectieuses des cathéters

Macroglobulinémie à forme pancytopénique . Rémission complète

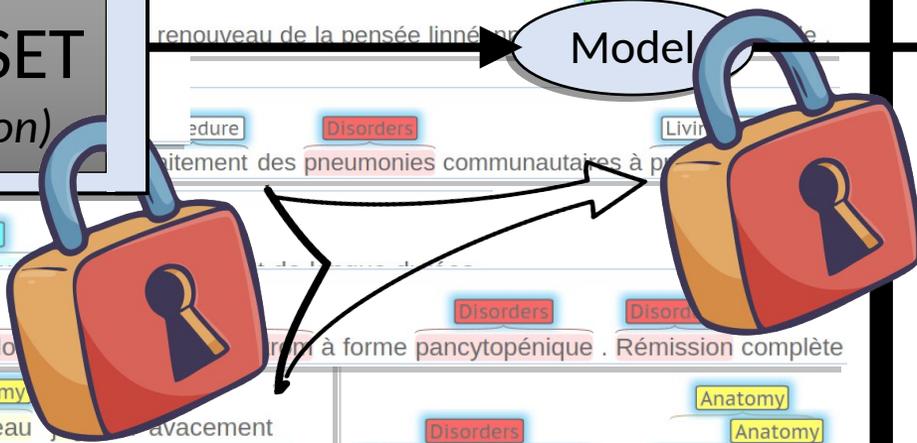
Réparation de la commissure des lèvres par glissement d' un lambeau

Allaitement maternel et nutrition du nouveau-né et du nourrisson

Cancers cutanés épithélioïdes . Epidemiologie , étiologie , anatomie pathologique , diagnostic , principes de traitement .

Labels: Disorders, CHEM, Procedure, PROC, Anatomy, DEVI, ANAT, Physiology, DISO, ANATOMY, LIVB, DISO, ANATOMY, PROCEDURE.

Learn to generalize
TRAINING SET
(train + validation)



Mésentérite rétractile simulant une appendicite aiguë

Mise au point d' une technique de dosage de l' acide folique et de l' acide p-aminobenzoylglutamique

Troubles du

Le diététicien , un acteur de soins au service de la gériatrie .

Entérocolites nécrosantes A propos d' un case associé a une agranulocytose

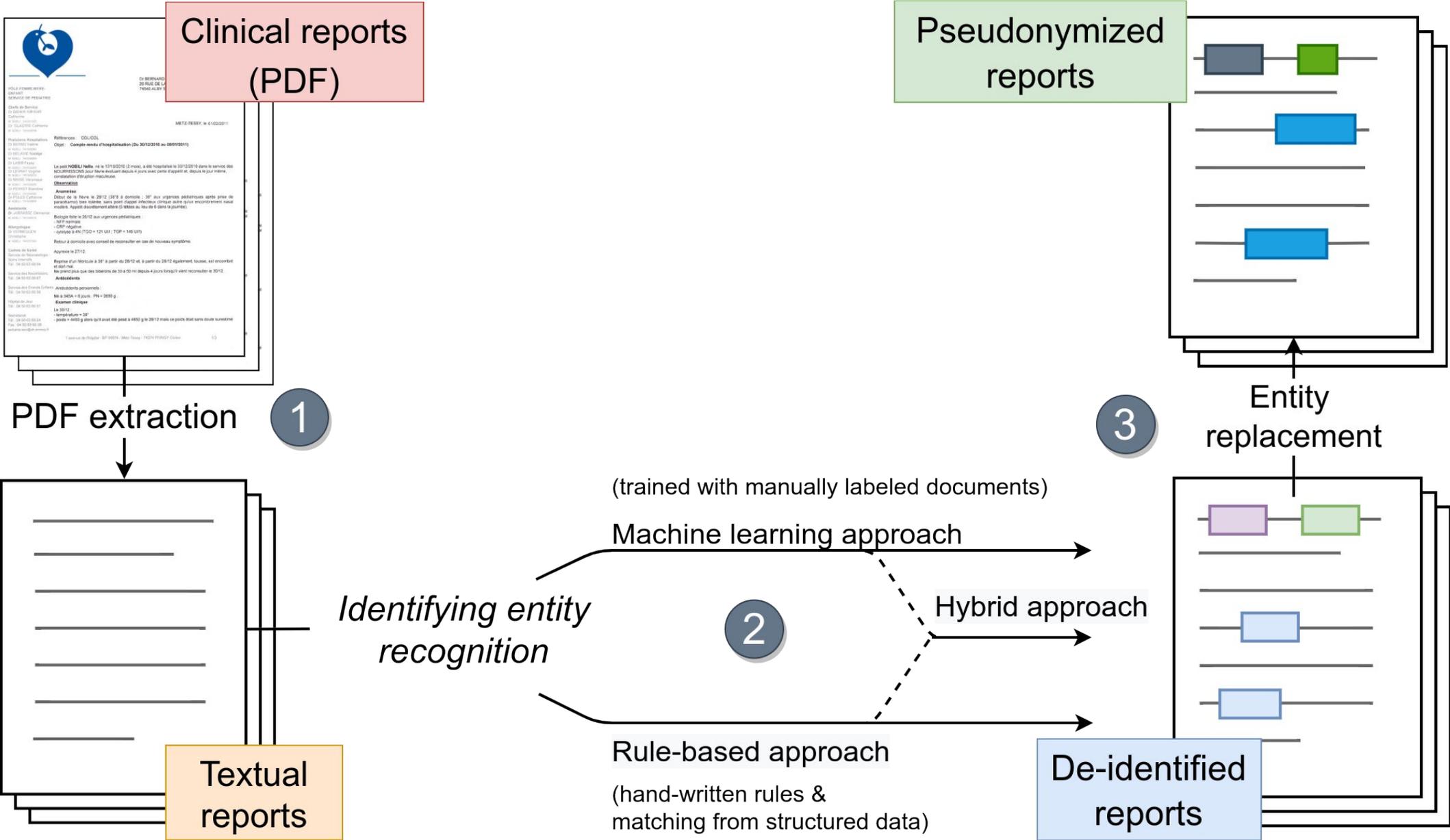
Tumeurs nerveuses parapharyngées (10 nouveaux cas)

Hydroa vacciniiforme : traitement par huiles de poisson (Maxepa (R)).

Labels: Disorders, CHEM, Chemicals & Drugs, DISO, ANATOMY, LIVB, PROCEDURE.

Evaluation
TEST SET

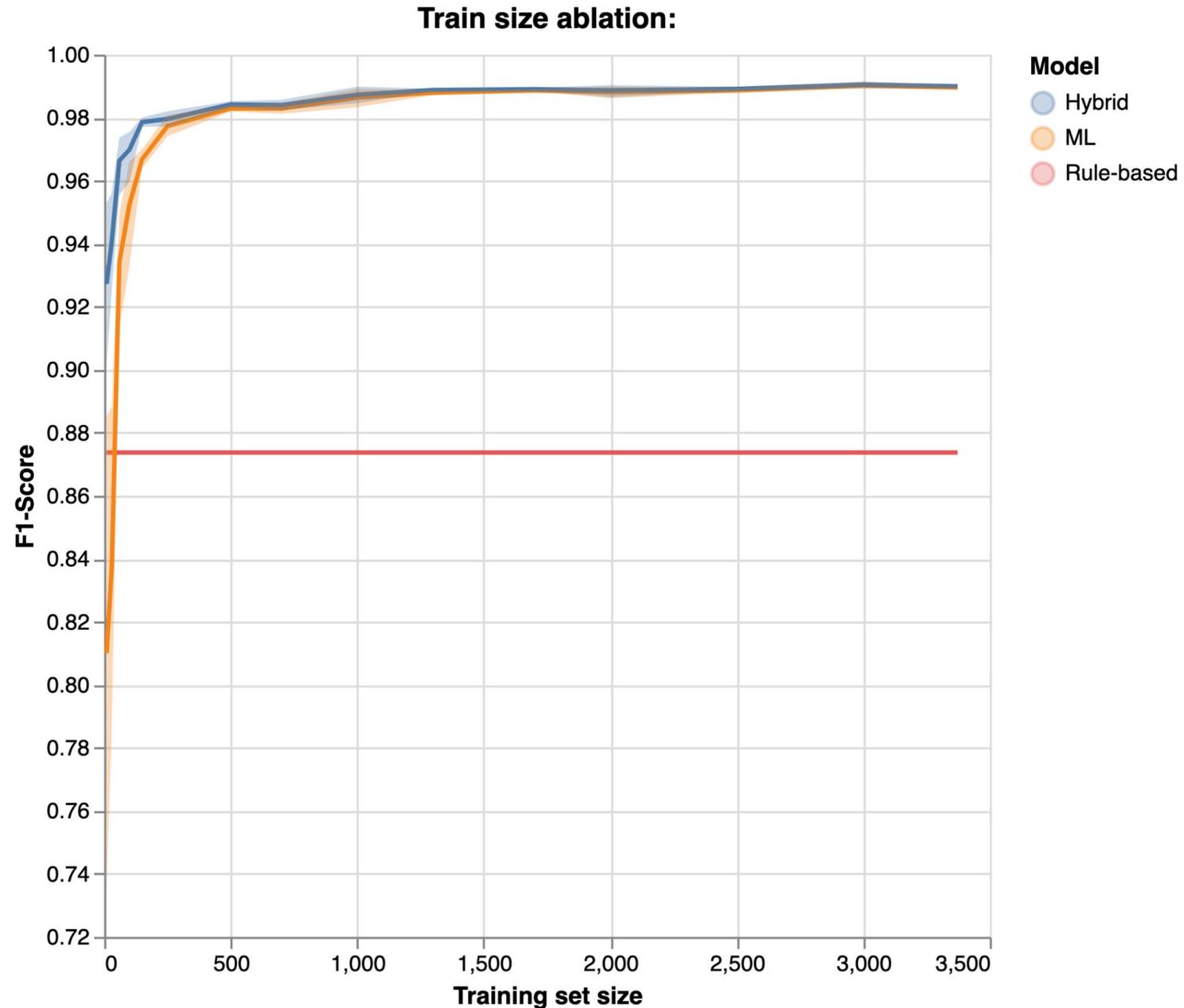
Pseudonymisation



Pseudonymisation

A **cookbook** to build your own pseudonymization pipeline :

- Training set size
- Document types
- Language models
- Methods (only ML or ML+rules)



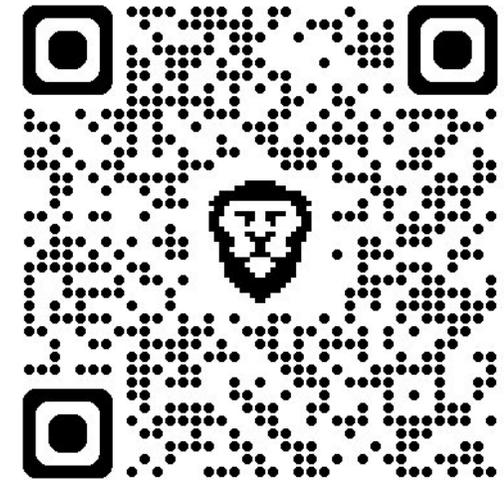
Pseudonymisation

EDS-Pseudo, an AP-HP library for pseudonymisation:

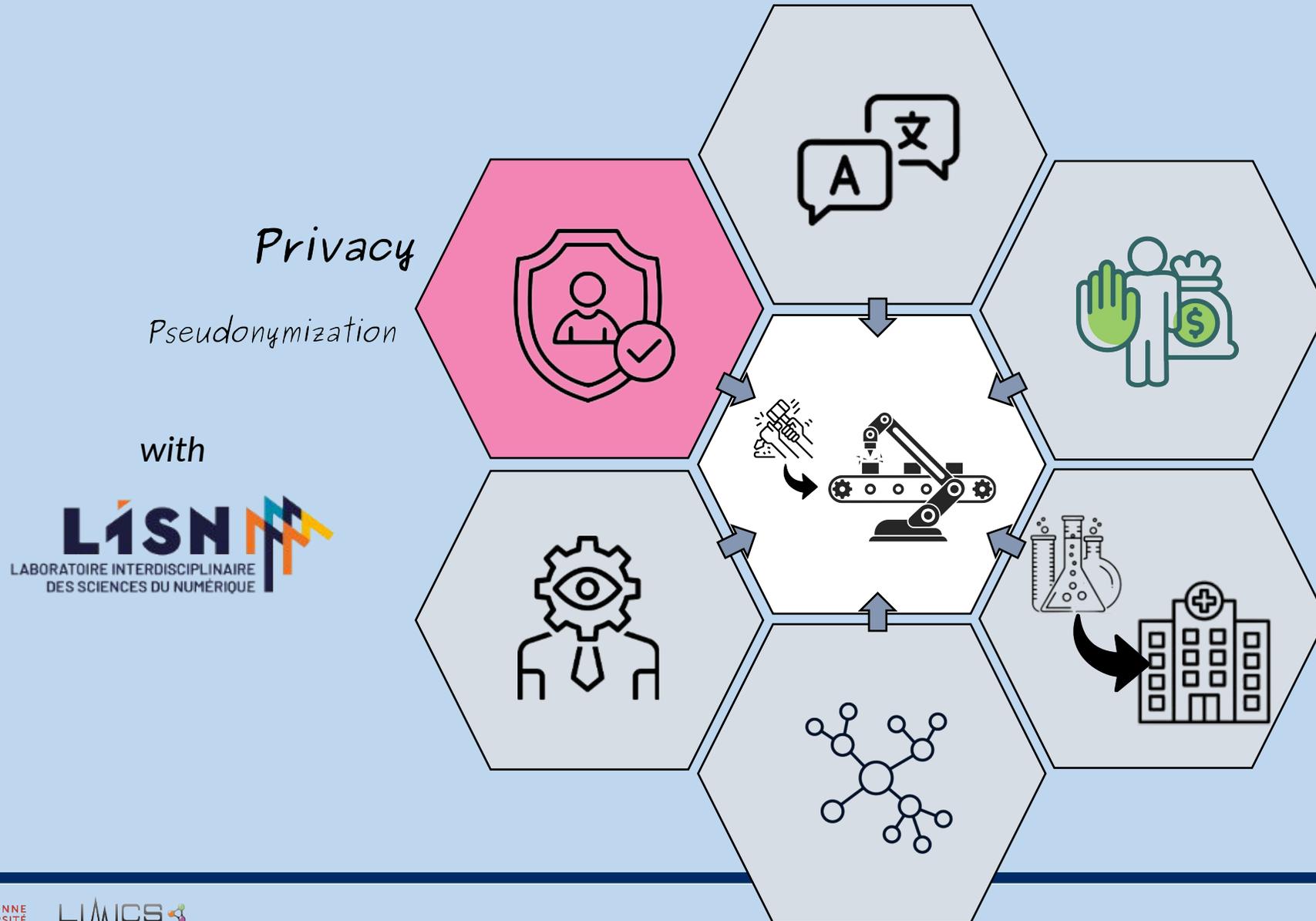
- Code to train your pipeline on your own data
- A **public model trained on fake data** (handwritten + GPT3 augmentation)

Label	Description
ADRESSE	Street address, eg 33 boulevard de Picpus
DATE	Any absolute date other than a birthdate
DATE_NAISSANCE	Birthdate
HOPITAL	Hospital name, eg Hôpital Rothschild
IPP	Internal AP-HP identifier for patients, displayed as a number
MAIL	Email address
NDA	Internal AP-HP identifier for visits, displayed as a number
NOM	Any last name (patients, doctors, third parties)
PRENOM	Any first name (patients, doctors, etc)
SECU	Social security number
TEL	Any phone number
VILLE	Any city
ZIP	Any zip code

<https://github.com/aphp/eds-pseudo>

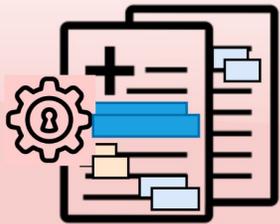


Example 2: Privacy-preserving models



Privacy-preserving mimic models

Private



Training



Teacher model

Sensitive labeled data

Preprocessing steps are totally independent of the private data

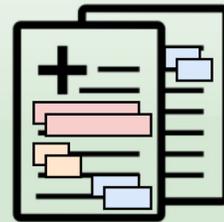
Model weights are reinitialized before training on the public data

Potential attackers could only use the labels produced on non-sensitive data by the private model, which we argue is insufficient to retrieve personal information from the sensitive data

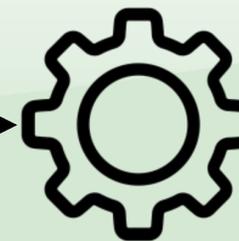
Shareable



Labeling



Training



Share



Unlabeled public data

Silver-labeled public data

Student Model

End users

Privacy-preserving mimic models

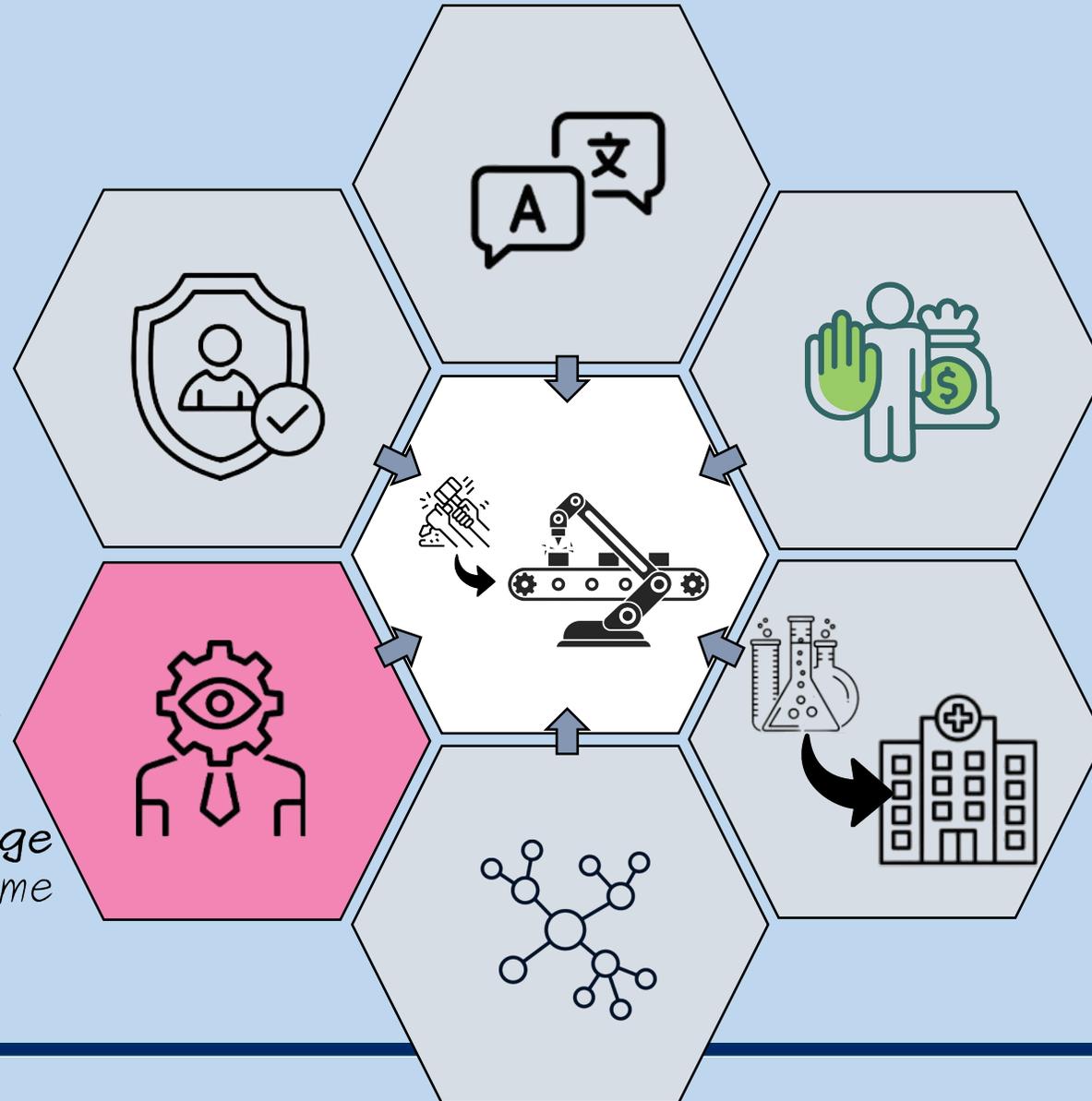
- Task : medical concepts in French
- Private data : MERLOT (L. Campillos et al., 2018)
- Public data : DEFT (R. Cardon et al., 2020)

	Precision	Recall
<i>baseline</i> → Dictionary-based model (UMLS)	0.25	0.17
Student (public) model	0.60	0.74
Teacher (private) model	0.86	0.86

skyline ↗

Can be used for preannotation or bootstrapping a noise-robust model

Example 3: Weak / distant supervision

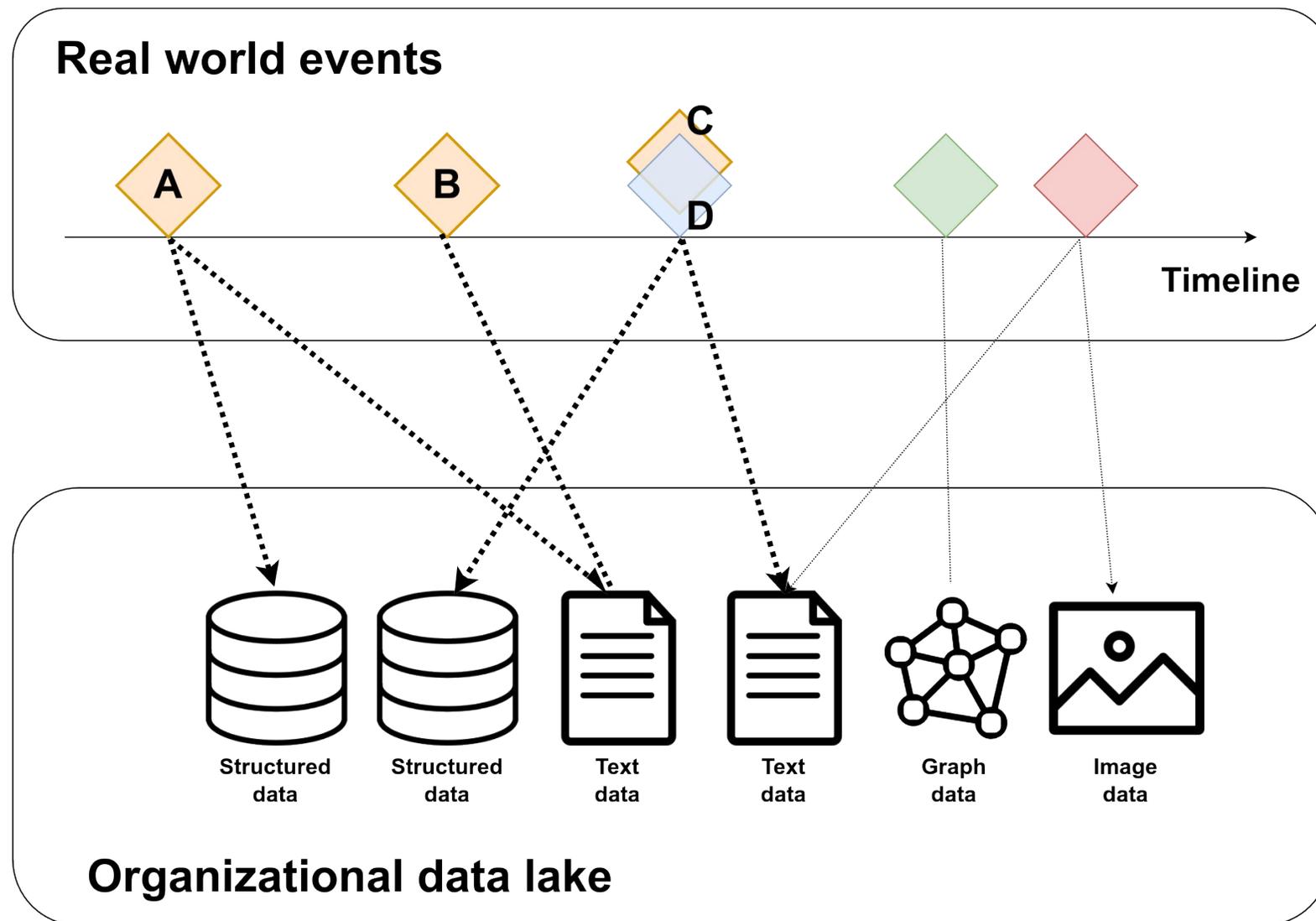


Saving human time

- Using weak or distant supervision*
- Using human knowledge without using human time*

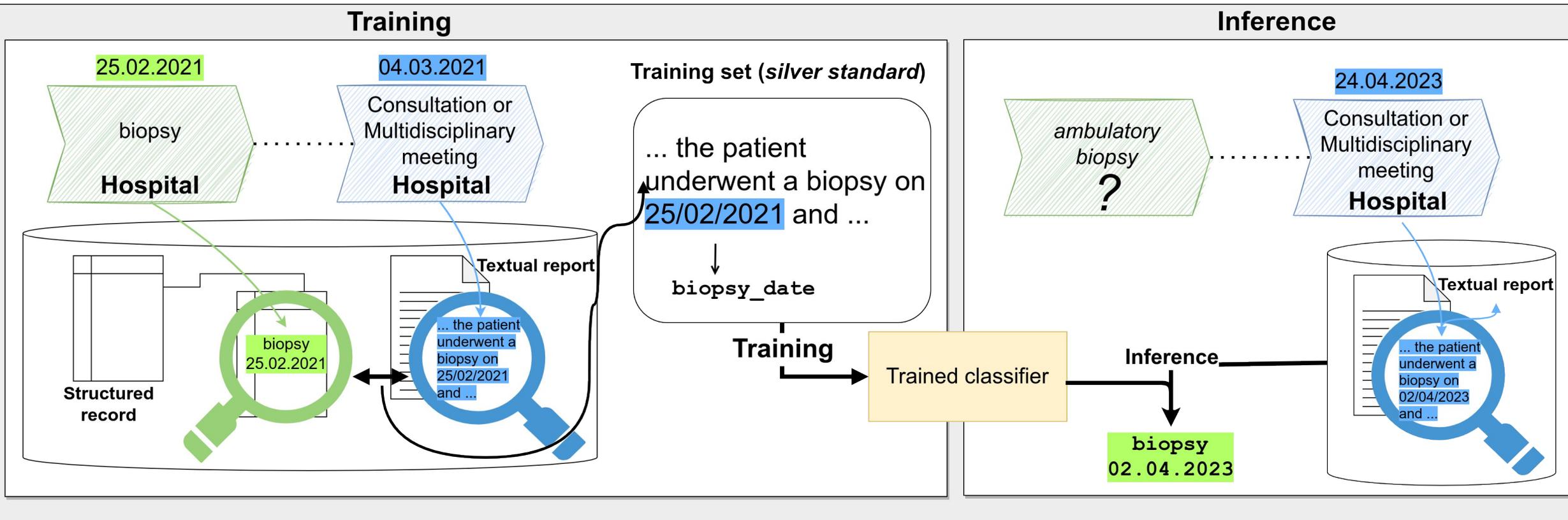
Weak / distant supervision

Information redundancy in an
clinical data lake /
data warehouse



Weak / distant supervision

How to get distant labels



Weak / distant supervision

Dealing with noisy labels introduced by the distant supervision

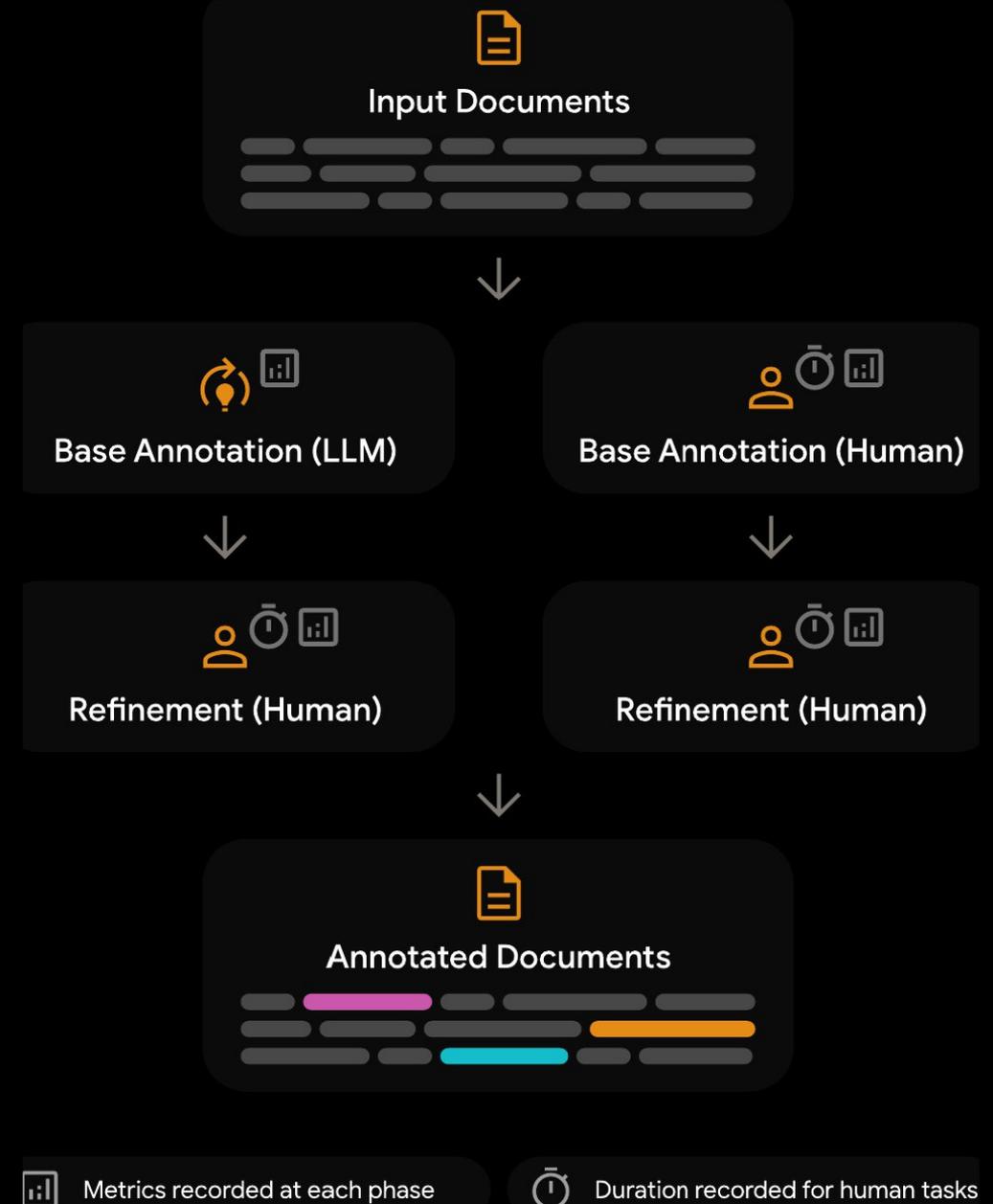
		Precision	Recall
<i>baseline</i> →	Regular cross-entropy	0.38	0.90
<i>three different ways to deal with noisy labels</i> →	Robust Loss Function (NCE-RCE)	0.61	0.90
	Sample Selection (O2U - CE)	0.56	0.93
	Label Refurbishment (LRT)	0.64	0.90
<i>skyline (supervised)</i> →	Expert label dataset (supervised learning)	0.69	0.93

LLM-based supervision

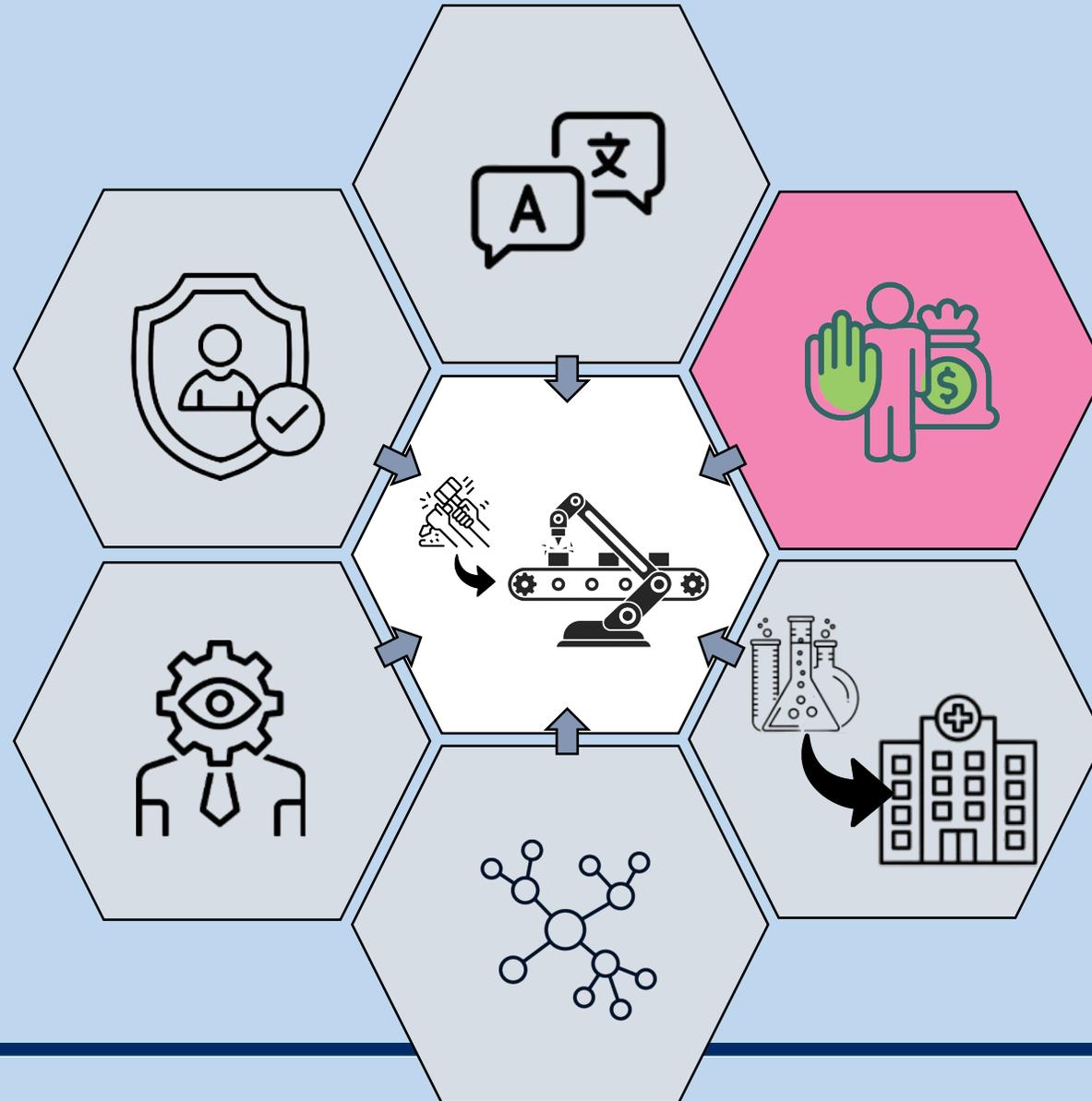
Example:

LLMs Accelerate Annotation for Medical Information Extraction.

Goel et al., *Google Research*



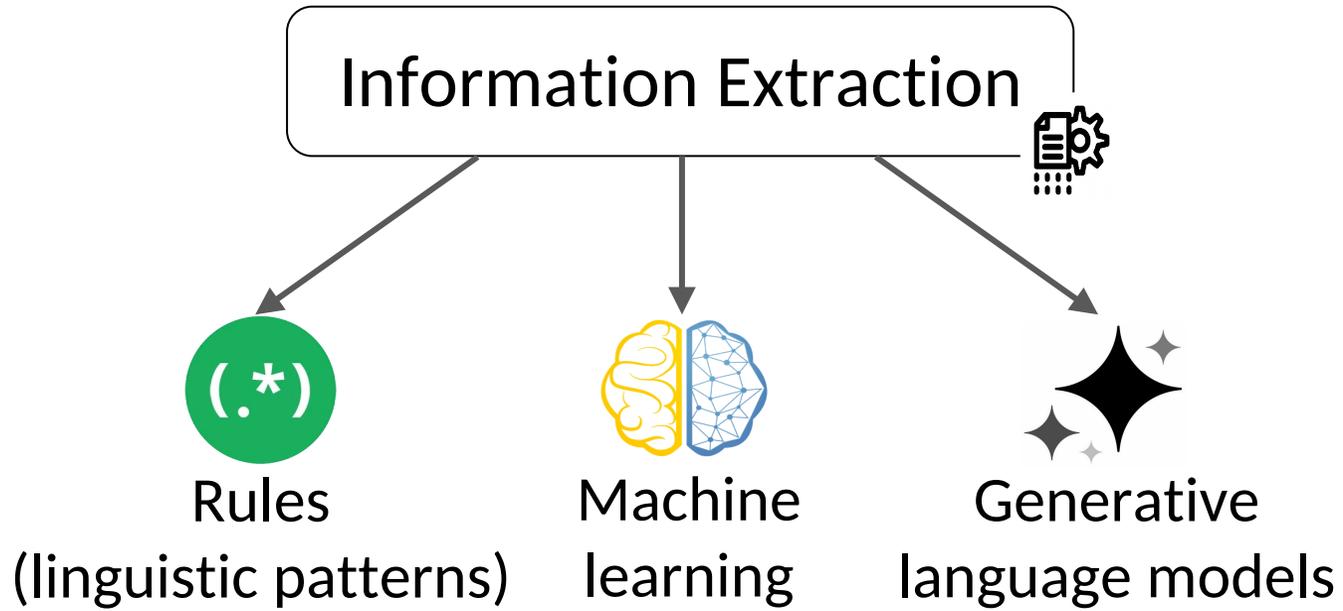
Example 4: Choose the lightest methods



Minimizing resources

Bring out the big guns only when necessary

Choose the lightest methods (REST)



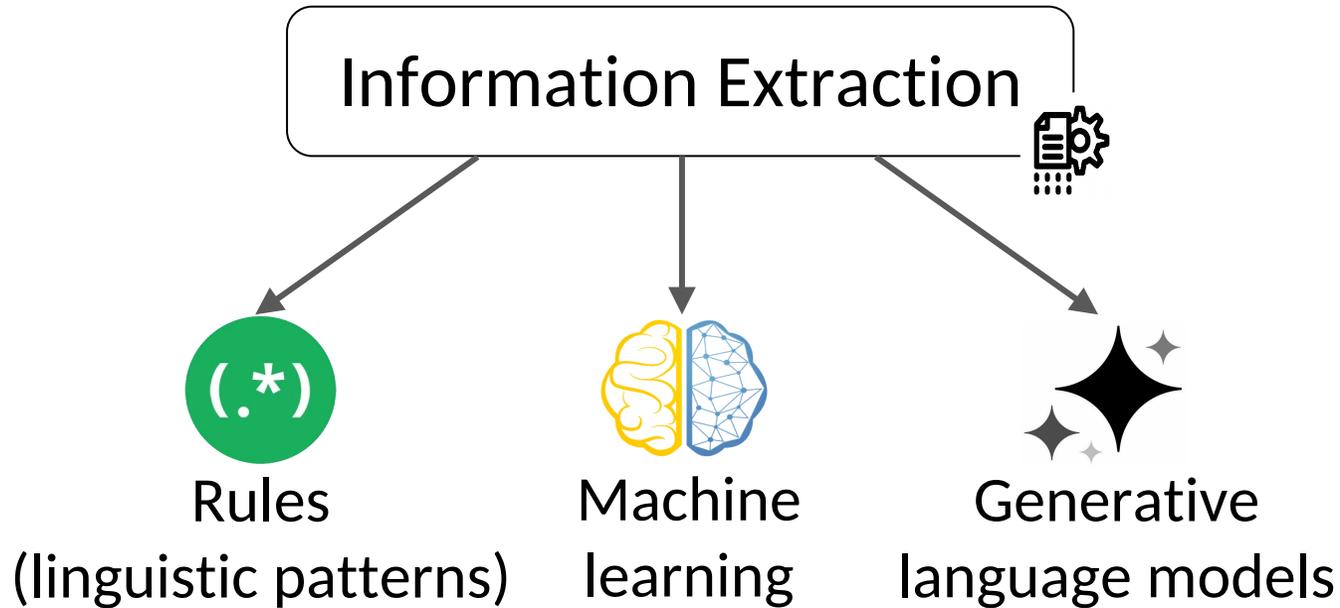
Sustainability

Transferability

Interpretability

*Human cost
(expert, IT)*

Choose the lightest methods (REST)



REST

I - Entity metrics results

id	entity	count	percentage
1	person	10	100%

II - Categories metrics results

category	count	percentage
category_1	10	100%

III - Summary table of found pattern

id	category	count	percentage
1	category_1	10	100%

II - Category creation

- Category: adreconome
- Category: grandes cellules
- Category: neuroendocrine
- Category: tumour metastatique

Categories distribution for the entity - Message Summary

A tool to visualize and predict feasibility of a rule-based system for an information extraction task

<https://github.com/GuiGuiBazin/REST-interface>

Choose the lightest methods (REST)

Concordancer

Word :

key	words before	word	words after	entity
0	...nt médical intéressant, a consulté en avril 2010 pour palpation d'une	tumeur	aux deux seins depuis plusieurs jours. exploration physi...	Signes_physiques
1	...ncé sans signe de maladie jusqu'à ce qu'en janvier 2016, une nouvelle	tumeur	soit palpée dans le sein gauche. une nouvelle biopsie a été...	Not annotated
2	...en raison de l'absence de mutations spécifiques dans l'échantillon de	tumeur	. l'approbation a été demandée pour une nouvelle ligne de tr...	Not annotated
3	...-exon 21, mutation t790m-exon 20. le cas a été présenté au comité des	tumeur	s thoraciques et, bien que la résection du nodule soit techn...	Not annotated
4	...ômes b intercurrents. aucune auto-palpation de lymphadénopathie ou de	tumeur	s à quelque niveau que ce soit. les parents du patient n'ont...	Not annotated
5	...iveau du pôle inférieur du testicule droit. » en cas de suspicion de	tumeur	testiculaire avancée, une orchidectomie radicale droite a é...	Not annotated
6	... pathologiquement, la lésion cicatricielle était compatible avec une	tumeur	à cellules germinales testiculaires brûlées. diagnostic ...	Histologie_tumorale
7	...ne	tumeur	à cellules germinales testiculaires brûlées. diagnostic	Topographie_du_primitif

Choose the lightest methods (REST)

Intra-entity linguistic homogeneity

entity	homogeneity
histologie_tumorale	0.79
traitement_specifique_du_cancer	0.79
signes_physiques	0.29
evolutivite_en_lien_avec_le_cancer	0.01
reponse_a_la_chimiotherapie	0.82
stade_metastatique_avec_localisations	0.6
statut_tabagique	0.62
atcd_geriatriques_et_medicaux_significatifs_pour_la_prise_en_charge	0.11
stade_oms_ecog_karnofsky	0.91
biomarqueurs_therapeutiques	0.67
topographie_du_primitif	0.57
symptomes	0.4

Choose the lightest methods (REST)

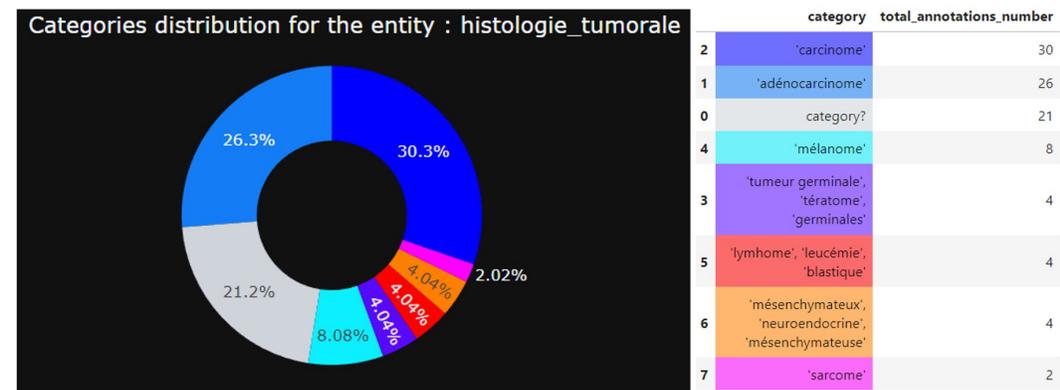
Precision and coverage of existing annotations

I - Entity metrics results

key	entity	homogeneity	TP	FP	FN	precision	precision_confidε	recall	recall_confidε
5	stade_metastatique_avec_localisations	0.6	45	26	38	0.63	[0.49, 0.76]	0.54	[0.42, 0.66]

II - Categories metrics results

category	raw highlights				corrected highlights			
	TP	FP	FN	precision	TP(corr)	FP(corr)	FN(corr)	precision
'métastase', 'métastatique'	22	11	0	0.67	22	11	0	0.67
'stade iv', 'stade 4', 'stade IV'	9	2	0	0.82	9	2	0	0.82
'm0', 'm1'	3	1	0	0.75	3	1	0	0.75
'implants péritonéaux', 'implant péritonéal'	3	0	0	1	3	0	0	1
'carcinose'	1	8	0	0.11	1	8	0	0.11
't.n.m.', 'pt.n.m.', 'ct.n.m.', 'pt..n.m.'	7	4	0	0.64	7	4	0	0.64



Choose the lightest methods

More generally, try to

- Choose the lightest methods
- Choose the lightest models
- Choose scalable methods
- Estimate your energy consumption / carbon impact

Example 5: Evaluation in clinical settings



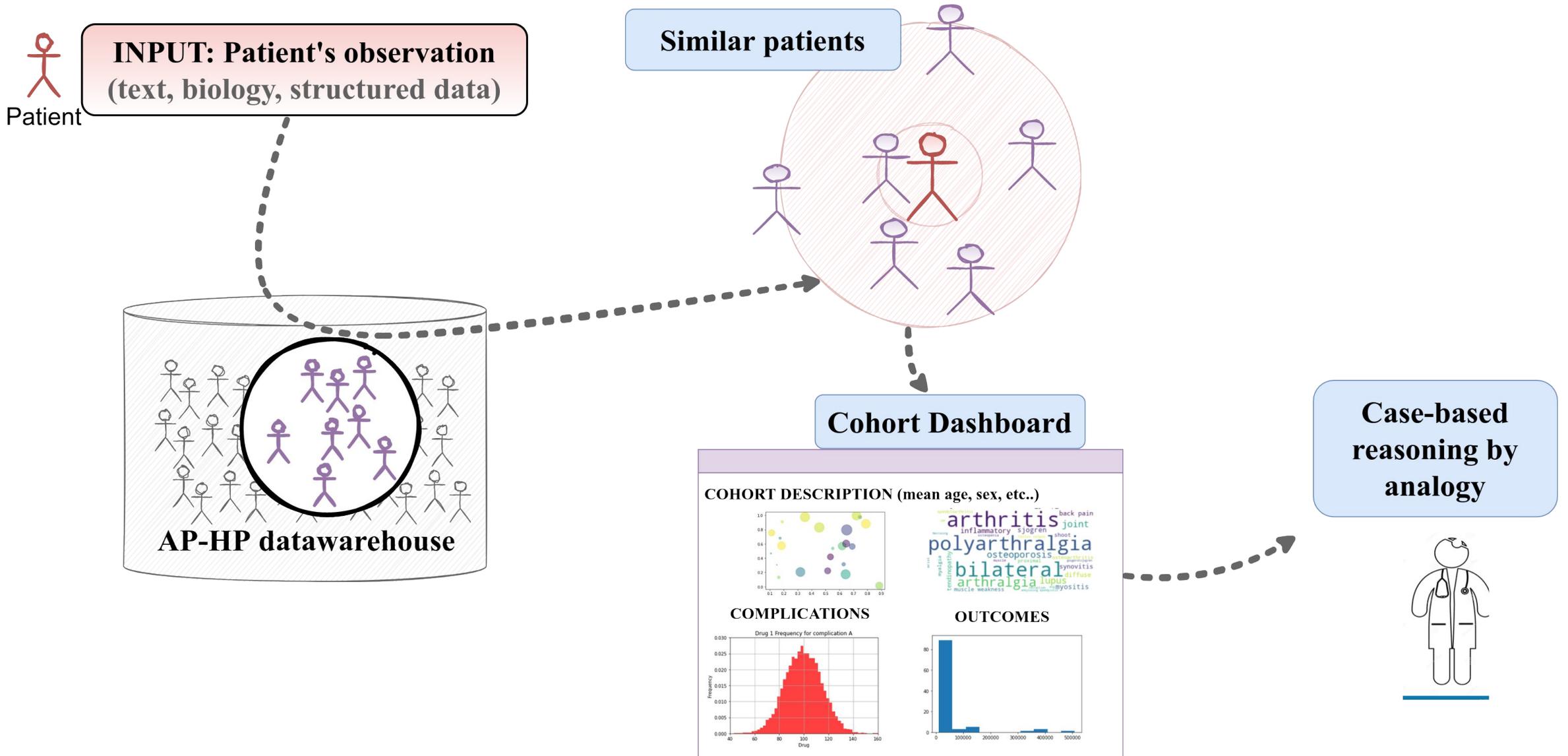
with



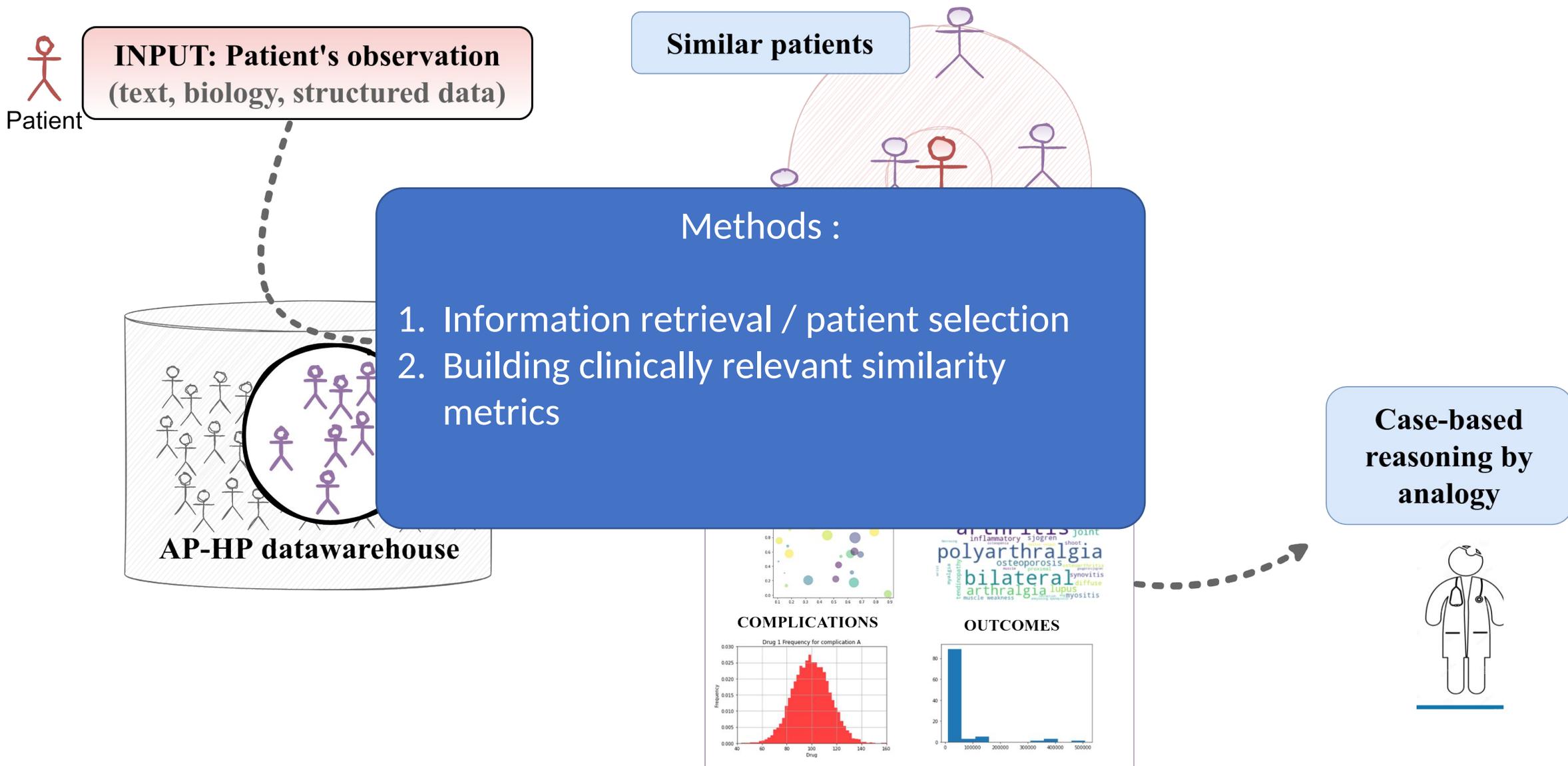
Evaluation in clinical settings

Interoperability with hospital information systems

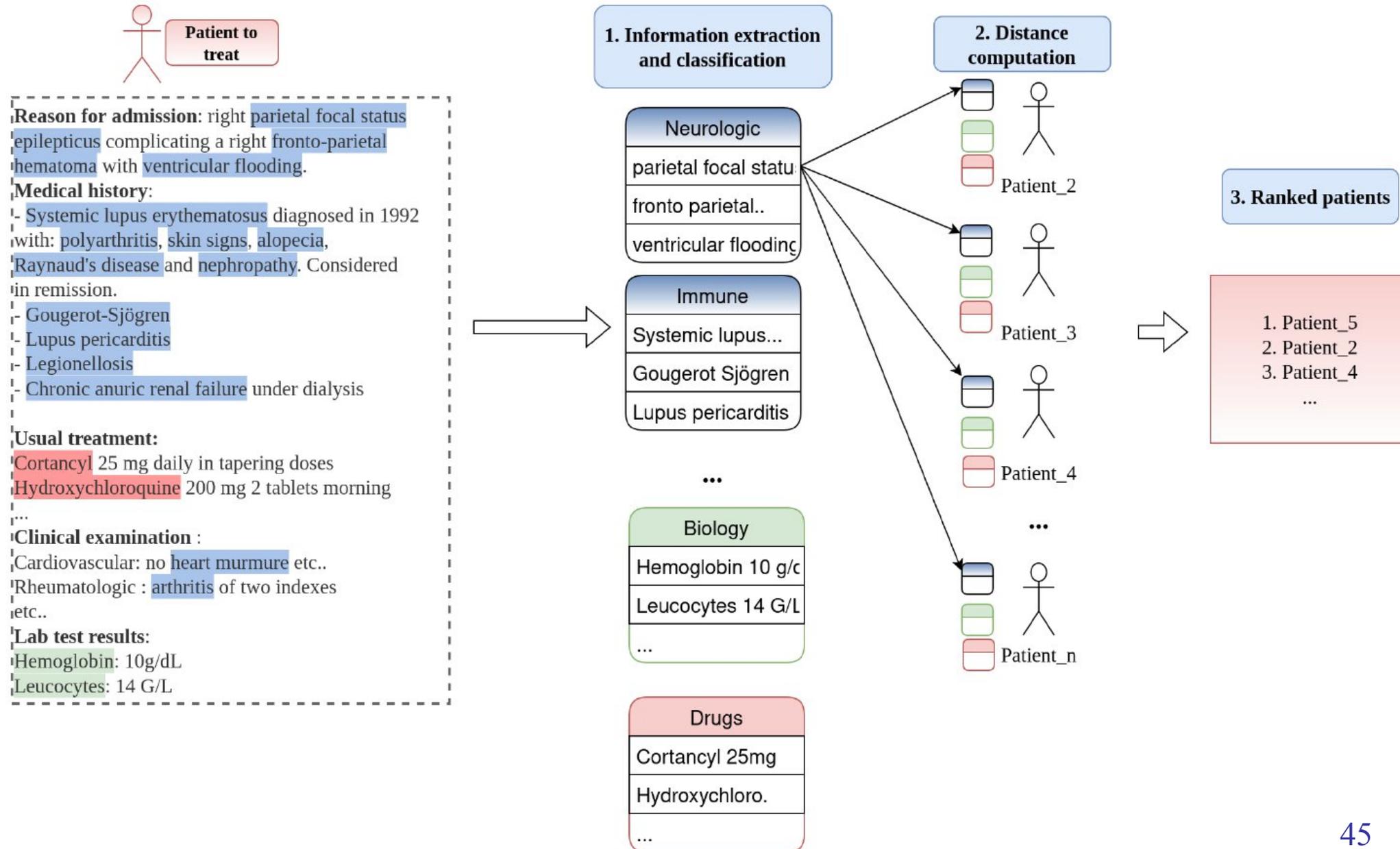
Use case 1, similar patient selection



Use case 1, similar patient selection



Use case 1, Building clinically relevant similarity metrics



Use case 1, Building clinically relevant similarity metrics

Nephritis in Systemic Lupus Erythematosus

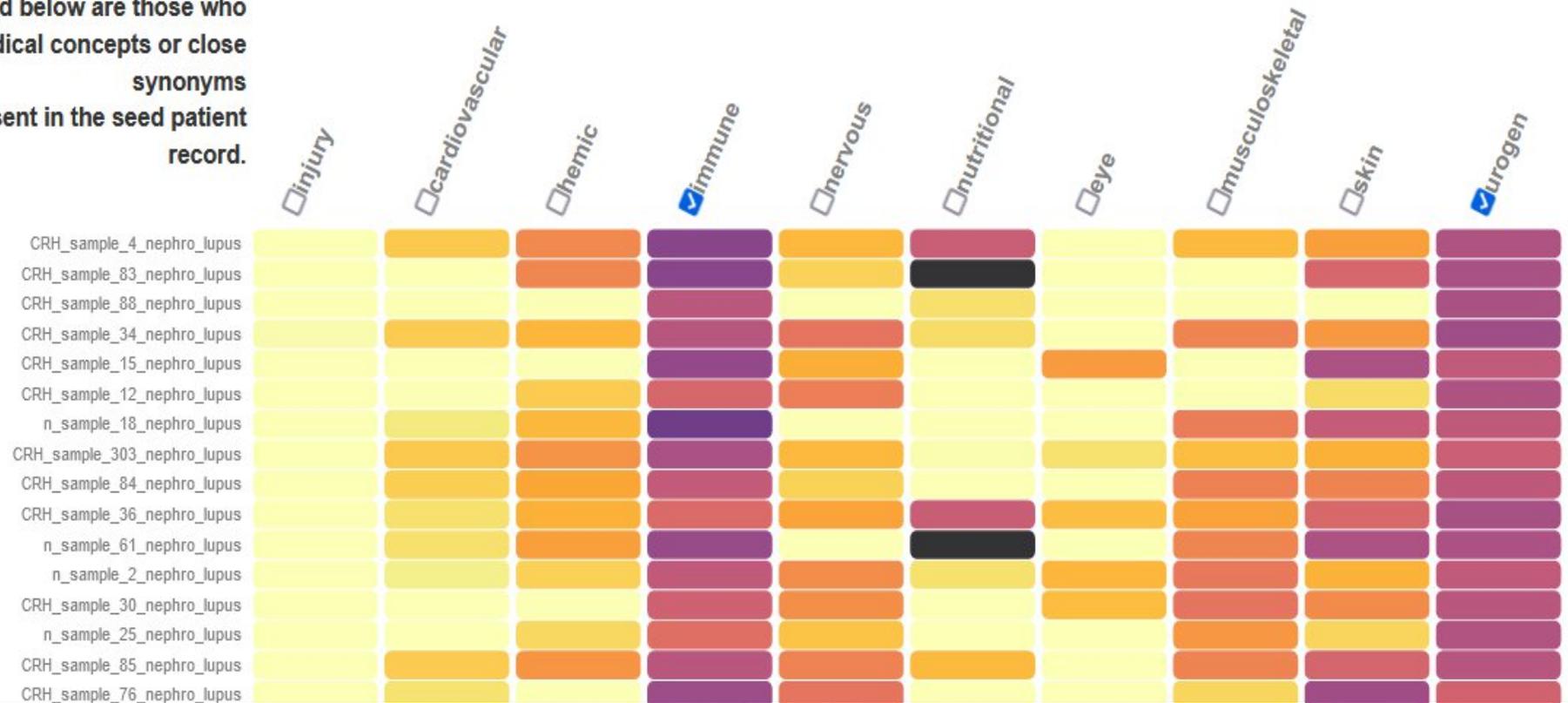
The patient for this use case is **CRH_val_sample_64**. The most relevant labels for this use case are **urogen,immune**.

The full content of the patient records cannot be shown in this demo for privacy reasons. The most relevant terms for the labels checked below are extracted automatically with the method described in the companion article.

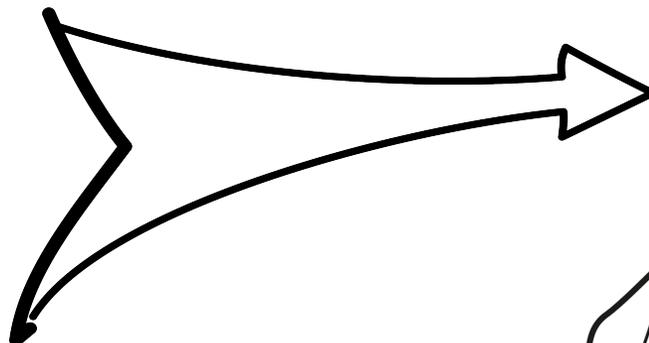
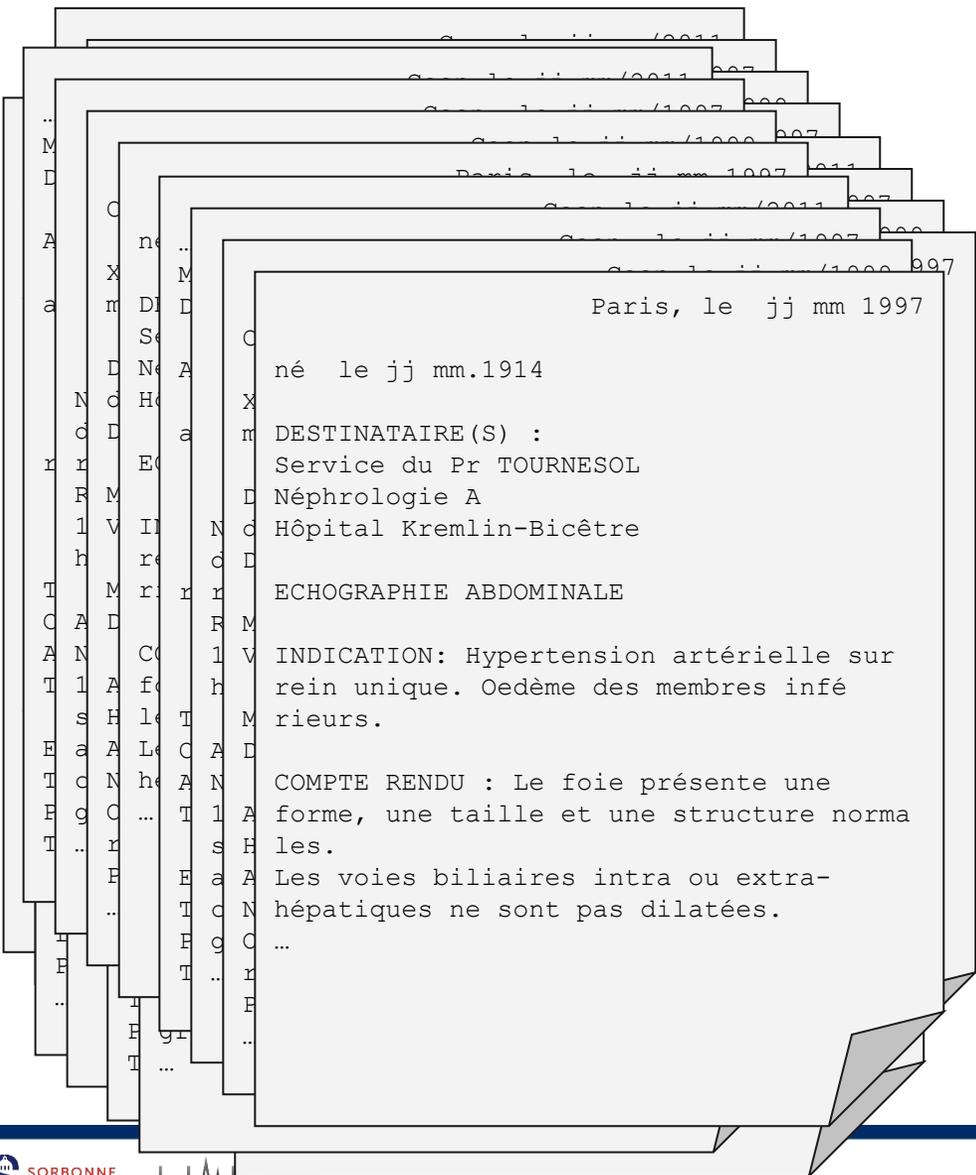
néphropathie lupique
goussière du loup
glomérulonéphrite lupique de classe v
syndrome néphrotique
lupus systémique
proléturie
glomérulonéphrite lupique
lupus systémique
glomér



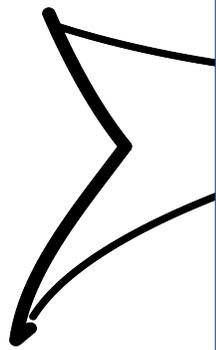
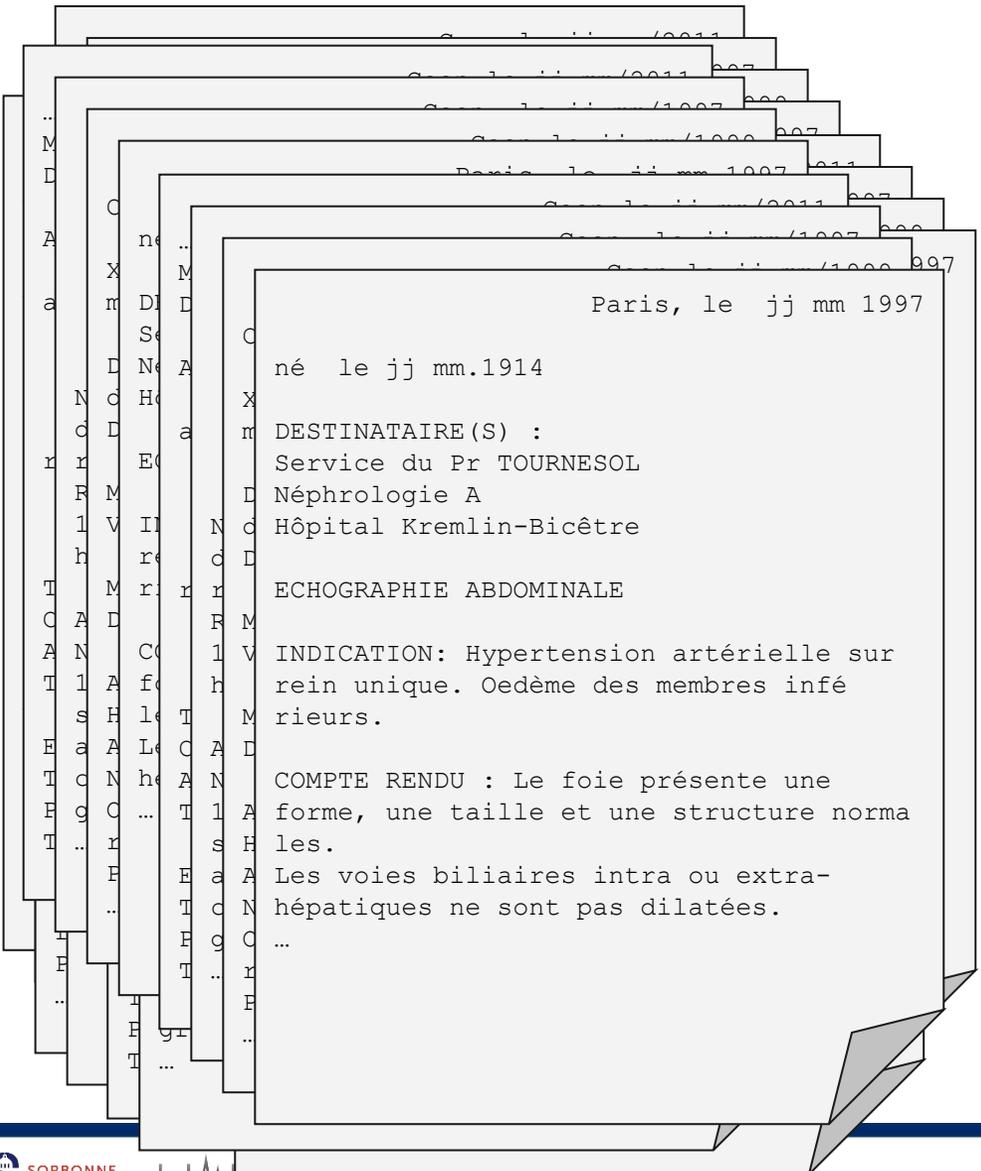
Most similar patients ranked below are those who share the same medical concepts or close synonyms
Shown labels are those present in the seed patient record.



Use case 2, ER patient summary



Use case 2, ER patient summary



Life Style

Patient divorcé
3 enfants : 19, 21 et 24 ans
Vit seul en appartement
Complètement autonome pour les actes de la vie quotidienne (juin 2024)
Performans Status 0 (juin 2024)
Travail comme conseiller en communication
Pas d'intoxication tabagique
Consommation éthylique : 1-2 verre de vin par jour
Dernier poids connu : 83 kg ;
Dernière taille connue : 175 cm

Surgical History

- **Résection recto-sigmoïdienne (janvier 2024)**
Pour adénocarcinome rectal infiltrant
Réalisée par le Pr Larzus en 1er opérateur
Par coelioscopie puis laparo-convertie
Anastomose colo-sus-anale mécanique
Complicquée d'une embolie pulmonaire post opératoire (cf antécédent médicaux)
- **Appendicectomie** (dans l'enfance)
- **Cure de hernie inguinale droite** (date inconnue)

Medical History

- **Hypertension artérielle essentielle**
- **Cardiopathie ischémique**
Diagnostiquée en mars 2018 sur un angor d'effort
Angioplastie et un stent actif de l'IVA proximale en mars 2018
Dernière ETT (mars 2024) : FEVG 60%, pas d'anomalie de la cinétique segmentaire [...]
- **Diabète de type 2**
Diagnostiqué en septembre 2015 et suivi en ville
Insulinodépendant depuis 2020
Dernière HbA1c (mai 2024) : 6.8%

Drug Allergies

- **ALLERGIE A L'AMOXICILLINE :**
Réaction anaphylactique en 2015 avec œdème de Quincke

Use case 2, ER patient summary

Methods : “Retrieval-Augmented Generation (RAG)”

1. Retrieve passages from the documents concerning each target section
2. Use a Large Language Model to generate a summary based on these passages

Life Style

Patient divorcé
3 enfants : 19, 21 et 24 ans
Vit seul en appartement
Complètement autonome pour les actes de la

Medical History

- **Hypertension artérielle essentielle**
- **Cardiopathie ischémique**
Diagnostiquée en mars 2018 sur un angor d'effort
Angioplastie et un stent actif de l'IVA proximale en mars 2018
Dernière ETT (mars 2024) : FEVG 60%, pas d'anomalie de la cinétique segmentaire [...]
- **Diabète de type 2**
Diagnostiqué en septembre 2015 et suivi en ville
Insulinodépendant depuis 2020
Dernière HbA1c (mai 2024) : 6.8%

Drug Allergies

- **ALLERGIE A L'AMOXICILLINE :**
Réaction anaphylactique en 2015 avec œdème de Quincke

Anastomose colo-sus-anales mécanique
Complicquée d'une embolie pulmonaire post opératoire (cf antécédent médicaux)

- **Appendicectomie** (dans l'enfance)
- **Cure de hernie inguinale droite** (date inconnue)

né le jj mm.1914

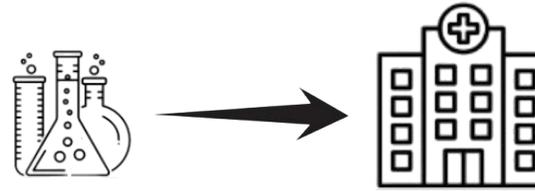
DESTINATAIRE(S) :
Service du Pr TOURNESON
Néphrologie A
Hôpital Kremlin-Bicêtre

ECHOGRAPHIE ABDOMINALE
INDICATION: Hypertension
rein unique. Oedème des
rieurs.

COMPTE RENDU : Le foie présente une
forme, une taille et une structure norma
les.

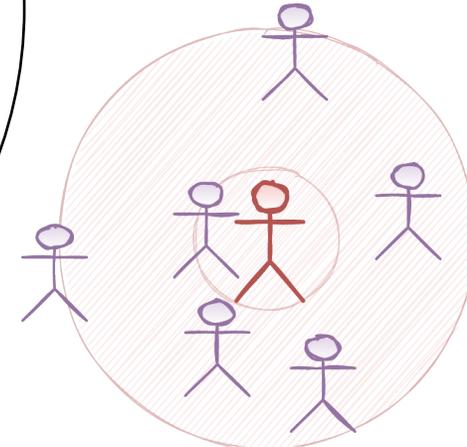
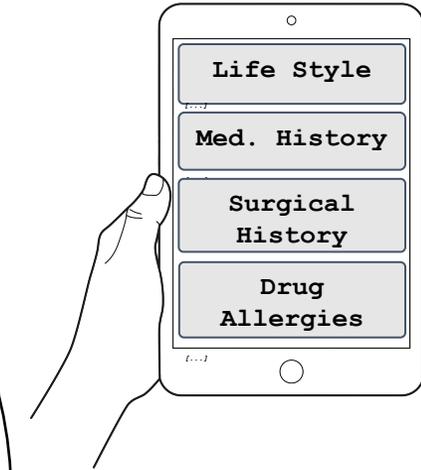
Les voies biliaires intra ou extra-
hépatiques ne sont pas dilatées.

Evaluation in clinical settings



- Is it usable?
- Is it useful?
- Is it so useful that patients are given better care?
- Is it never harmful?
- Is the result a “medical device”?
- Who is responsible for missed information?
- ...

- Interoperability issues (plug the system into the information system)
- Privacy issues (keep data and computing power inside the hospital)
- Computing resources issues (big data is... big)



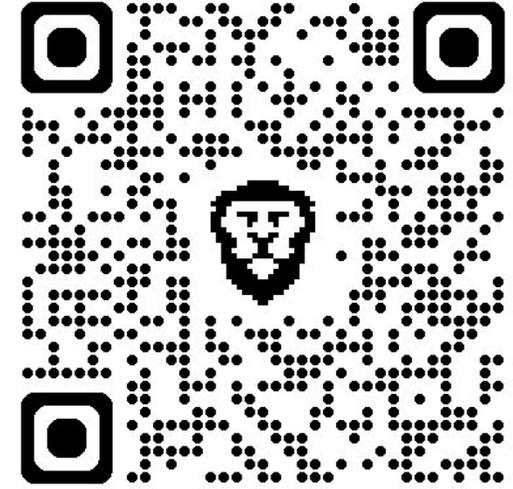
Example 6: edsnlp *et al.*, the AP-HP “EDS” software suite



ASSISTANCE
PUBLIQUE  HÔPITAUX
DE PARIS
Entrepôt de Données de Santé

The AP-HP “EDS” software suite

- Named entity recognition
- Normalization to terminologies (concept unique identifiers)
- Patient classification



EDS-NLP

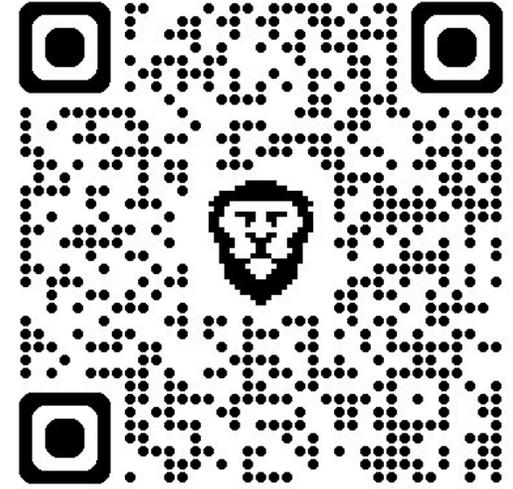
<https://github.com/aphp/edsnlp>

EDS-NLP is a collaborative NLP framework that aims primarily at extracting information from French clinical notes. At its core, it is a collection of components or pipes, either rule-based functions or deep learning modules. These components are organized into a novel efficient and modular pipeline system, built for hybrid and multitask models. We use [spaCy](#) to represent documents and their annotations, and [Pytorch](#) as a deep-learning backend for trainable components.

EDS-NLP is versatile and can be used on any textual document. The rule-based components are fully compatible with spaCy's components, and vice versa. This library is a product of collaborative effort, and we encourage further contributions to enhance its capabilities.

Check out our interactive [demo](#) !

The AP-HP “EDS” software suite



EDS-PDF

<https://github.com/aphp/edspdf>

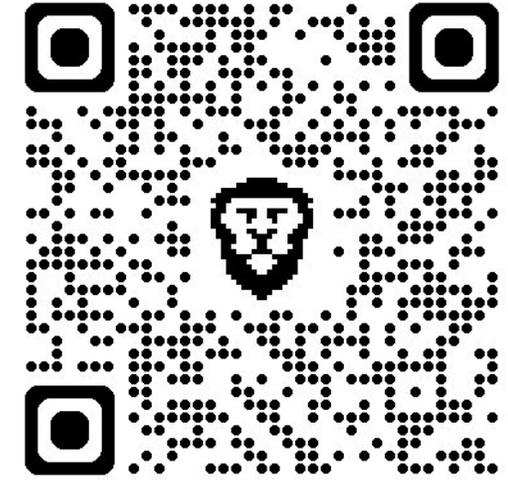
EDS-PDF provides a modular framework to extract text information from PDF documents.

You can use it out-of-the-box, or extend it to fit your specific use case. We provide a pipeline system and various utilities for visualizing and processing PDFs, as well as multiple components to build complex models: complex models:

-  [Extractors](#) to parse PDFs (based on [pdfminer](#), [mupdf](#) or [poppler](#))
-  [Classifiers](#) to perform text box classification, in order to segment PDFs
-  [Aggregators](#) to produce an aggregated output from the detected text boxes
-  Trainable layers to incorporate machine learning in your pipeline (e.g., [embedding](#) building blocks or a [trainable classifier](#))

Visit the  [documentation](#) for more information!

The AP-HP “EDS” software suite



EDS-Pseudo

<https://github.com/apHP/eds-pseudo>

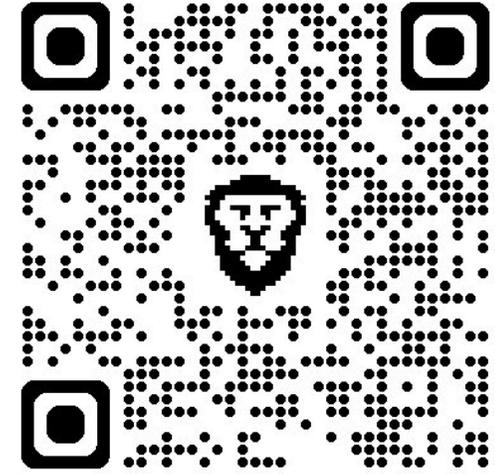
The EDS-Pseudo project aims at detecting identifying entities in clinical documents, and was primarily tested on clinical reports at AP-HP's Clinical Data Warehouse (EDS).

The model is built on top of [edsnlp](#), and consists in a hybrid model (rule-based + deep learning) for which we provide rules ([eds-pseudo/pipes](#)) and a training recipe [train.py](#) .

We also provide some fictitious templates ([templates.txt](#)) and a script to generate a synthetic dataset [generate_dataset.py](#) .

The AP-HP “EDS” software suite

ASSISTANCE PUBLIQUE  HÔPITAUX DE PARIS
Entrepôt de Données de Santé



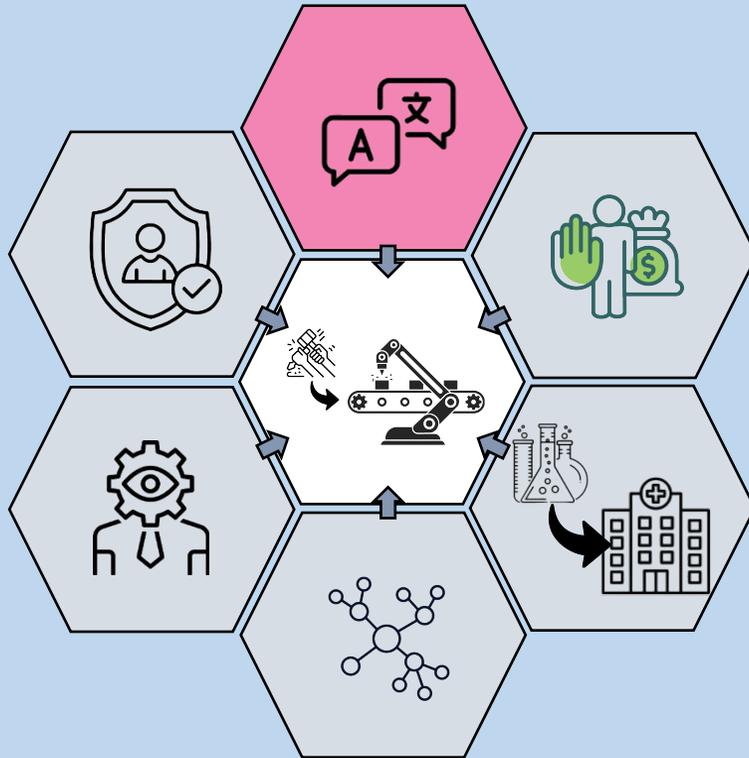
tests **passing** docs **passing** pypi **v0.1.8** python **>= 3.7.1 | < 3.8** code style **black** coverage **94%** DOI **10.5281/zenodo.11636968**

eds-scikit is a tool to assist data scientists working on the AP-HP's Clinical Data Warehouse. It is specifically targeted for OMOP-standardized data. Its main goals are to:

- Ease access and analysis of data
- Allow a better transfer of knowledge between projects
- Improve research reproducibility

<https://github.com/aphp/eds-scikit>

And also...



with



Institut Pierre Louis d'Épidémiologie et de Santé Publique
Pierre Louis Institute of Epidemiology and Public Health

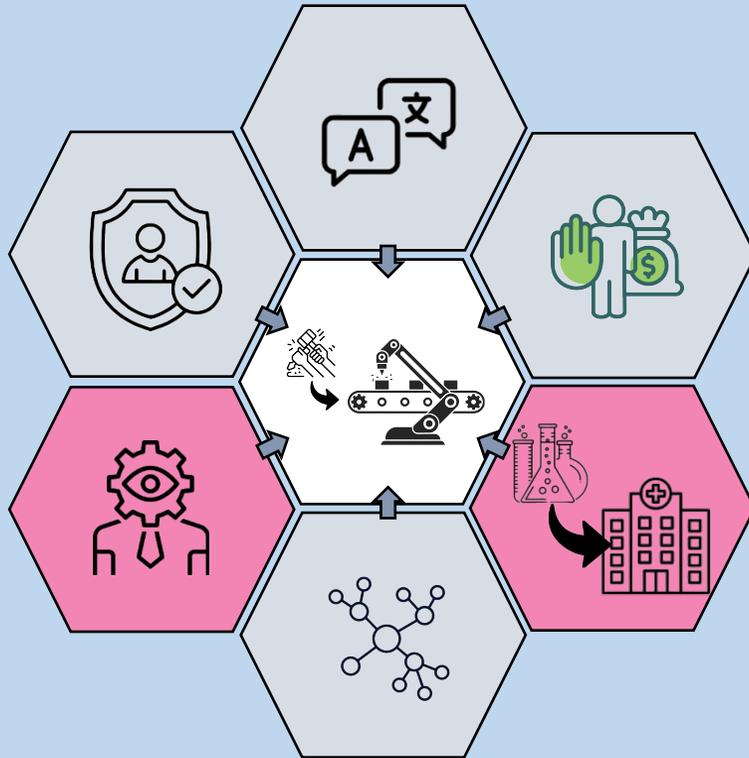
Translation

Is it a good idea to translate into English to benefit from the richer resources of that language?

(spoiler: it's complicated)

Gérardin et al. **Impact of translation on biomedical information extraction: an experiment on real-life clinical notes.** JMIR Medical Informatics, 2024

And also...



with



UNIVERSITÉ
DE GENÈVE

Generative LLMs

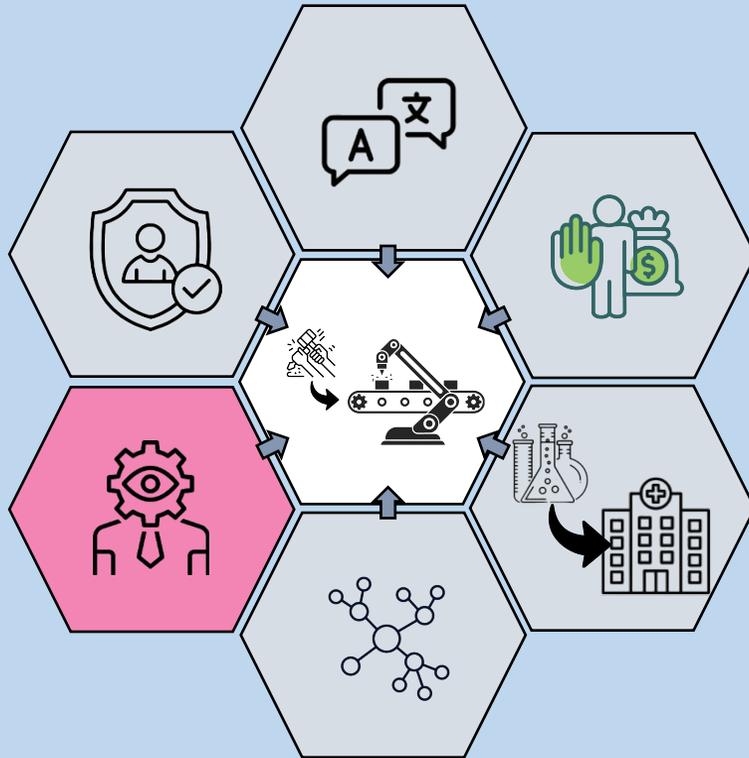
Performance evaluation of generative LLMs in different configurations

Naguib et al. **Few-shot clinical entity recognition in English, French and Spanish: masked language models outperform generative model prompting.** EMNLP 2024

Zaghir et al. **Prompt Engineering Paradigms for Medical Applications: Scoping Review.** JMIR, 2024

Tai et al. **Clinical trial cohort selection using Large Language Models on n2c2 Challenges.** Submitted 2025

And also...



with



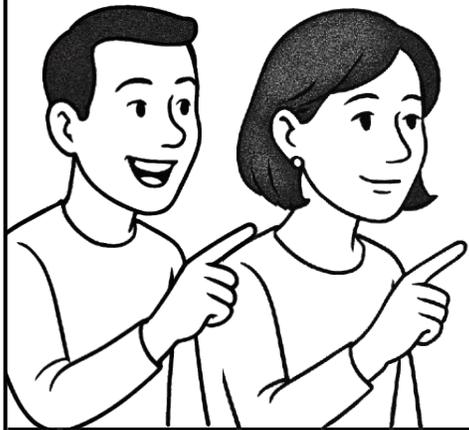
Active learning & Few-shot

Can we

- optimize the choice of annotated documents
- pre-annotate
- learn from few samples to save human time?

Naguib et al. Few-shot clinical entity recognition in English, French and Spanish: masked language models outperform generative model prompting. EMNLP 2024

Naguib et al. Stratégies d'apprentissage actif pour la reconnaissance d'entités nommées en français. TALN 2023



Thank you!

Questions ?



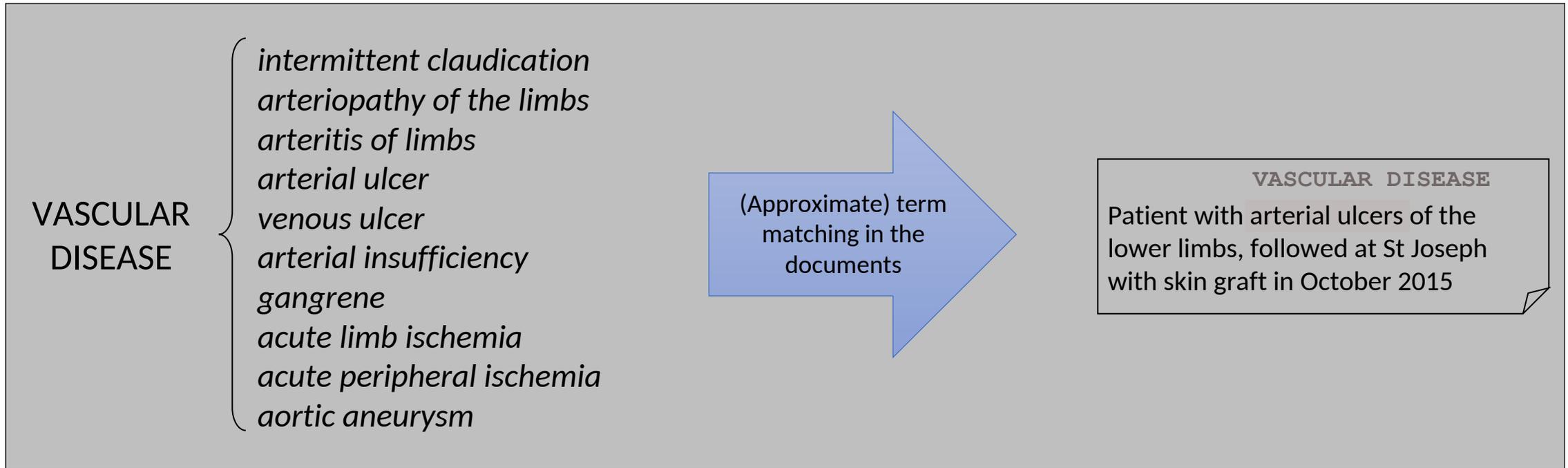
Backup slides...

Methods

- 1. Rule-based systems**
2. Supervised learning systems
3. Generative, large language models
4. Retrieval-Augmented Generation

Rule-based systems

1. Terminological approach



Rule-based systems

1. Terminological approach
2. **Additional rules**



Rule-based systems

1. Terminological approach
2. Additional rules
3. Trickier rules

Emmanuelle Kempf et al., 2023

secondary, progressive,
fractured or suspected
ostecondensation

```
ost[ée]ocondensa. {1,20} (suspect|secondaire|[ée]vulsive)| (l[ée]sion|anomalie|image).  
{1,20}os.  
{1,30} (suspect|secondaire|[ée]vulsive)| os. {1,30} (l[ée]sion|anomalie|image). {1,20}  
(suspect|secondaire|[ée]vulsive)| (l[ée]sion|anomalie|image).  
{1,20};L[I,Y]tique| (l[ée]sion|anomalie|image). {1,20}condensant. {1,20}  
(suspect|secondaire|[ée]vulsive)| fracture. {1,30}  
(suspect|secondaire|[ée]vulsive)| ((l[l[ée]sion|anomalie|image|nodule). {1,80}  
(secondaire))| ((l[l[ée]sion|anomalie|image|nodule)s.{1,40}suspec?ts?).
```

post-operative
anatomopathological tumor
stage (pTNM)

```
([ycpP]{1,2}\s? (T([01234x]|is)[abcdx]?),\s {0,2} [ycp] {0,2}\s? (N[xo01234\+][abcdx]?)*\s?  
(M[o01]? [\+x]?))| ((T([01234x]|is)[abcdx]?),\s {0,2} [ycp] {0,2}\s? (N[xo01234\+][abcdx]?)  
\s?  
(M[o01]? [\+x]?))
```

Rule-based systems: Pros & Cons

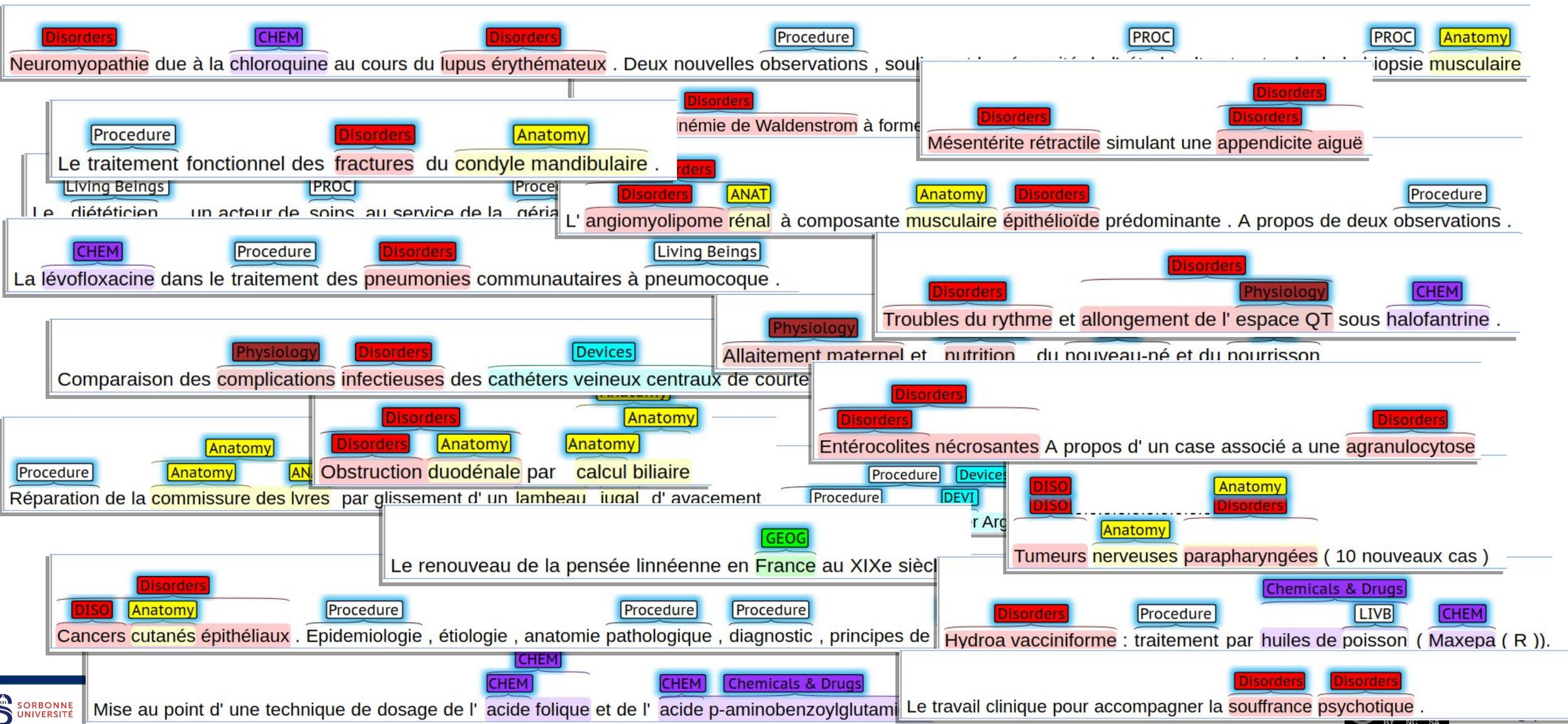
	Rules
General performance	Yellow
Ease of implementation	Green
Need for human expertise	Yellow
Explainability	Green
Material resources	Green
Energy consumption	Green
Ease of maintenance	Yellow
Generalization to a different problem/context	Red

Methods

1. Rule-based systems
- 2. Supervised learning systems**
3. Generative, large language models
4. Retrieval-Augmented Generation

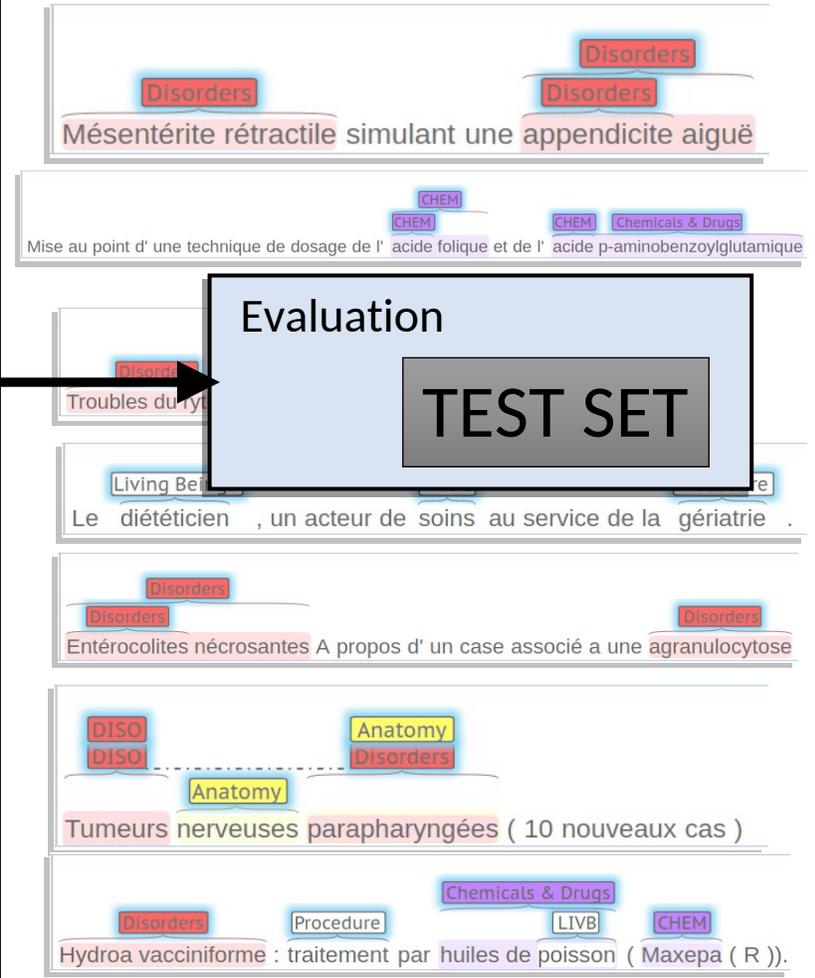
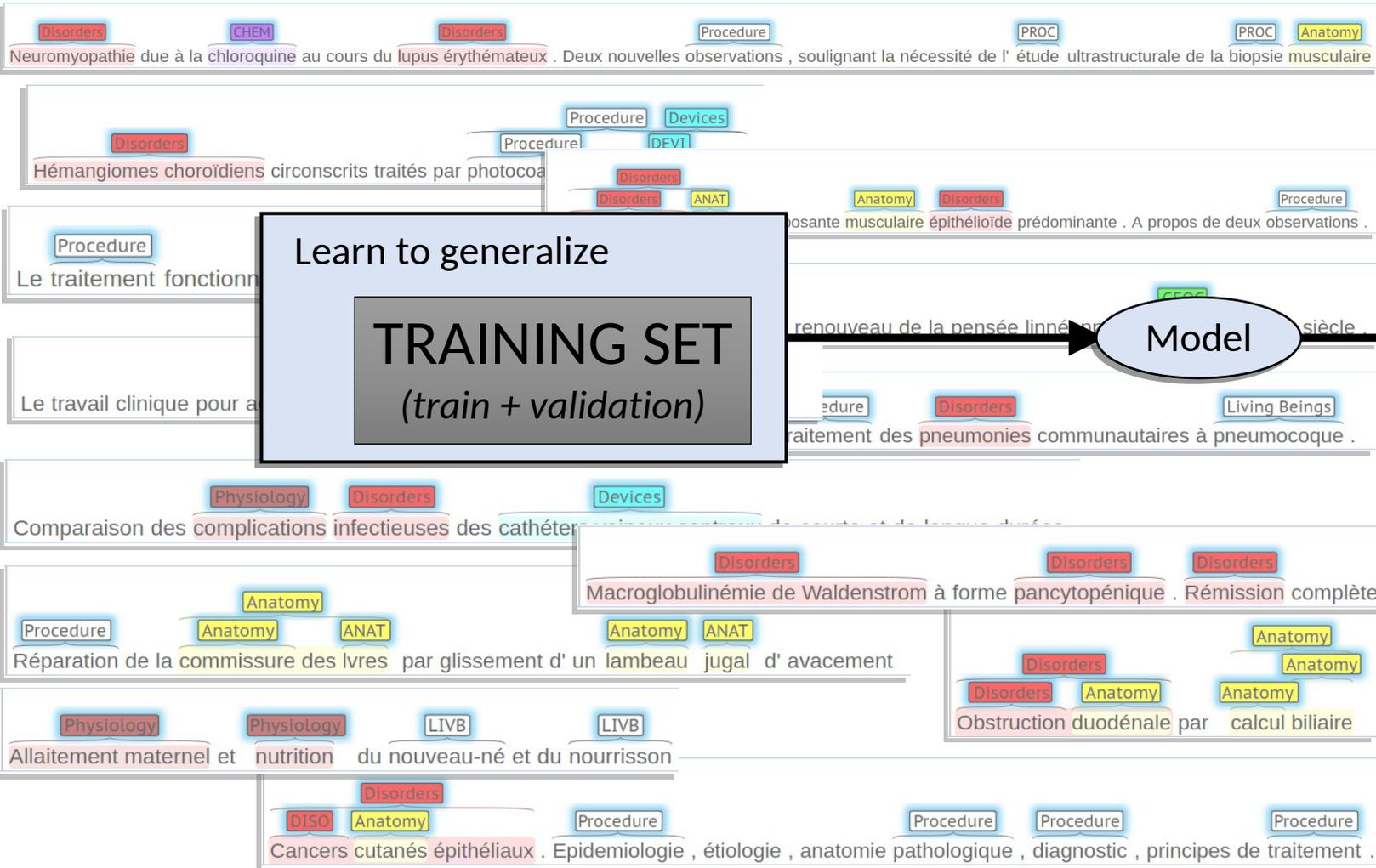
Supervised learning systems

Névél, A., Grouin, C., Leixa, J., Rosset, S., & Zweigenbaum, P. (2014).
The Quaero French Medical Corpus: A Ressource for Medical Entity Recognition and Normalization.



Supervised learning systems

Névéol, A., Grouin, C., Leixa, J., Rosset, S., & Zweigenbaum, P. (2014).
The Quaero French Medical Corpus: A Ressource for Medical Entity Recognition and Normalization.



Supervised learning systems: Pros & Cons

	Rules	Supervised learning
General performance	Yellow	Green
Ease of implementation	Green	Green
Need for human expertise	Yellow	Red
Explainability	Green	Yellow
Material resources	Green	Yellow
Energy consumption	Green	Yellow
Ease of maintenance	Yellow	Yellow
Generalization to a different problem/context	Red	Red

Methods

1. Rule-based systems
2. Supervised learning systems
- 3. Generative, large language models**
4. Retrieval-Augmented Generation

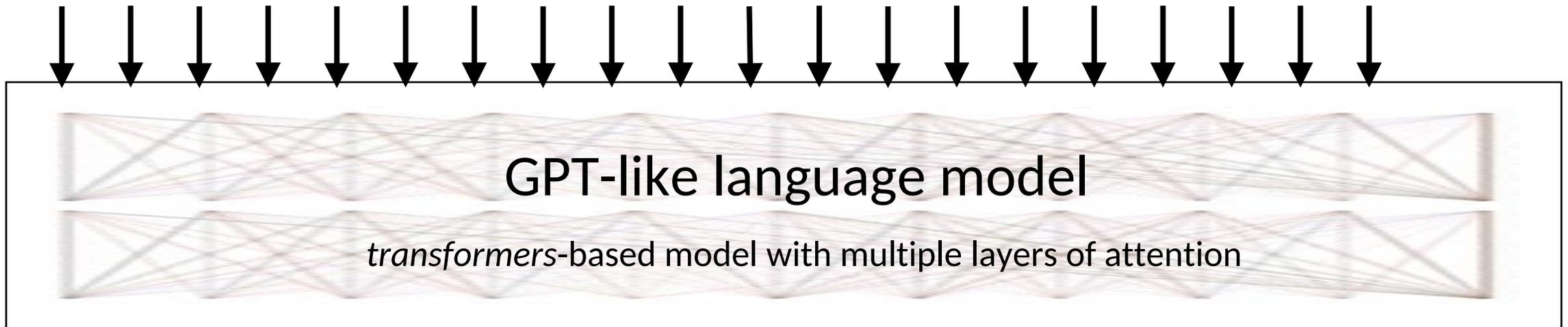
Large Language Models

- Generative large language models (LLMs) are trained to produce human-like text.
- They simulate human understanding by predicting and generating text based on the input they receive.
- They can be finetuned for different tasks, e.g rewriting a prompt in another styles.

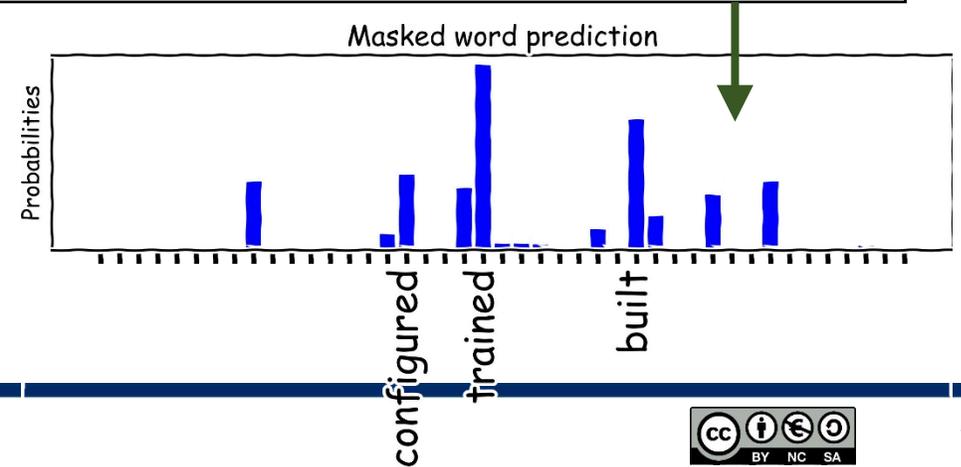
Auto-regressive language models (*decoder-style, e.g. GPT*)

User: What is a large language model?

Assistant: A large language model is a type of artificial intelligence (AI) model designed to understand and generate human-like language. These models are



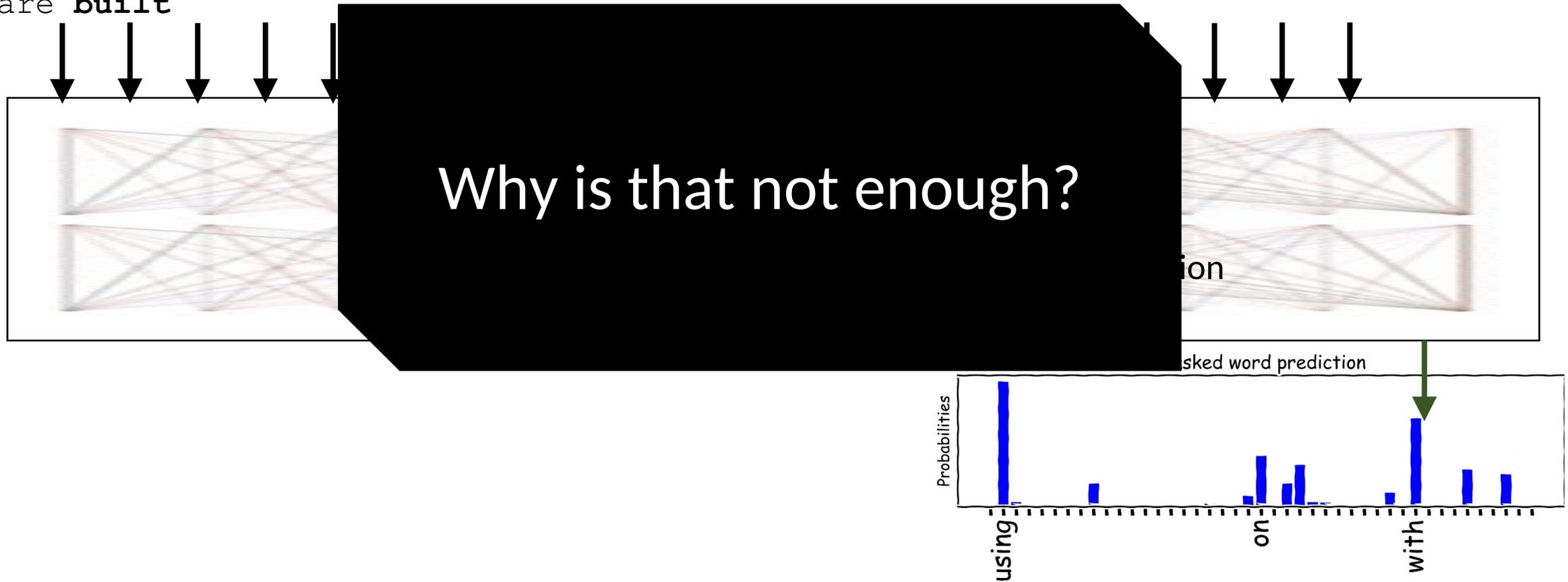
Next tokens are randomly selected from the predicted distribution, introducing variability in the generated output



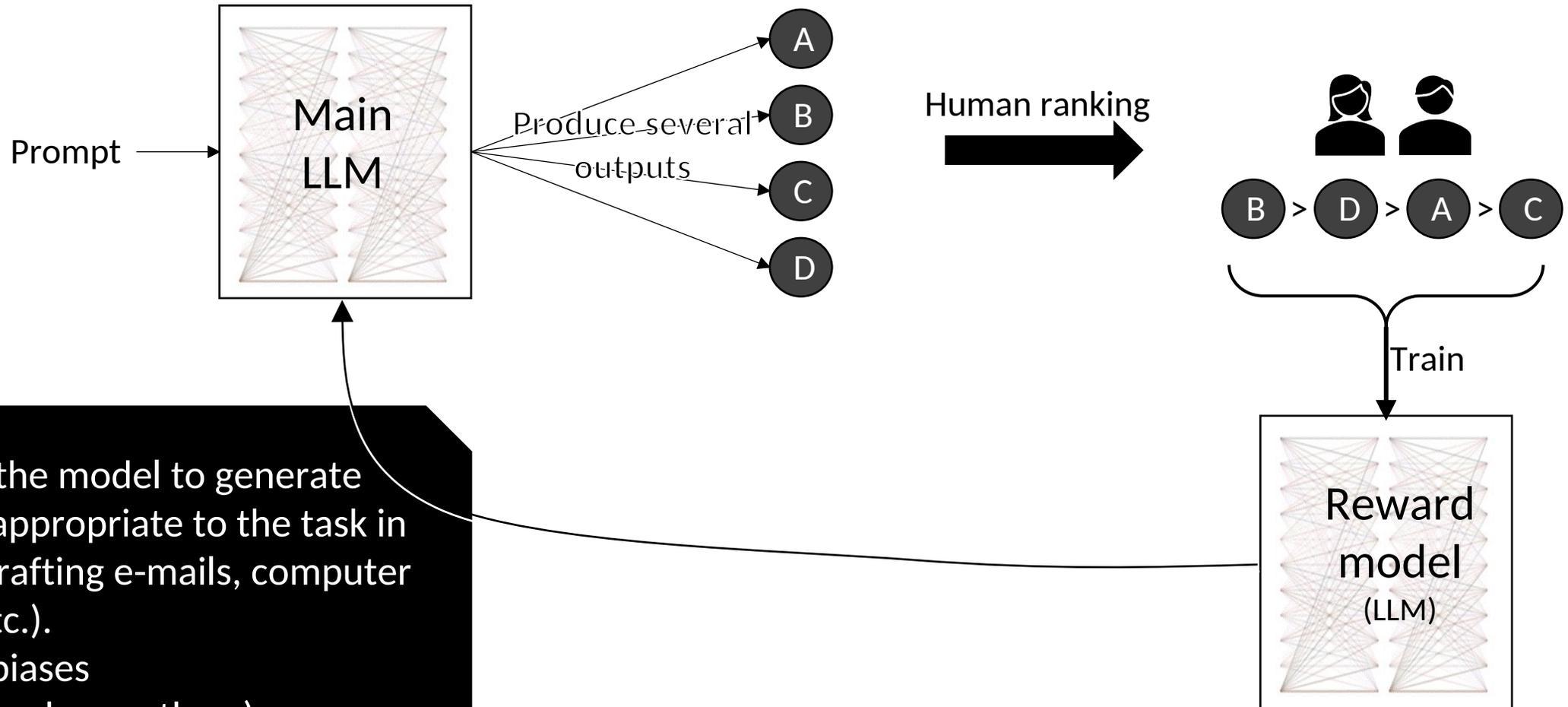
Auto-regressive language models (*decoder-style, e.g. GPT*)

User: What is a large language model?

Assistant: A large language model is a type of artificial intelligence (AI) model designed to understand and generate human-like language. These models are **built**



Reinforcement Learning with Human Feedback



- Adapts the model to generate results appropriate to the task in hand (drafting e-mails, computer code, etc.).
- Avoids biases (and introduces others)

Why are LLMs so good?

- LLMs are good at producing **relevant**, **well-written** and **convincing** information, thanks to:
 - **Representation** learning
 - **Contextual** understanding and attention mechanisms
(capture long-range dependencies and relationships in the data)
 - **Scalability** and massive amount of training data
(wide range of linguistic nuances and topics)
 - Massive finetuning with **human feedback**
(produce the right kind of results for the task prompted by the user)
- LLMs are NOT good at:
 - **Factuality / knowledge**
 - **High precision**
 - **Humor, creativity, originality**

How can we use LLMs?

- Direct prompt: *“How do genetic mutations in the ALMS1 gene contribute to the pathophysiology of Alström Syndrome?”*
 - ▶ only knowledge from the pretraining
- Prompt with persona: *“You’re an assistant specialized in research on rare diseases. How do genetic mutations in the ALMS1 gene contribute to the pathophysiology of Alström Syndrome?”*
 - ▶ allows to guide the answer and its style
- Prompt with document: *“Given this document, how do genetic mutations in the ALMS1 gene contribute to the pathophysiology of Alström Syndrome?”*
 - ▶ knowledge from the pretraining + supporting document

How can we use LLMs?

- Few-shot prompting: “

Classify the following rare diseases into their appropriate categories: Genetic Disorder or Neurodegenerative Disorder

Input: "Alström Syndrome"

Output: "Genetic Disorder"

Input: "Batten Disease"

Output: "Neurodegenerative Disorder"

Input: "Huntington's Disease"

Output: "Neurodegenerative Disorder"

Input: "Marfan Syndrome"

Output: "Genetic Disorder"

Input: "Parkinson's Disease"

Output: "Neurodegenerative Disorder"

Input: "Cystic Fibrosis"

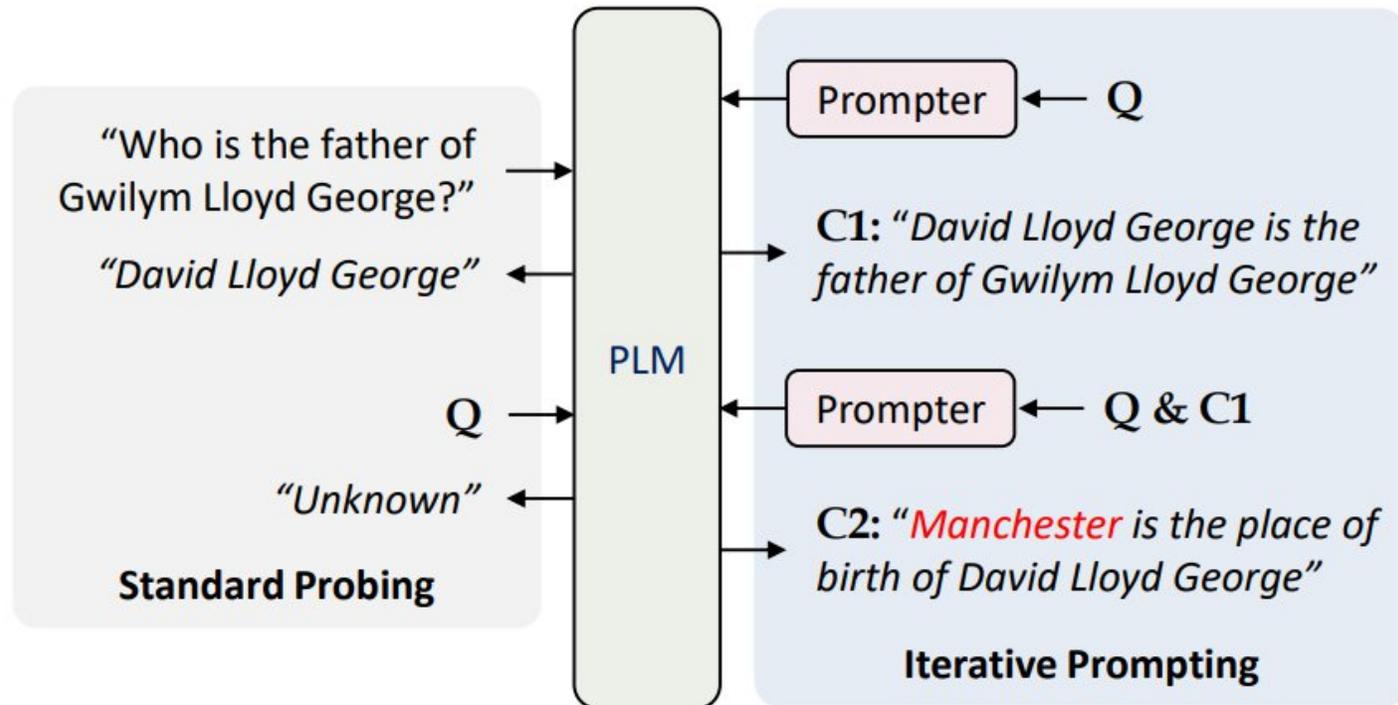
Output:

”

How can we use LLMs?

- Prompt chaining:

Q: "What is the place of birth of Gwilym Lloyd George's father?"
(Answer: **Manchester**)



Wang, B., Deng, X., & Sun, H. (2022). Iteratively Prompt Pre-trained Language Models for Chain of Thought. *Conference on Empirical Methods in Natural Language Processing*.

How can we use LLMs?

- Chain of thoughts:

StrategyQA

Q: Yes or no: Would a pear sink in water?

A: The density of a pear is about 0.6 g/cm^3 , which is less than water. Thus, a pear would float. So the answer is no.

Date Understanding

Q: The concert was scheduled to be on 06/01/1943, but was delayed by one day to today. What is the date 10 days ago in MM/DD/YYYY?

A: One day after 06/01/1943 is 06/02/1943, so today is 06/02/1943. 10 days before today is 05/23/1943. So the answer is 05/23/1943.

Sports Understanding

Q: Is the following sentence plausible? "Joao Moutinho caught the screen pass in the NFC championship."

A: Joao Moutinho is a soccer player. The NFC championship is part of American football, not soccer. So the answer is no.

SayCan (Instructing a robot)

Human: How would you bring me something that isn't a fruit?

Explanation: the user wants something to eat that isn't a fruit. An energy bar is not a fruit, so I will bring the user an energy bar.

Plan: 1. find(energy bar) 2. pick(energy bar) 3. find(user) 4. put(energy bar) 5. done().

Last Letter Concatenation

Q: Take the last letters of the words in "Lady Gaga" and concatenate them.

A: The last letter of "Lady" is "y". The last letter of "Gaga" is "a". Concatenating them is "ya". So the answer is ya.

Coin Flip (state tracking)

Q: A coin is heads up. Maybelle flips the coin. Shalonda does not flip the coin. Is the coin still heads up?

A: The coin was flipped by Maybelle. So the coin was flipped 1 time, which is an odd number. The coin started heads up, so after an odd number of flips, it will be tails up. So the answer is no.

Wei, J., Wang, X., Schuurmans, D., Bosma, M., Chi, E.H., Xia, F., Le, Q., & Zhou, D. (2022). Chain of Thought Prompting Elicits Reasoning in Large Language Models. *Conference on Neural Information Processing Systems*.

How can we use LLMs?

- Self-Consistency
- Generated Knowledge Prompting
- Tree of Thoughts (ToT)
- Automatic Prompt Engineer (APE)
- Active-Prompt
- Directional Stimulus Prompting
- Program-Aided Language Models
- ... (see for example <https://www.promptingguide.ai/>)

Generative LLMs: Pros & Cons

	Rules	Supervised learning	Promise of gLLMs
General performance	Yellow	Green	Grey with ?
Ease of implementation	Green	Green	Green
Need for human expertise	Yellow	Red	Green
Explainability	Green	Yellow	Yellow
Material resources	Green	Yellow	Red
Energy consumption	Green	Yellow	Red
Ease of maintenance	Yellow	Yellow	Yellow
Generalization to a different problem/context	Red	Red	Yellow



A thorough evaluation of these models
is always necessary!

Whatever the method (rules, ML, LLMs...),
build a test set for the evaluation

Do not prompt public or API-based LLMs
with sensitive data!

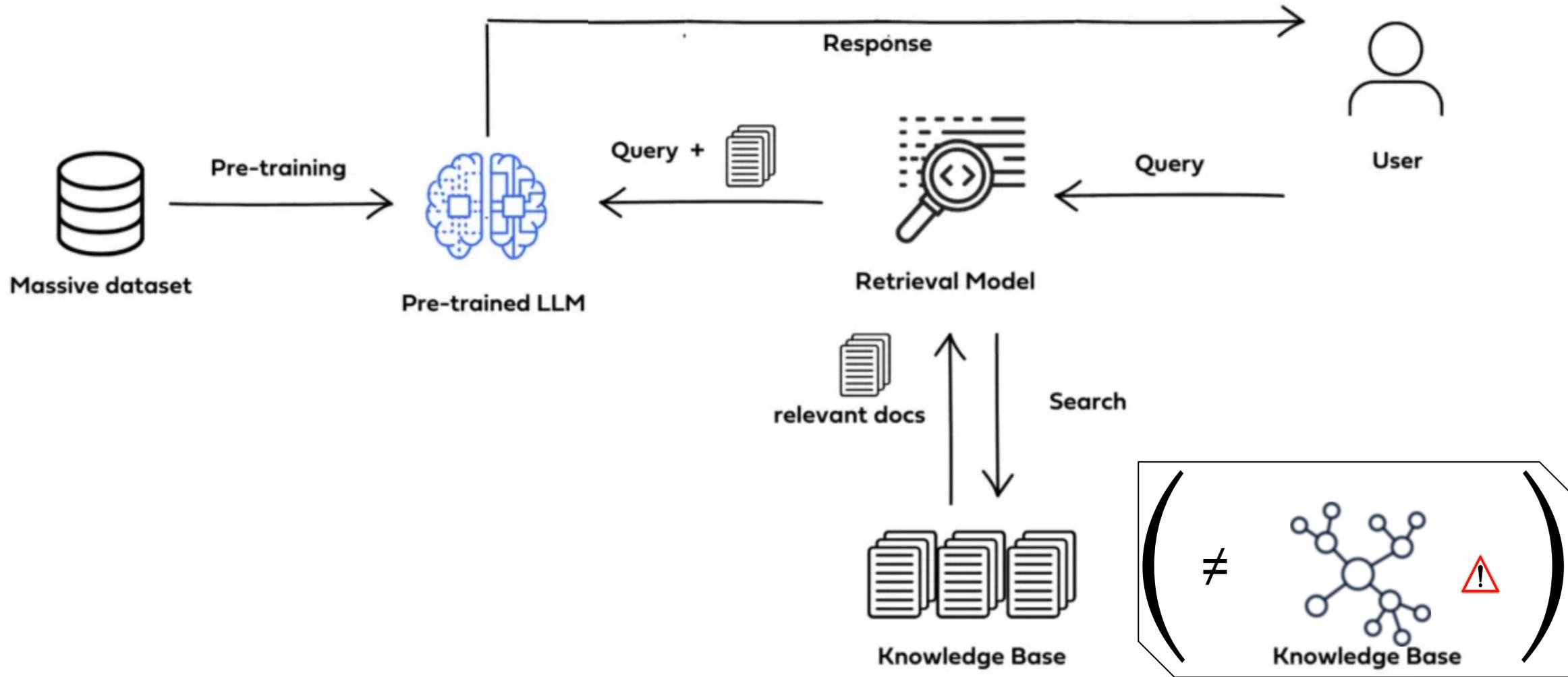


Methods

1. Rule-based systems
2. Supervised learning systems
3. Generative, large language models
- 4. Retrieval-Augmented Generation**

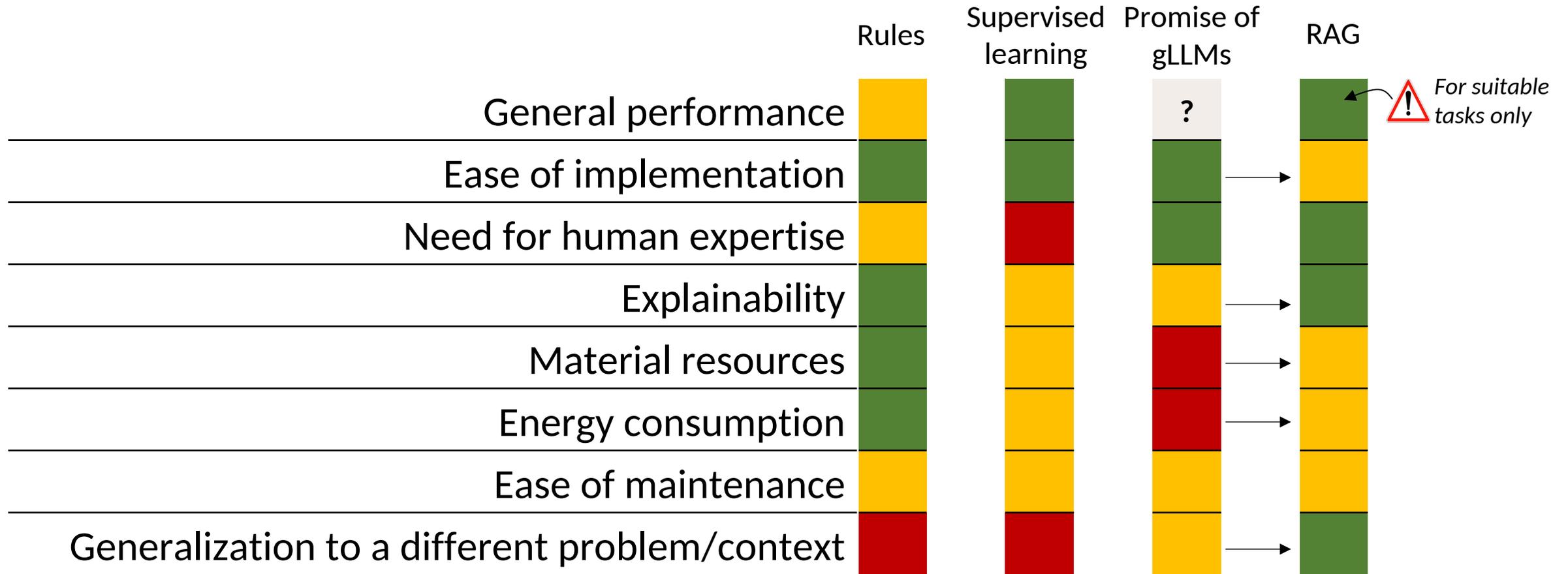
The new search engine

Retrieval-Augmented Generation

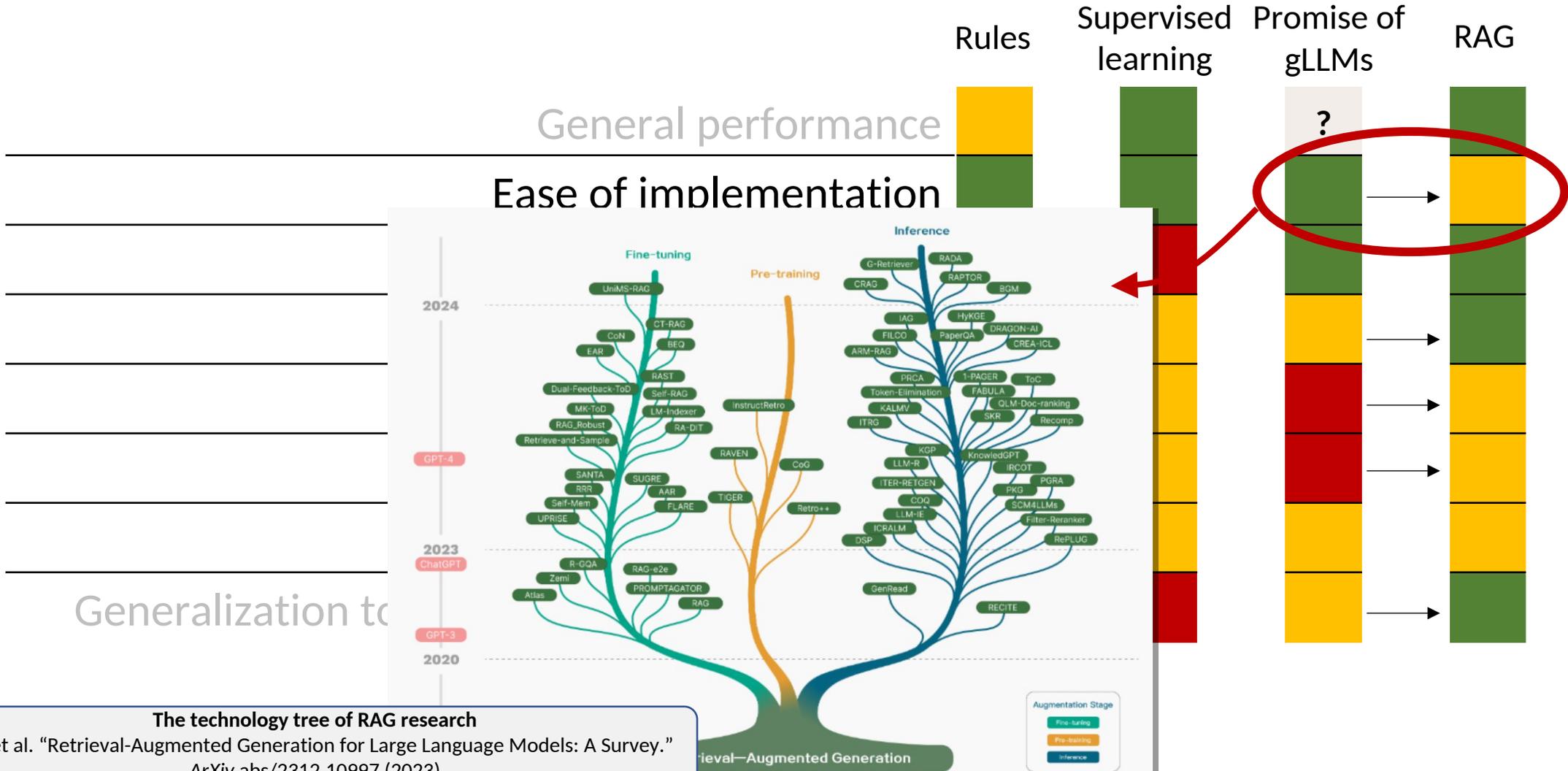


<https://medium.com/@krtarunsingh/introduction-to-retrieval-augmented-generation-rag-and-its-transformative-role-in-ai-c07e35da7f01>

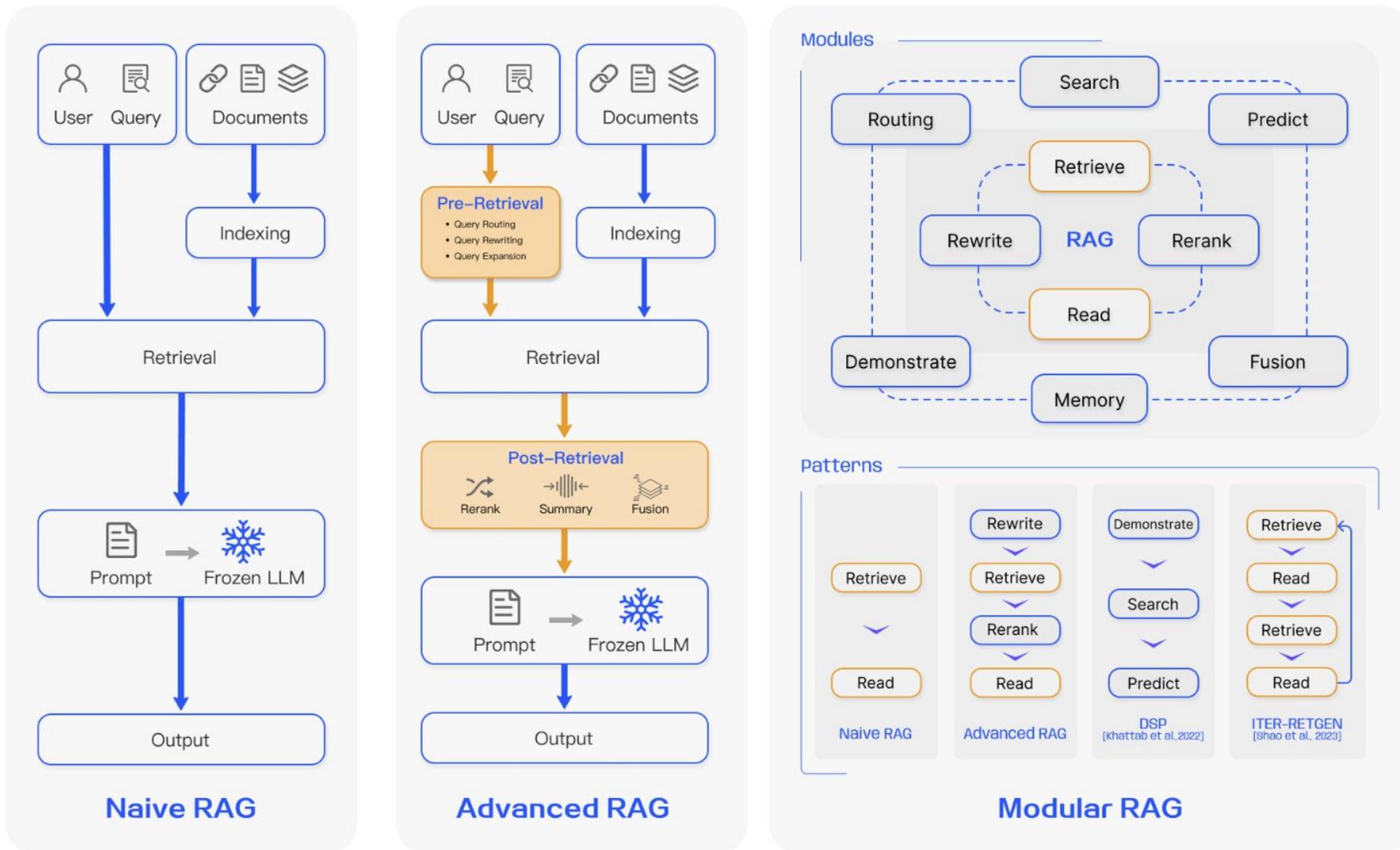
RAG: Pros & Cons



RAG: Pros & Cons



RAG



Gao et al. "Retrieval-Augmented Generation for Large Language Models: A Survey."
ArXiv abs/2312.10997 (2023)

