

*Inria*

# Projet RadioNLP

INRIA-SED LYON

Salim Sadoune

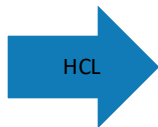
## Présentation de projet RadioNLP

- Le projet RadioNLP est créé dans le cadre de la collaboration entre le pôle de développement IA-HCL et l'équipe INRIA pour les applications de traitement du langage naturel (NLP) en radiologie.

### Objectif

Développer une méthode de traitement du langage naturel basée sur les **transformateurs**, adaptée à la radiologie pour **l'analyse et l'extraction des informations** dans les comptes rendus de scanner thoraciques;

- **Détection des pathologies**
- **Reconnaissances de négation**
- **Reconnaissances de l'incertitude**
- ...

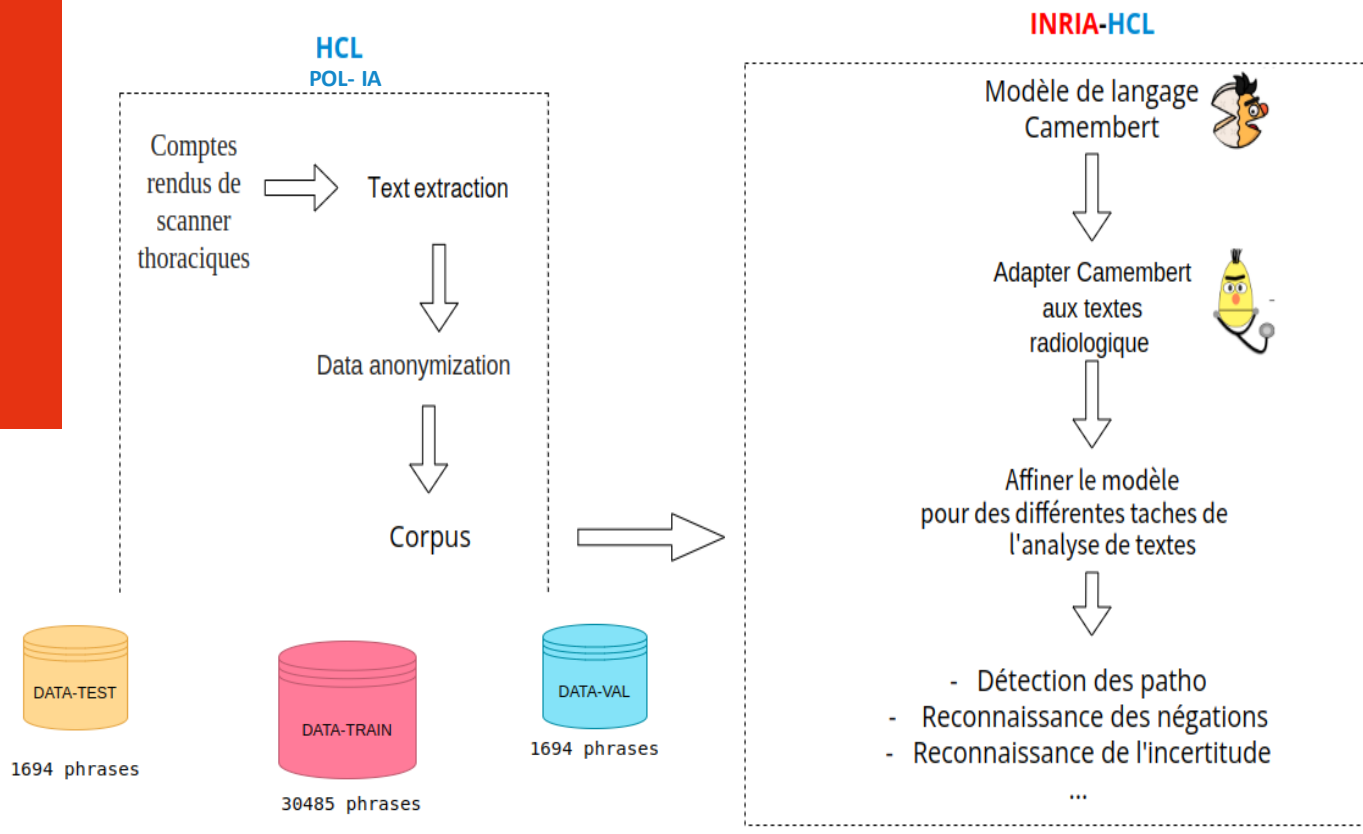


Intégrer les techniques de traitement automatique du langage dans les outils d'analyse de données du système d'information patients des HCL.

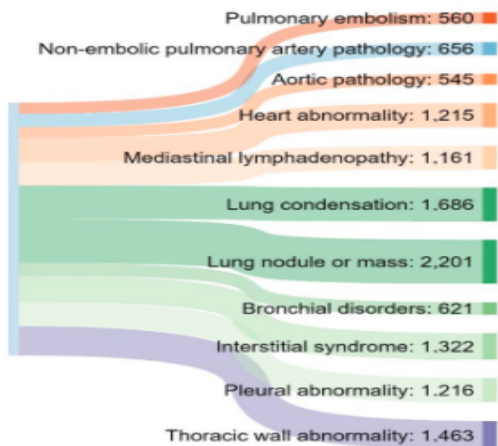
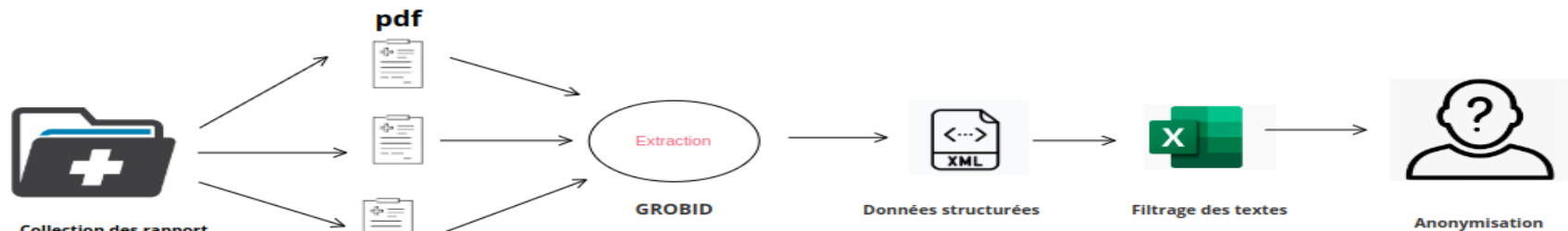


Intégrer le travail dans un projet open source pour la communauté scientifique.

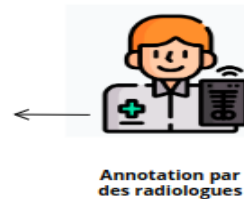
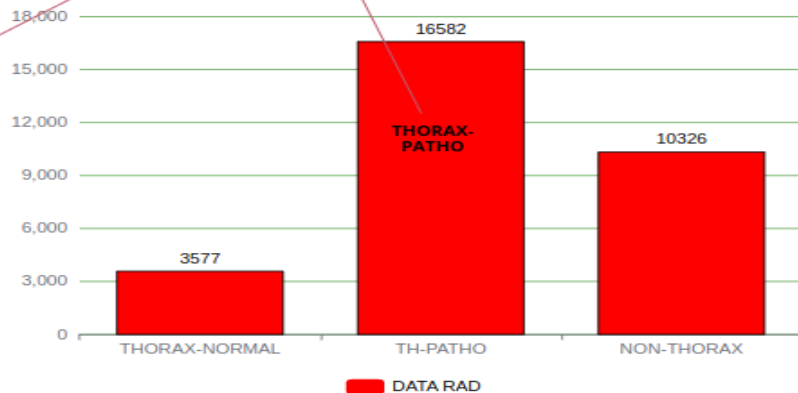
# PLAN



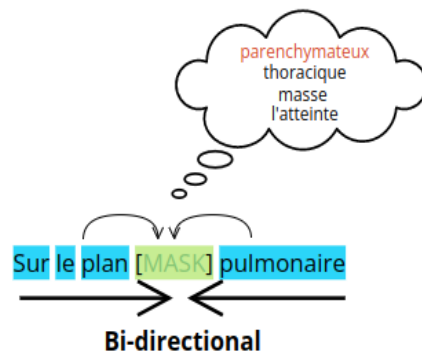
# Préparation de corpus



23 pathologies




## Adapter le modèle de langage camembert à la radiologie



MASK Camembert

**Radiology Free-Text**

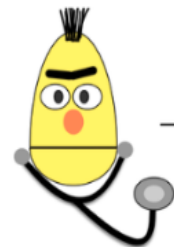
Ces micronodules ne sont pas suspects  
Nodules solides n'excédant pas 4 mm apical droit  
absence de signe de dissection aoAorte athéromateuse  
A l'étage thoracique : Sur le plan parenchymateux pulmonaire  
A son voisinage on peut décrire ce jour une lésion excavée de 21 mm de diamètre qui était punctiforme  
Aspect stable du nodule sous-pleural du segment antérieur du lobe moyen, mesuré à 9 mm ; pas d'apparition de nouvelle lésion  
Aorte : Prothèse aortique en place avec absence de fuite péri anastomotique  
On ne reviendra pas sur le montage de Fontan



Masked-Language Modeling

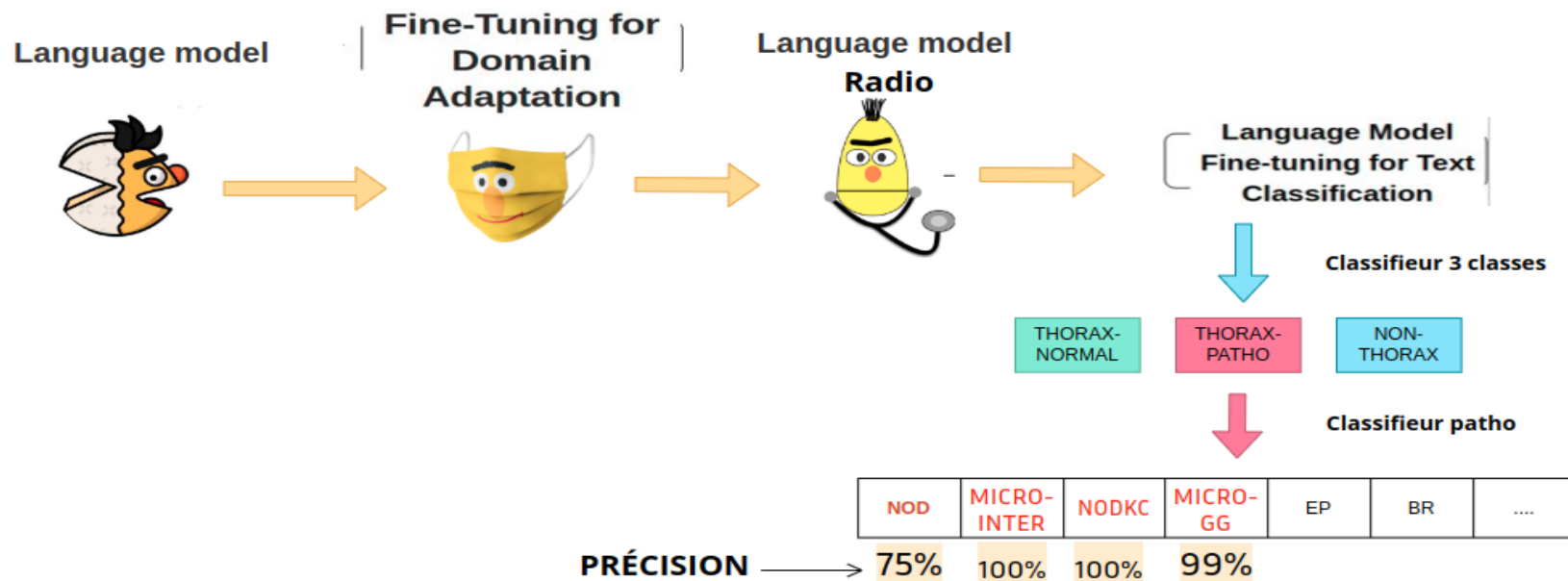


Camembert Radio



# Détection des pathologies thoraciques

- Affiner le modèle de Camembert sur la tâche de classification des phrases

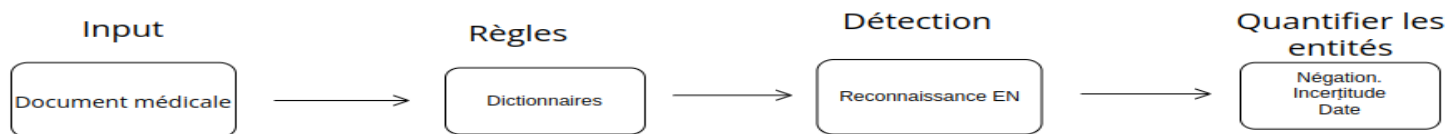


# Extraction d'informations dans les données radio-cliniques

# Approche basée sur les règles

## EDS-NLP:

extraction d'informations ( négation, l'incertitude, famille, date...) à partir des textes cliniques en français



"Motif de prise en charge : probable pneumopathie a **COVID19**, "  
 "sans difficultés **respiratoires**"  
 "Le père du patient est **asthmatique**."  
 Motif de prise en charge : probable pneumopathie a un coronavirus "

variante_lexicale	étiquette	négation	hypothèse	famille
COVID-19	covid	Faux	Vrai	Faux
respiratoires	respiratoire	Vrai	Faux	Faux
asthmatique	respiratoire	Faux	Faux	Vrai

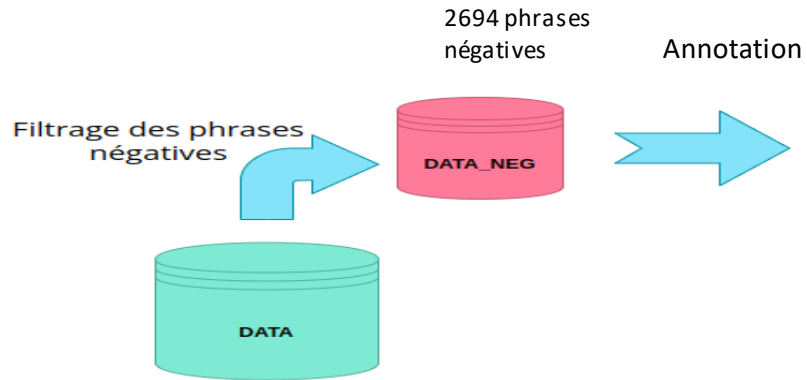
# Application pour la reconnaissance de négation sur les textes radiologiques

- Absence d'atteinte alvéolaire, interstitielle diffuse ou nodulaire suspecte
- |        |        |
|--------|--------|
| Indice | Portée |
|--------|--------|

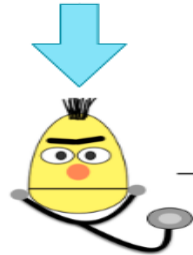
## Limitation

- La portée de négation est la phrase entière ou qu'une partie
- Généraliser la méthode.

# Approche basée sur l'apprentissage automatique



Sentence 1	Nodule calcifié sous-pleural du lobe inférieur droit, <b>pas</b> de nouveau nodule décelé
Sentence 2	<b>absence</b> de signe de dissection aortique : Aorte athéromateuse
Sentence 3	Stigmates de résection au niveau du culmen, <b>il n'y a pas</b> de nodule pulmonaire suspect visible

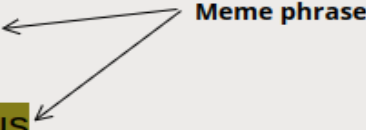


Token classification

## Evaluation de modèle

precision	recall	F1-score	accuracy
0.94	0.9189	0.9293	0.9698

## Exemples

1. Multiples micronodules péribronchovasculaires d'aspect branché diffus au niveau des 2 champs pulmonaires en en faveur d'une surinfection bronchique **pas** d'épanchement pleural
  2. Doute sur un ulcère pénétrante de l'aorte descendante non bien évaluable en **absence** de produit de contraste
  3. les micronodules **ne** sont **pas** suspects
  4. Certaines personnes **n'** attrapent **pas** le Covid-19
  5. Certaines personnes **n'** attrapent **pas** le Coronavirus
- 
- Meme phrase

# Futures pistes de recherche

- Construire un modèle de langage large adaptés à la radiologie.
- Affiner le modèle sur les différentes taches de NLP
- Généraliser et améliorer l'approche automatique pour l'extraction des informations