

# Explicabilité de l'intelligence artificielle : de l'importance de la contextualisation

**Océane Fiant**

Post-doctorante

Costech, Université de Technologie de Compiègne

[oceane.fiant@utc.fr](mailto:oceane.fiant@utc.fr)

# Introduction

---

## **Exigences légales pour les intelligences artificielles destinées à un usage médical :**

- Explicabilité (RGPD, 2016).
- Sûreté et performances relativement à l'usage attendu (règlement européen 2017/745 relatif aux dispositifs médicaux, 2021)

## **Très peu de dispositifs satisfont ces exigences :**

- Opacité liée à la complexité des données et des modèles, mais également au secret industriel (Burrell 2016).
- Biais liés aux données, par exemple à des variables confondantes (Winkler et al. 2019).



# Introduction

---

## **Comment fiabiliser ces dispositifs ?**

- Agir sur les données d'entraînement : mettre à disposition des concepteurs de grandes quantités de données étiquetées ou non (Health Data Hub), mettre l'accent sur leur préparation, etc.
- Agir sur l'opacité : rendre les modèles explicables pour pouvoir prévenir et corriger leurs défaillances.

L'explicabilité est une nécessité technique, légale et éthique, ainsi qu'un préalable à l'adoption de ces dispositifs en médecine (confiance).

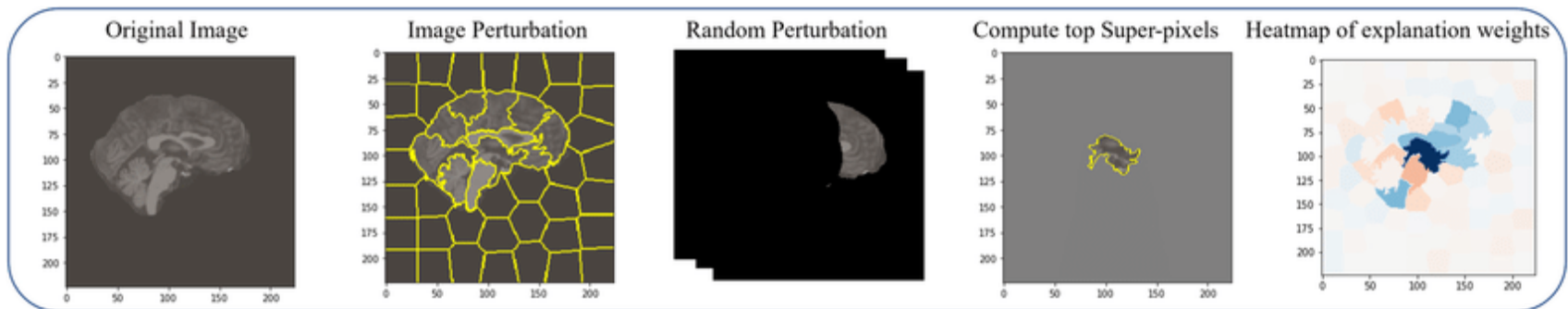


# L'explainable artificial intelligence

---

## Deux catégories d'approches :

- *Ante-hoc* : modèles explicables par nature (par ex. régression linéaire).
- *Post-hoc* : méthodes permettant de parvenir à une explication d'un modèle opaque. Par exemple, *local interpretable model-agnostic explanations* (Ribeiro *et al.* 2016) :



(Aghaei *et al.* 2022)



# Limites pour l'utilisateur

---

## Les explications proposées font-elles sens pour le médecin ?

- Les approches *ante-hoc* supposent que les médecins ont la compétence technique pour comprendre, voire corriger un modèle explicable par nature.
  - Un dispositif médical n'est utilisé que dans la mesure où il s'intègre aux pratiques médicales existantes et apporte à un bénéfice à la prise en charge des patients.
- Les approches *post-hoc* supposent que les médecins intègrent à leur décision des éléments dont les mécanismes causaux leur échappent et que la confiance qu'ils accordent à ces éléments provient d'une compréhension limitée de ces mécanismes causaux.
  - Les médecins utilisent déjà des boîtes noires validées par des essais randomisés contrôlés et des méta-analyses.

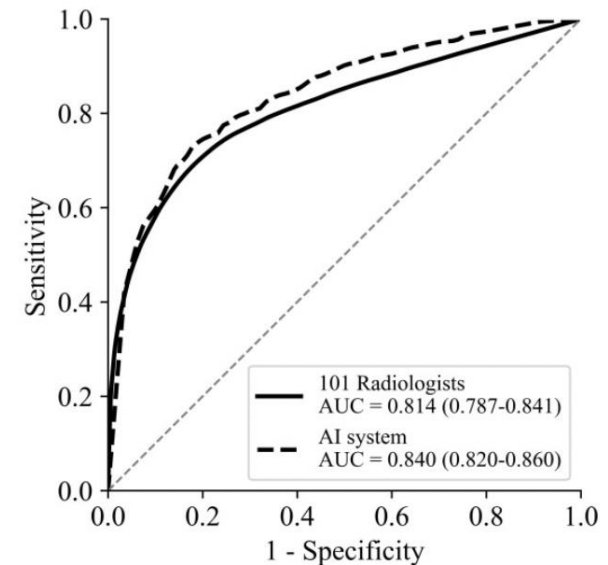


# Implication des utilisateurs

---

**L'approche dominante en matière de conception semble être une approche en termes de supplantation du médecin dans certaines de ses tâches :**

- Prépondérance dans la littérature de références à la supériorité des IA par rapport aux médecins en termes de performances/rapidité (Liu *et al.* 2019). Il ne faut toutefois pas prendre ces résultats au pied de la lettre (Nagendran *et al.* 2020).
- Si l'IA surpasse le médecin dans certaines tâches, il semble rationnel et donc pleinement légitime de déléguer à l'IA ces tâches pour lesquelles elle est plus performante.



(Rodriguez-Ruiz et al. 2019)



# Inadaptation aux pratiques et contextes effectifs

---

Les médecins ont rarement voix au chapitre en ce qui concerne la conception des IA qui leur sont destinées. Les pratiques et contextes effectifs sont alors peu pris en compte.

## **Exemple de Watson for Oncology (Tupalesa et Di Nucci 2020) :**

Un taux de concordance entre les recommandations de Watson et les protocoles locaux atteignant parfois 96 %, mais de l'ordre de 30 % à Copenhague :

- Watson a été développé à partir des recommandations américaines et ignore les recommandations nationales et des préférences locales.
- Watson accorde beaucoup d'importance aux données d'essais cliniques américains, qui ne sont pas représentatifs des populations danoises. Pour les oncologues Danois, les essais américains sont de moins bonne qualité que les essais danois.



# Conclusion

---

La validité des dispositifs d'IA est liée aux contextes, il est donc important que les dispositifs destinés à un usage médical aient une validité externe étendue (différentes populations, différents contextes).

Plus fondamentalement, la conception des IA devrait tenir compte des contextes dans lesquels ces dispositifs ont vocation à s'insérer, ce qui implique une connaissance fine de ces contextes par les concepteurs de ces dispositifs.



Merci de votre attention.

**Océane Fiant**

Post-doctorante

Costech, Université de Technologie de Compiègne

[oceane.fiant@utc.fr](mailto:oceane.fiant@utc.fr)