

A white line-art graphic of a stethoscope, positioned on the left side of the title. The earpieces are at the top, and the chest piece is at the bottom, with the tubing looping around.

# Reconnaissance des Maladies

Animé par:

**OURARA  
YASSINE**



# Sommaire

---

Introduction

---

Dataset

---

Pré-traitement / Vectorisation

---

Modèle intelligent

---

Code

---

Conclusion

NLP\_project

# Introduction

---



# Les débuts

---

//01

L'historique de la reconnaissance de maladies en NLP remonte à plusieurs décennies. Les premières recherches dans ce domaine ont commencé dans les années 1980 et 1990, lorsque les chercheurs ont commencé à explorer les possibilités de l'utilisation de l'intelligence artificielle (IA) pour améliorer le diagnostic des maladies.

■■■

---

Au fil des ans, la reconnaissance de maladies en NLP est devenue de plus en plus sophistiquée, grâce aux avancées dans les techniques d'apprentissage automatique et les méthodes de traitement du langage naturel. Les systèmes de reconnaissance de maladies en NLP sont maintenant capables d'analyser des quantités massives de données textuelles liées à la santé, telles que des notes de consultation, des dossiers médicaux et des articles scientifiques.

# Plus loin

---

Aujourd'hui, la reconnaissance de maladies en NLP est utilisée dans une variété de contextes médicaux, allant du diagnostic de maladies rares ou complexes à l'analyse de grandes bases de données de santé pour découvrir de nouvelles tendances et de nouveaux traitements. Les modèles de reconnaissance de maladies en NLP sont également utilisés pour améliorer la surveillance des maladies infectieuses et pour aider les professionnels de la santé à suivre la progression des maladies chez les patients.

# Exemple

---

Amazon Comprehend Medical est un service qui utilise le NLP pour extraire les états pathologiques, les médicaments et les résultats des traitements à partir des notes des patients, des rapports d'essais cliniques et d'autres dossiers médicaux électroniques.



?

---

Quelles sont les différentes  
techniques utilisées pour la  
reconnnaissances de  
maladies ?



NLP\_project

# Extraction d'entités médicales

# Extraction d'entités médicales <sup>//05</sup>

---

## Pattern matching

- Utilisation de motifs prédéfinis pour identifier des entités médicales spécifiques dans les textes.
- Les motifs peuvent être des expressions régulières, des règles linguistiques ou des dictionnaires de termes médicaux.
- Avantages : simplicité, rapidité d'implémentation, adaptabilité à des cas spécifiques.
- Limites : dépendance des motifs prédéfinis, difficulté à couvrir tous les cas possibles, sensibilité aux variations linguistiques et aux erreurs de correspondance.

## Rule based methods

- Utilisation des règles définies manuellement pour extraire des entités médicales.
- Les règles peuvent être fondées sur des connaissances expertes ou sur des patrons syntaxiques spécifiques.
- Avantages : flexibilité pour inclure des connaissances spécifiques, interprétabilité des règles définies.
- Limites : besoin d'une expertise humaine pour la définition des règles, difficulté à couvrir toutes les variations possibles, sensibilité aux erreurs de règle.

# Extraction d'entités médicales <sup>//06</sup>

---

## Apprentissage supervisé

- Entraînement d'un modèle d'apprentissage automatique en utilisant des exemples annotés de textes médicaux.
- Les exemples annotés comprennent des entités médicales marquées, permettant au modèle de généraliser et de prédire les entités dans de nouveaux textes.
- Avantages : capacité d'apprentissage à partir de données, performances potentiellement élevées.
- Limites : besoin d'un ensemble de données annotées de haute qualité, exigences en termes de ressources de calcul et de puissance de traitement.

## Apprentissage non supervisé

- Extraire des entités médicales à partir de données non annotées sans utiliser d'exemples annotés.
- Les techniques non supervisées comprennent le clustering, la détection d'anomalies, l'extraction de termes clés, etc.
- Avantages : indépendance des données annotées, potentiel pour découvrir des entités inconnues, adaptabilité à des cas spécifiques.
- Limites : difficulté à évaluer la qualité de l'extraction sans annotations, possibilité d'erreurs d'extraction, besoin de validation humaine pour garantir la précision.

NLP\_project

# Applications

---

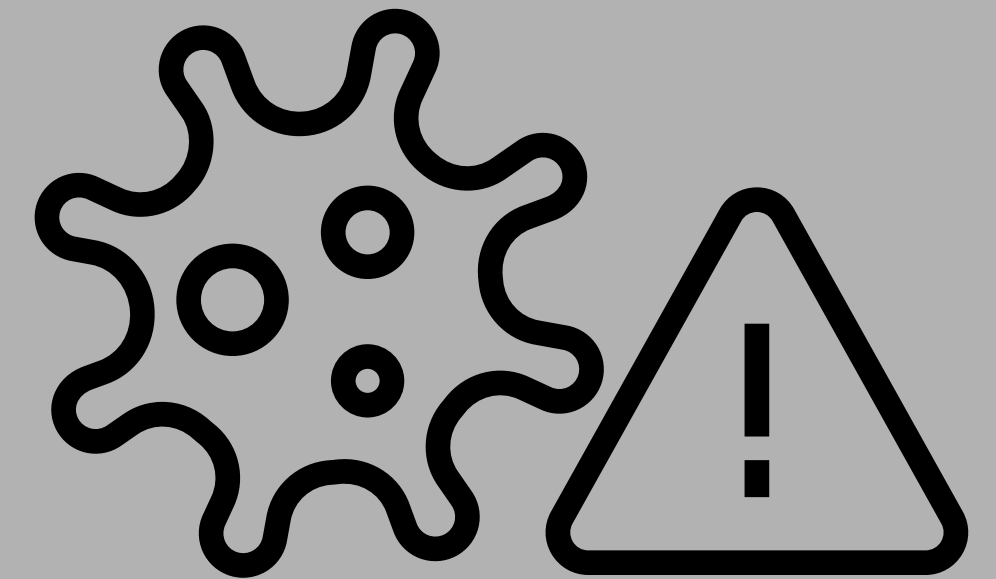
# Applications

---

//07

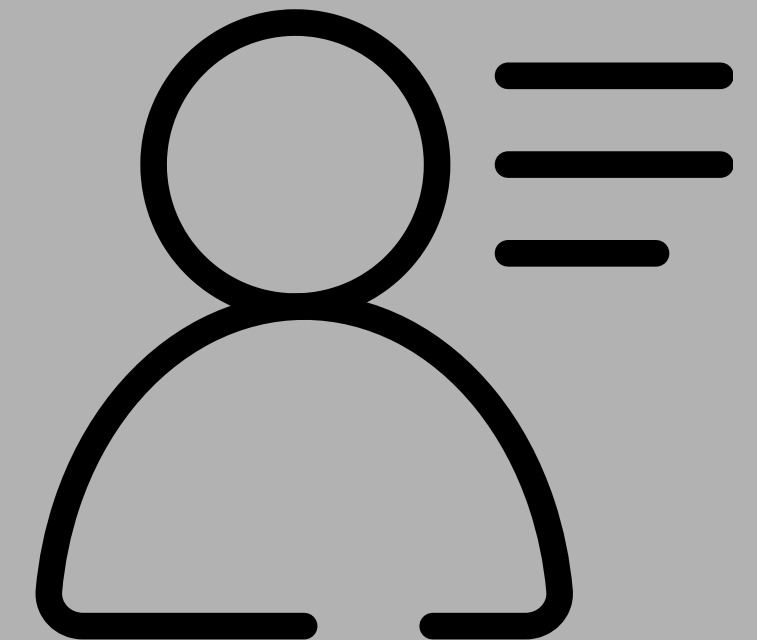
## Surveillance des épidémies

Une étude publiée dans "**Nature Biotechnology**" en novembre 2020 par Wynants a utilisé des techniques de **NLP** pour développer un modèle de prédiction de la COVID-19. Les chercheurs ont analysé les symptômes rapportés par les patients via une application mobile, en extrayant et en analysant les informations textuelles. Le modèle de prédiction développé était capable d'identifier les individus à haut risque d'être atteints de la COVID-19.



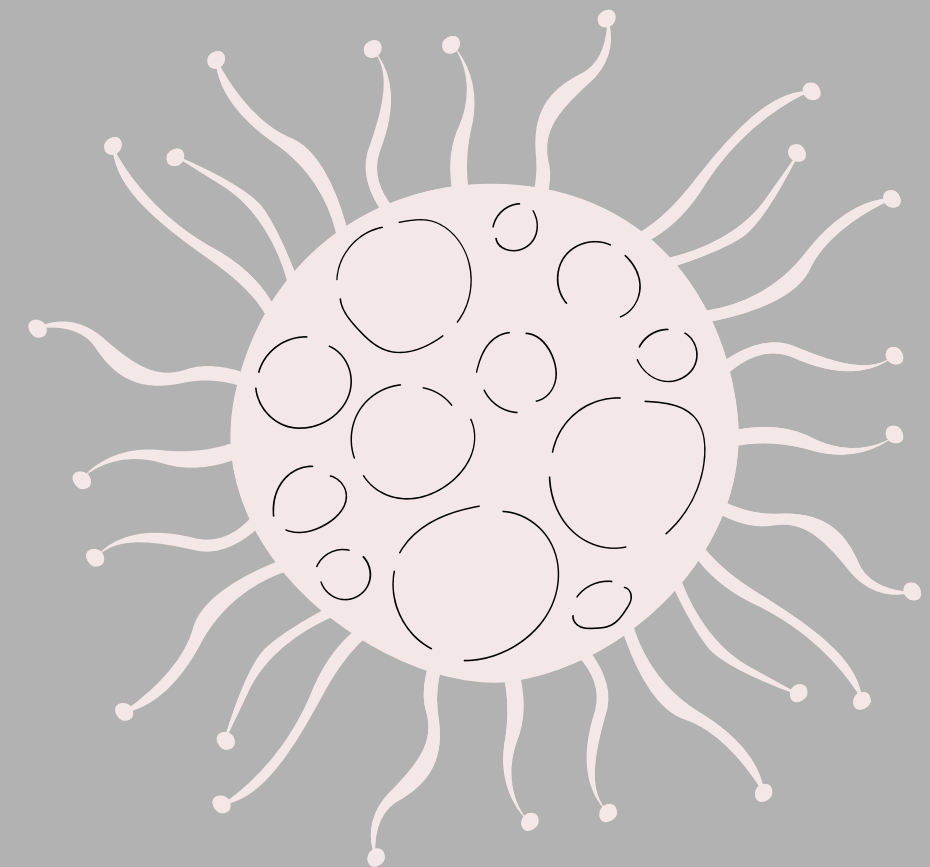
## Compréhension des plaintes des patients

Une étude publiée en mars 2021 dans la revue "**Artificial Intelligence in Medicine**" par Jiang a utilisé des techniques de **NLP** pour analyser les descriptions de symptômes des patients et aider à poser des diagnostics précis pour les maladies dermatologiques. Le système basé sur la reconnaissance de maladies en NLP a montré une précision élevée dans l'identification des maladies cutanées.



## Détection de cancers

Une étude publiée en janvier 2022 dans la revue "**Nature Medicine**" par Esteva. a utilisé des techniques de **NLP** pour extraire des informations à partir de rapports d'imagerie médicale dans le cadre de la détection du cancer du sein. L'étude a montré que le système basé sur la reconnaissance de maladies en NLP surpassait les radiologues humains en termes de précision de détection.





?

---

Sur quoi le médecin s'appuie  
pour établir un diagnostic ?

# Motivation

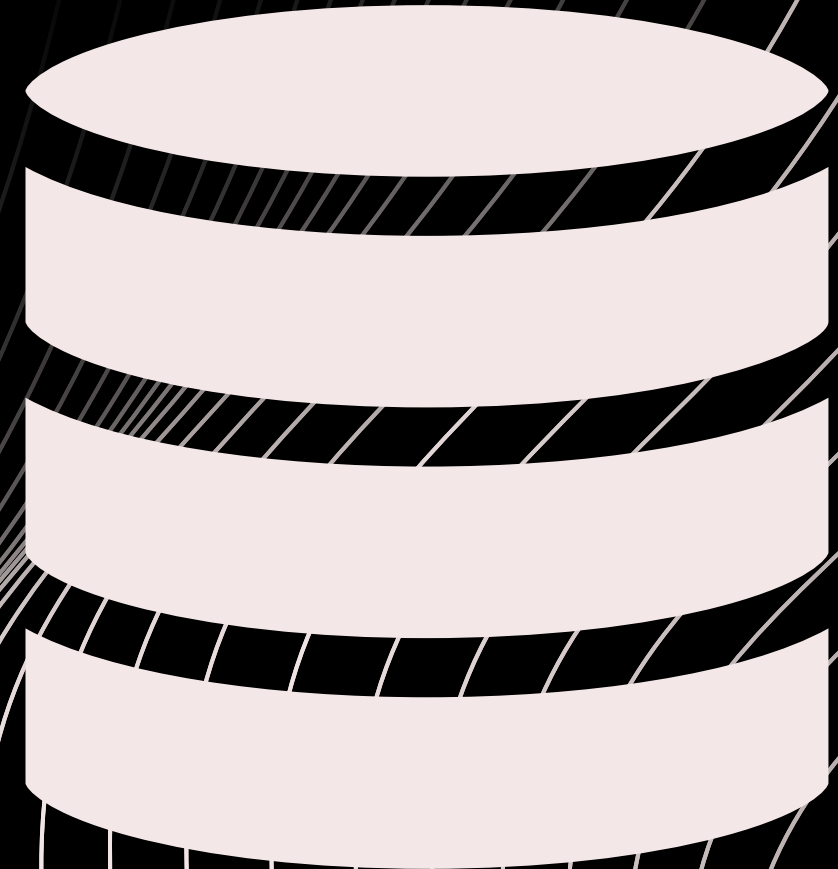
---

Le NLP aide les ordinateurs à interpréter le langage humain et à y répondre. Outre l'observation visuelle, la description des symptômes par le patient est l'un des principaux éléments sur lesquels s'appuie un médecin pour établir un diagnostic ou réduire les possibilités. Si les algorithmes du NLP peuvent aider à répondre aux questions de sélection initiale, les médecins peuvent passer moins de temps à trier et à demander des informations générales. Au lieu de cela, ils peuvent prescrire des tests et examiner des problèmes spécifiques.

NLP\_project

# Dataset

---



# Symptom2Disease

---

L'ensemble de données se compose :  
1200 lignes et 2 colonnes .

	Unnamed: 0	label	text
0	0	Psoriasis	I have been experiencing a skin rash on my arm...
1	1	Psoriasis	My skin has been peeling, especially on my kne...
2	2	Psoriasis	I have been experiencing joint pain in my fing...

Dataset

label : contient les étiquettes de la maladie

text : contient les descriptions des symptômes en langage naturel.

Les 24 maladies suivantes ont été couvertes dans l'ensemble de données :

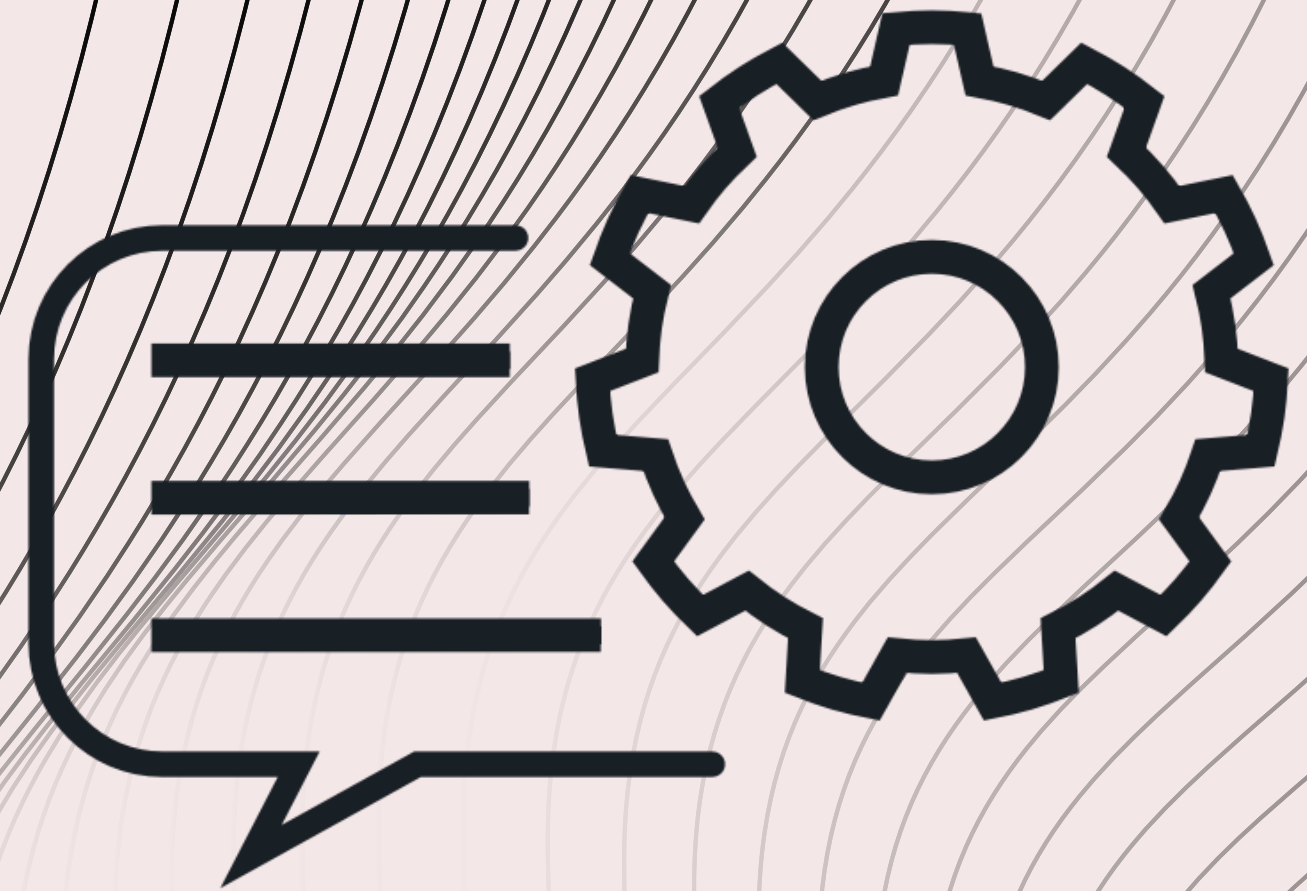
Psoriasis, varices, typhoïde, varicelle, impétigo, dengue, infection fongique, rhume, pneumonie, hémorroïdes dimorphiques, arthrite, acné, asthme bronchique, hypertension, migraine, spondylose cervicale, jaunisse, paludisme, infection des voies urinaires, allergie, reflux gastro-œsophagien, réaction médicamenteuse, ulcère gastroduodénal, diabète.

NLP\_project

# Pré-traitement

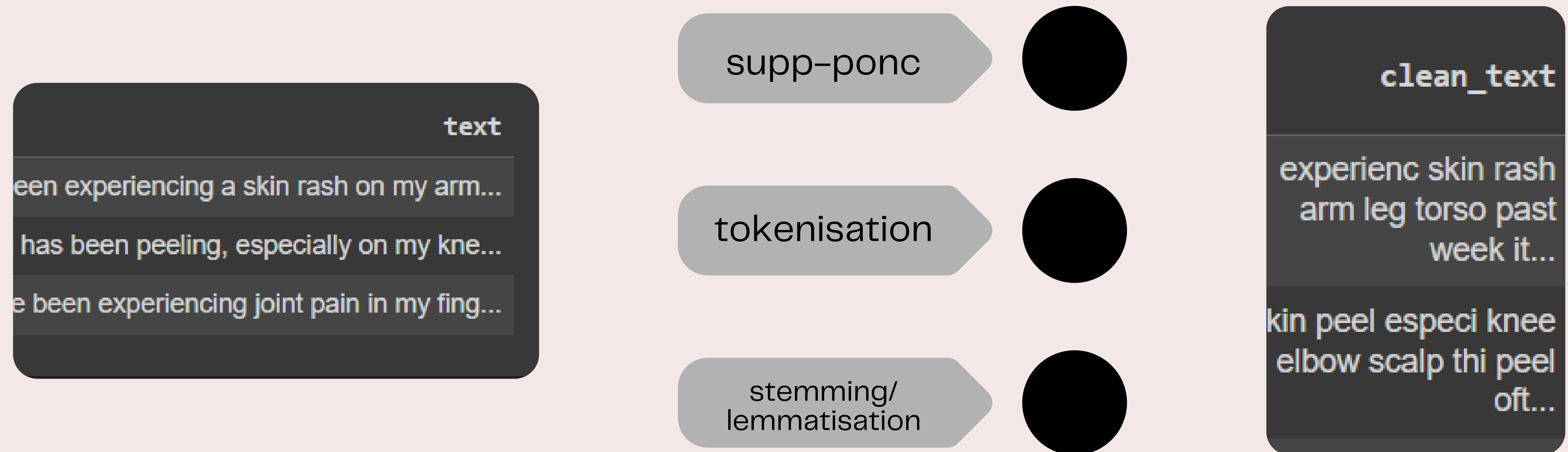
---

# Vectorisation



# Pré-traitement

---





# Vectorisation

---

//14

## TF-IDF

---

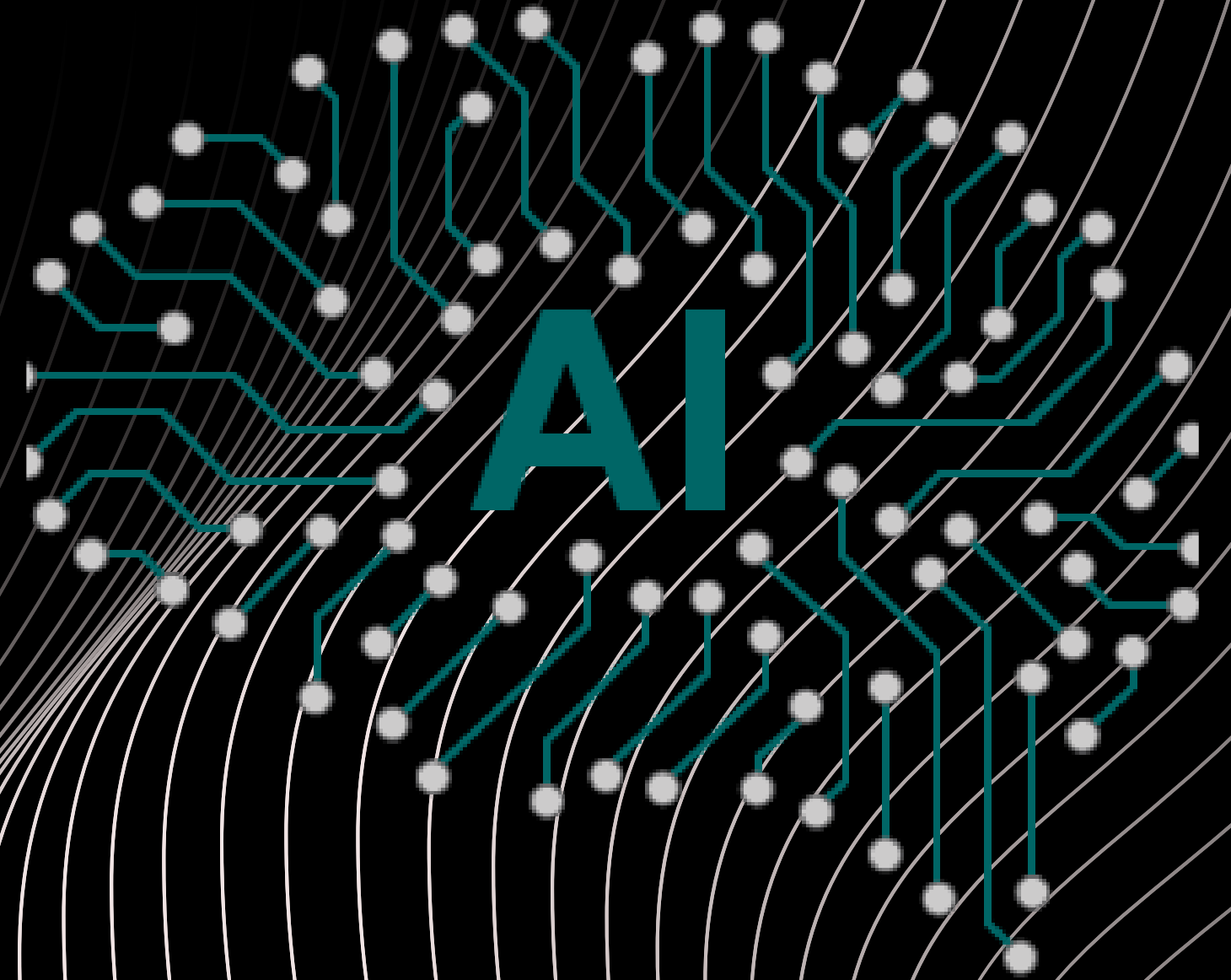
On a besoin de savoir la fréquence de chaque mot et l'importance du terme dans l'ensemble de la phrase afin de prédire la maladie que décrit le texte du patient.

Donc un vecteur Tf-Idf a été ajusté au texte et un objet `OneHotEncoder(get_dummies)` a été adapté aux maladies (labels).

NLP\_project

# Modèle AI

---



- On divise les données en ensembles d'entraînement et de test..10% des données seront utilisées comme ensemble de test, tandis que le reste sera utilisé comme ensemble d'entraînement.

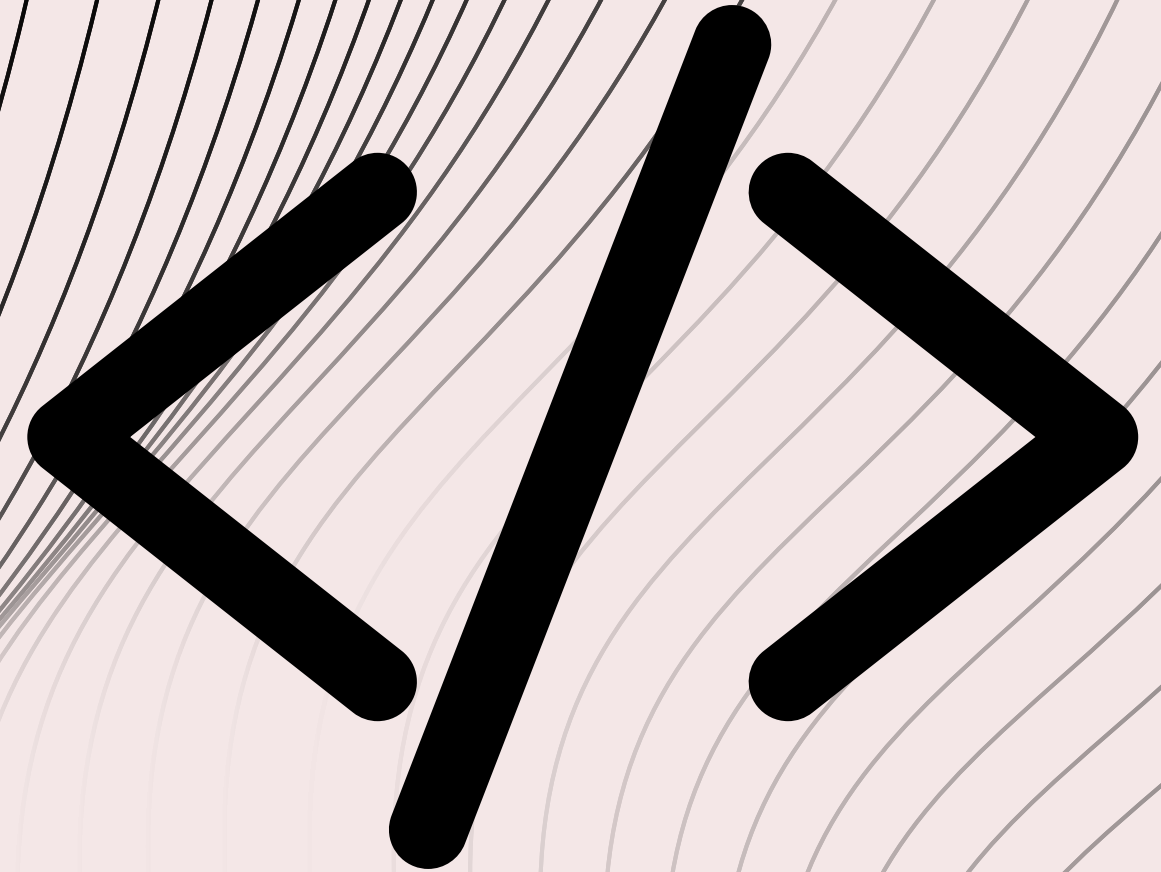
- Le modèle est créé en utilisant `Sequential()`, ce qui signifie que les couches sont empilées séquentiellement les unes sur les autres.
- Le modèle a trois couches denses. La première couche a 128 neurones avec une fonction d'activation "ReLU." La couche suivante est une couche de dropout, qui désactive aléatoirement certains neurones pendant l'entraînement pour prévenir le surapprentissage.
- Ensuite, il y a une deuxième couche dense de 64 neurones avec une fonction d'activation ReLU, suivie d'une autre couche de dropout.
- Enfin, la dernière couche dense a un nombre de neurones égal au nombre d'étiquettes de sortie (`len(y_train[0])`) et utilise une fonction d'activation softmax pour prédire les intentions de sortie.

- Evaluation du modèle et ajustement des poids du réseau neuronal
- Entraînement du modèle en effectuant 100 époques d'entraînement.

NLP\_project

# Code

---



# Projet\_YassineOurara\_RecM

---

//18



Code source: [ici](#)



# Conclusion

---

