

# WP XAI : activités et perspectives

Journées MAIA - Boulogne-sur-mer, 23 mai 2024



# Expliquer ... Vaste programme !

- ▶ C'est nécessaire !
  - ▶ L'IA à base de ML **impacte nos activités quotidiennes**

# Expliquer ... Vaste programme !

- ▶ **C'est nécessaire !**

- ▶ L'IA à base de ML **impacte nos activités quotidiennes**
- ▶ Mais les modèles ML ont des **limites**
  - ▶ Assurer des prédictions 100% correctes : impossible !
  - ▶ **Sensibilité aux données** (qualité, quantité), *garbage in, garbage out* ...
  - ▶ Les modèles ML les plus performants en terme de prédiction sont des **boîtes noires**
  - ▶ Il faut permettre à l'utilisateur de **décider quoi faire des prédictions fournies**

# Expliquer ... Vaste programme !

## ▶ C'est nécessaire !

- ▶ L'IA à base de ML **impacte nos activités quotidiennes**
- ▶ Mais les modèles ML ont des **limites**
  - ▶ Assurer des prédictions 100% correctes : impossible !
  - ▶ **Sensibilité aux données** (qualité, quantité), *garbage in, garbage out* ...
  - ▶ Les modèles ML les plus performants en terme de prédiction sont des **boîtes noires**
  - ▶ Il faut permettre à l'utilisateur de **décider quoi faire des prédictions fournies**

## ▶ ... mais c'est difficile !

- ▶ Données brutes et données symboliques
- ▶ Plusieurs types et formats pour les explications
- ▶ Choix multi-critère

## Co-12 Properties

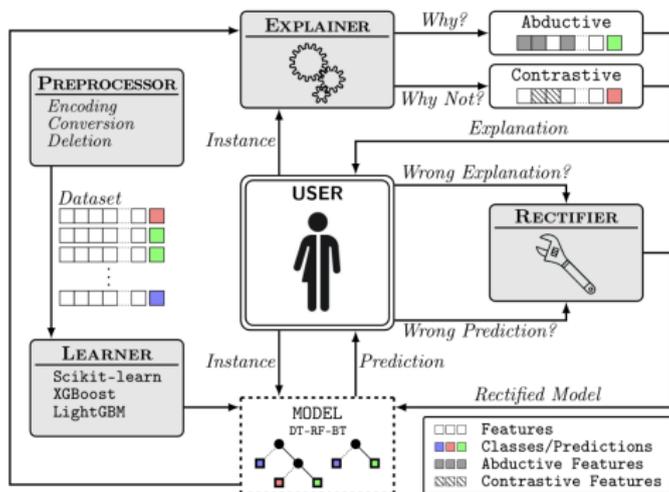
<b>Correctness</b> Match between model and explanation. 	<b>Completeness</b> How much of the model is explained? 	<b>Consistency</b> Robustness to small changes in model and implementation. $g(x) = g(x)$	<b>Continuity</b> Robustness to small changes input. $g(x) = g(x')$
<b>Contrastivity</b> Discriminative to other events or targets? $g(x \text{Cat}) \neq g(x \text{Dog})$	<b>Covariate Complexity</b> Complexity of features in the explanation 	<b>Compactness</b> Size of the explanation 	<b>Composition</b> Presentation format 
<b>Confidence</b> Probability information available? $p = ?$	<b>Context</b> Useful for users? 	<b>Coherence</b> Match with domain knowledge. $g(x) =$ 	<b>Controllability</b> Can user influence explanation? $g(x)$ 

Explanation / Model / User

[Nauta et al., ACM Computing Survey 2023]

- ▶ **Nombreux critères, certains antagonistes**
- ▶ **Dépendance à l'utilisateur**  
(disposer d'un modèle de celui-ci, variété des utilisateurs)
- ▶ **Co-construire des explications avec l'utilisateur via une **interaction complexe****

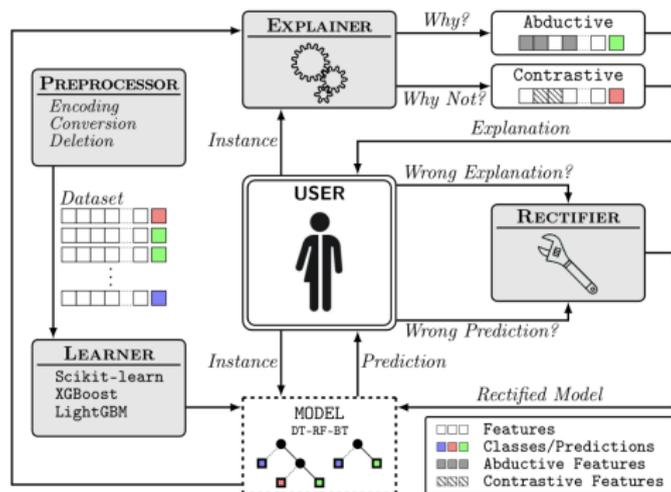
# CRIL: développement de PyXAI



User interaction with PyXAI.

- ▶ **Explications formelles**
- ▶ Modèles à base d'arbres (DT, RF, BT)
- ▶ Théorie du domaine (connaissances), préférences utilisateur
- ▶ Correction du modèle (rectification)

# CRIL: développement de PyXAI

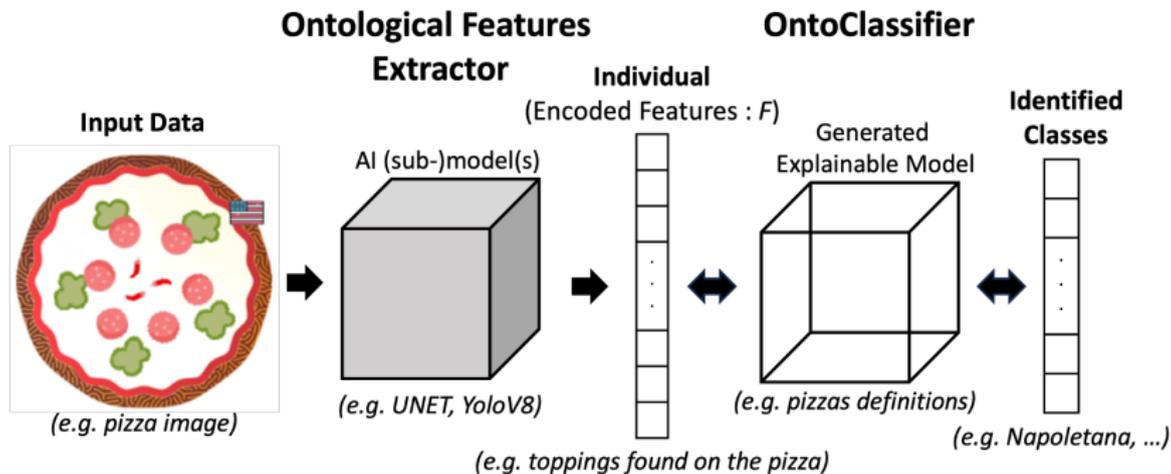


User interaction with PyXAI.

- **Explications formelles**
- Modèles à base d'arbres (DT, RF, BT)
- Théorie du domaine (connaissances), préférences utilisateur
- Correction du modèle (rectification)
- **Calcul d'explications formelles ... mais utilisant des exemples !**

[IJCAI'24]

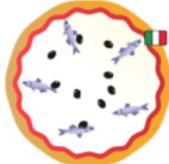
# LISIC : approche hybride pour une classification ontologiquement explicable



# OntoClassifier

## Multi Label Classification

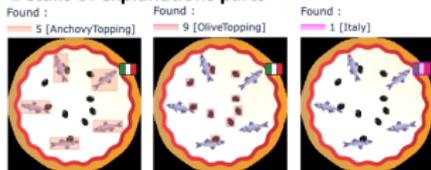
GenerousPizza  
Napoletana  
NonVegetarianPizza  
RealItalianPizza



## Raw Explanations

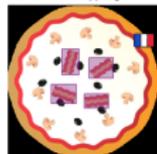
This is a Neapolitana because :  
hasTopping SOME AnchovyTopping is True  
found 5 AnchovyTopping  
hasTopping SOME OliveTopping is True  
found 9 OliveTopping  
hasTopping ONLY AnchovyTopping or OliveTopping is True  
hasCountryOfOrigin = Italy is True  
found Italy

## Details of explanations parts



GenerousPizza  
LaReine  
NonVegetarianPizza  
RealFrenchPizza

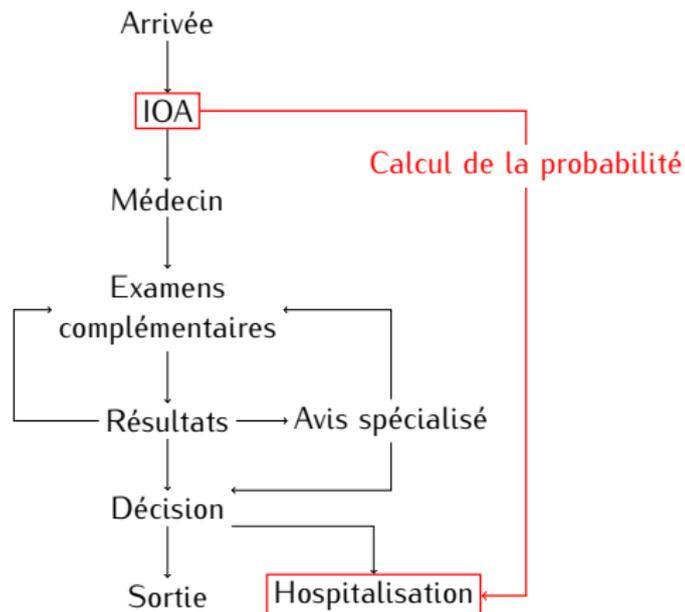
This is NOT a SpicyVegePizza because :  
This is NOT a VegetarianPizza because :  
hasTopping SOME MeatTopping is True  
found 4 HamTopping



4 [HamTopping]

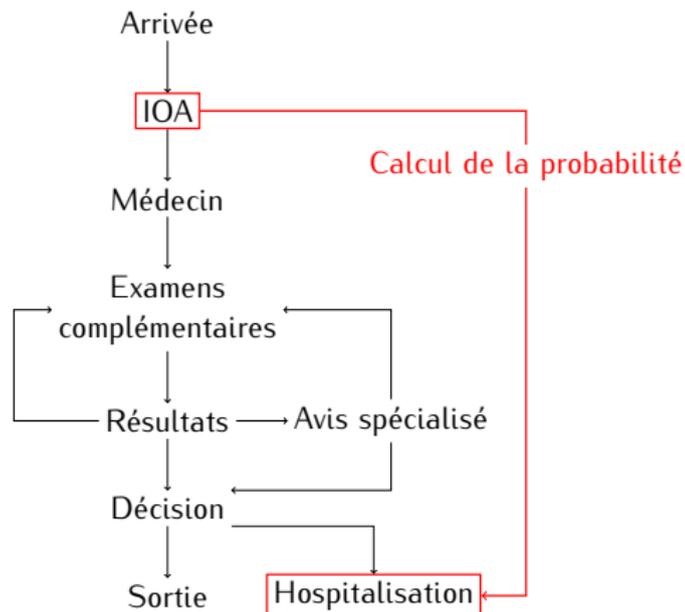
This is NOT a SpicyPizza because :  
hasTopping SOME SpicyTopping is False  
found nothing

# XAI – UPJV Laboratoire MIS

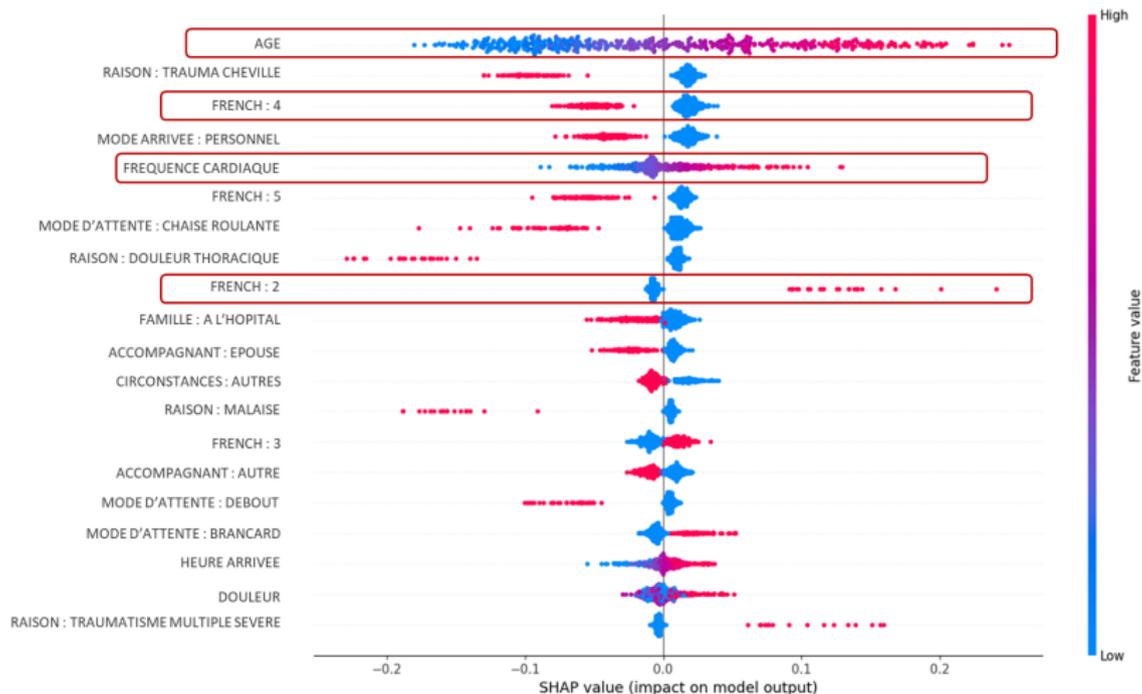




Probabilité  
d'hospitalisation brute  
n'est pas acceptable pour  
un médecin

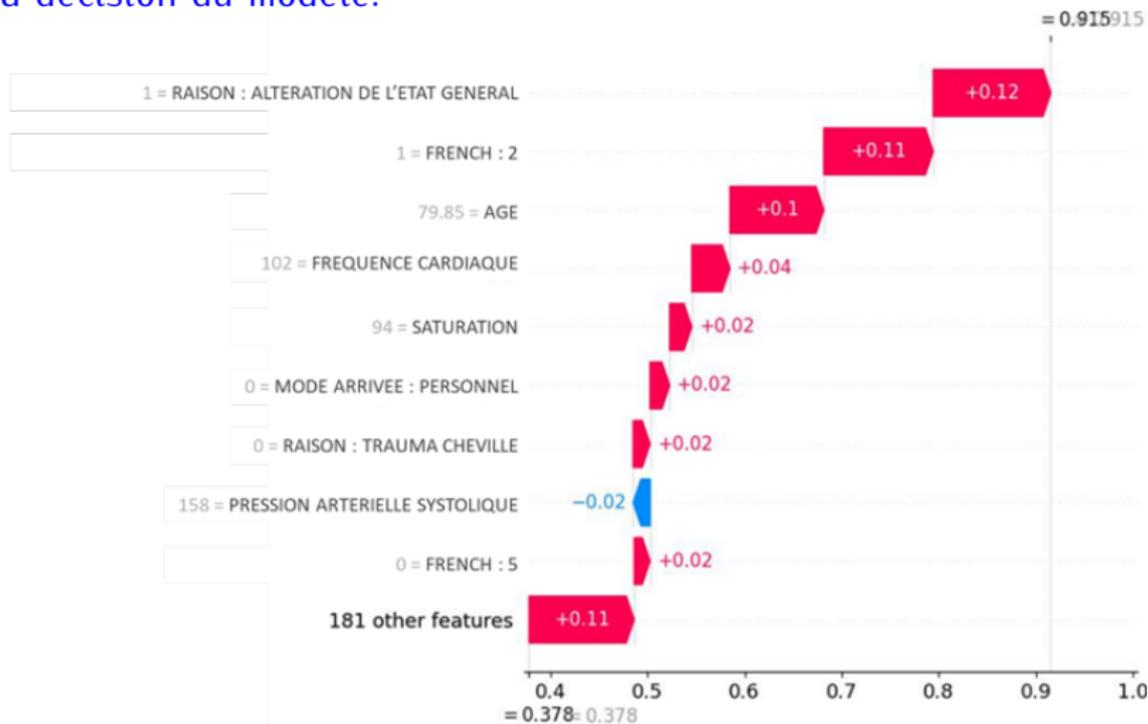


Fournir une explication globale sur l'impact des variables sur la décision permet de faire comprendre au médecin comment le modèle raisonne



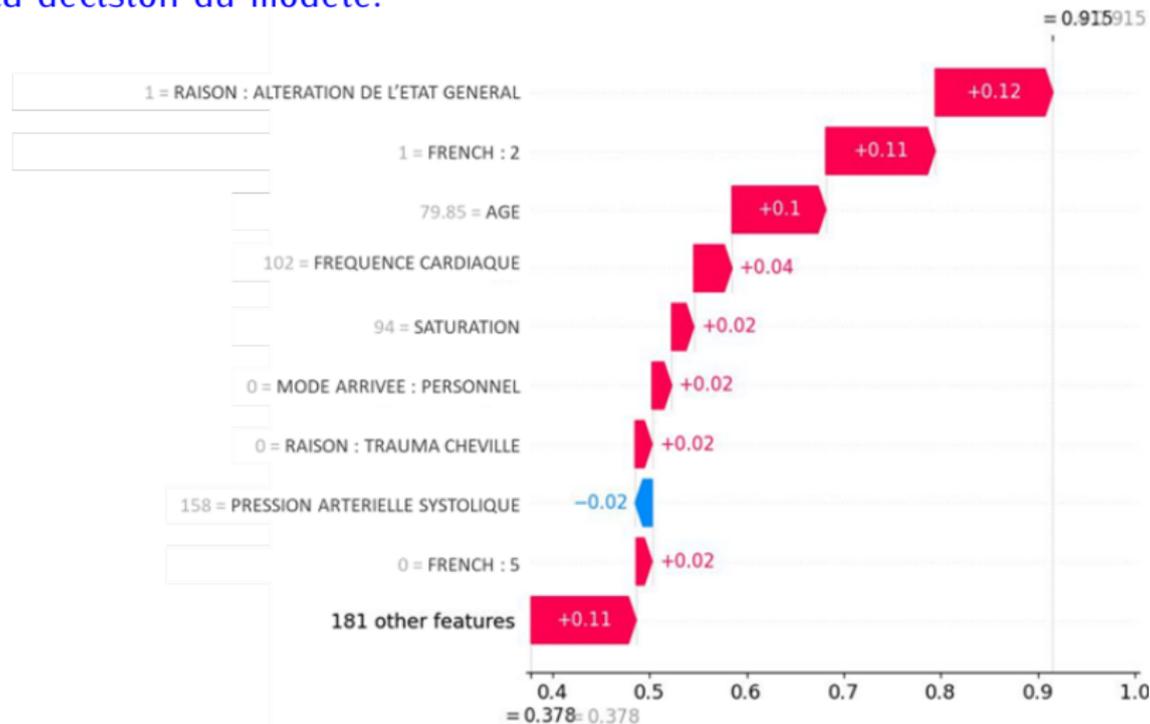
# XAI – MIS UPJV

Fournir une explication de l'instance pour justifier au médecin de la décision du modèle.



# XAI – MIS UPJV

Fournir une explication de l'instance pour justifier au médecin de la décision du modèle.



Problème : ce graphique reste peu lisible pour le corps médical.

# Perspectives de collaboration

- ▶ **IA neuro-symbolique : utiliser données et connaissances**
- ▶ Exploiter des connaissances pour **mieux prédire**
- ▶ Exploiter des connaissances pour **mieux expliquer** les prédictions
- ▶ **Plusieurs combinaisons** données / connaissances à étudier
- ▶ Gestion de l'incertitude et décision
- ▶ Choisir le modèle ML / ses paramètres / les attributs à considérer selon le contexte en exploitant des connaissances

## *We need you!*

- ▶ Besoin de **connaissances**, d'expertise liées à des applications spécifiques pour les **injecter dans nos IA**
- ▶ Comment mesurer la **qualité des explications du point de vue de l'utilisateur ?**
- ▶ **Venez travailler avec nous !**