ECUE21.2 Science des données

Bonnes pratiques

Chloé-Agathe Azencott

Center for Computational Biology (CBIO)
Mines Paris PSL – Institut Curie – INSERM U900
PSL Research University & PR[AI]RIE, Paris, France

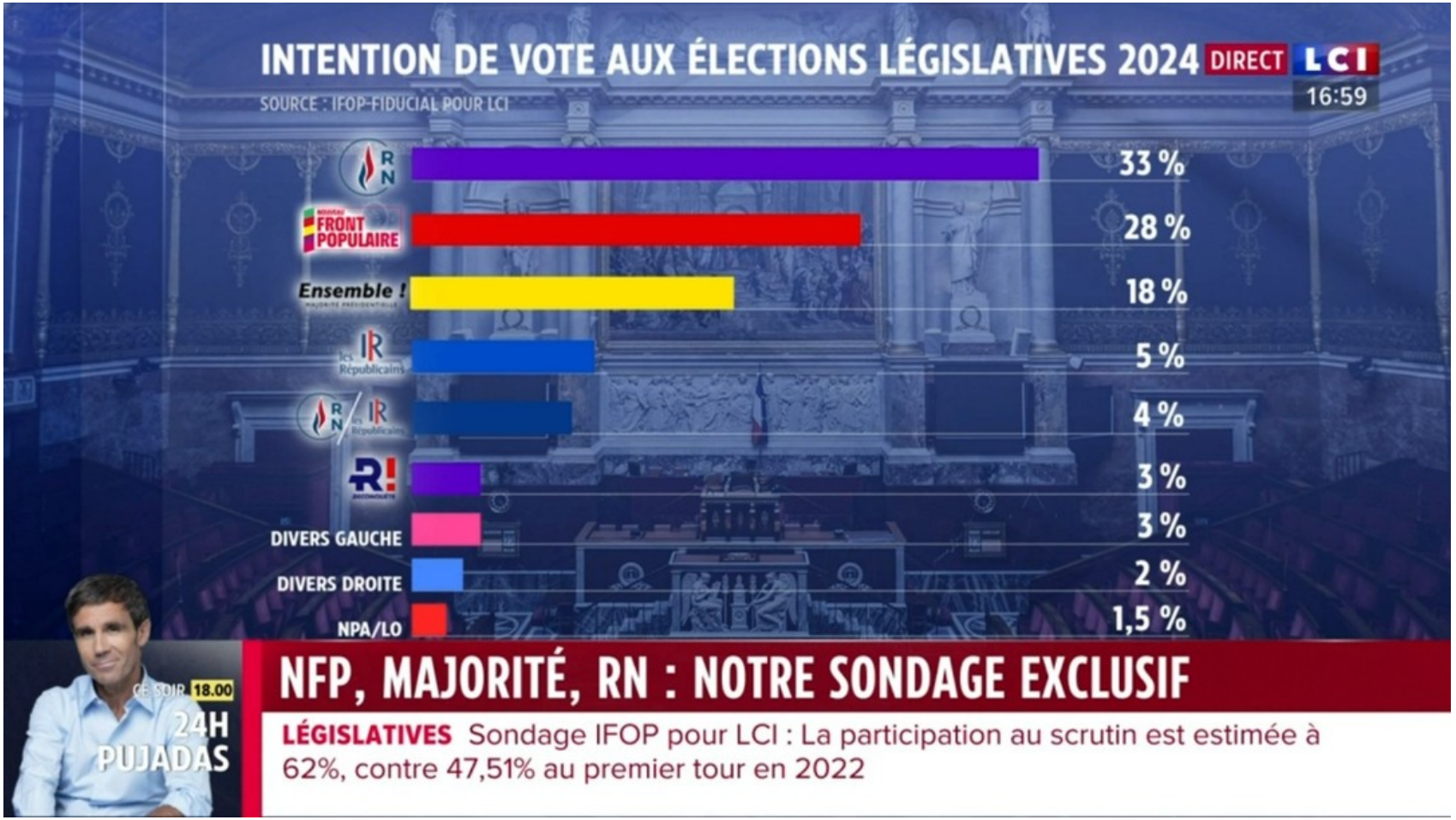
Juin 2024

http://cazencott.info chloe-agathe.azencott@minesparis.psl.eu @cazencott@lipn.info

Dataviz

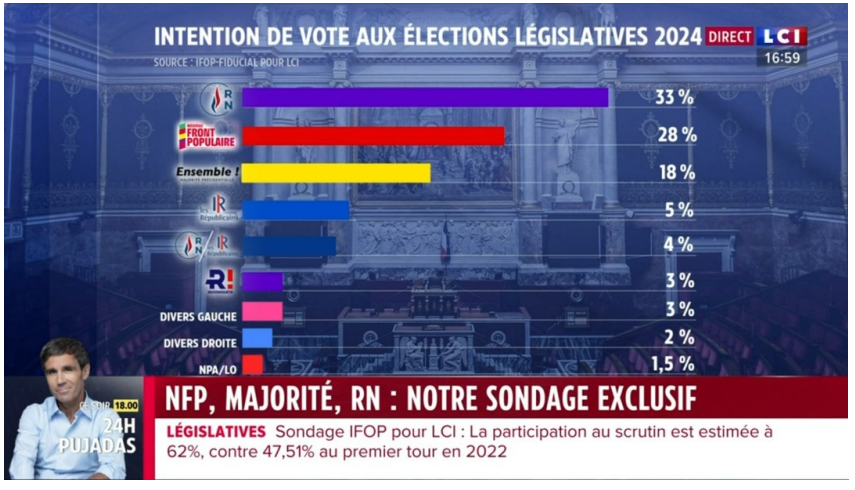
1. Visualisation de données

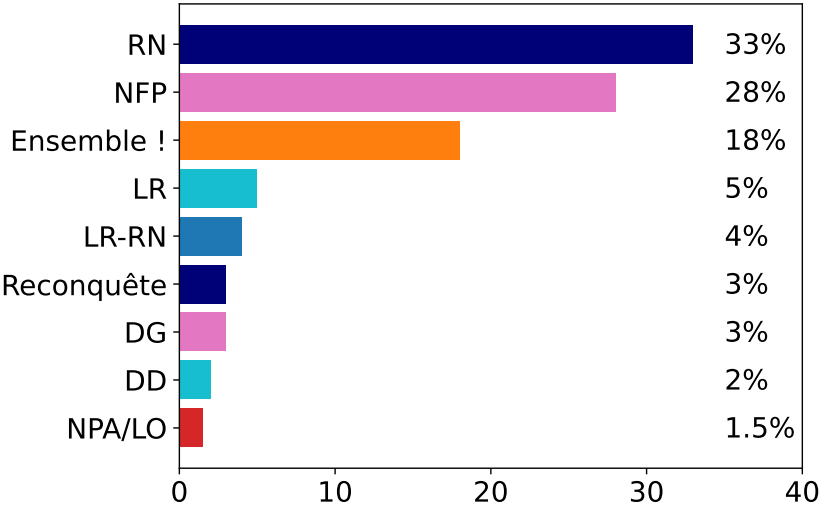
1. Exemple 1



TF1, 17 juin 2024, 16h56

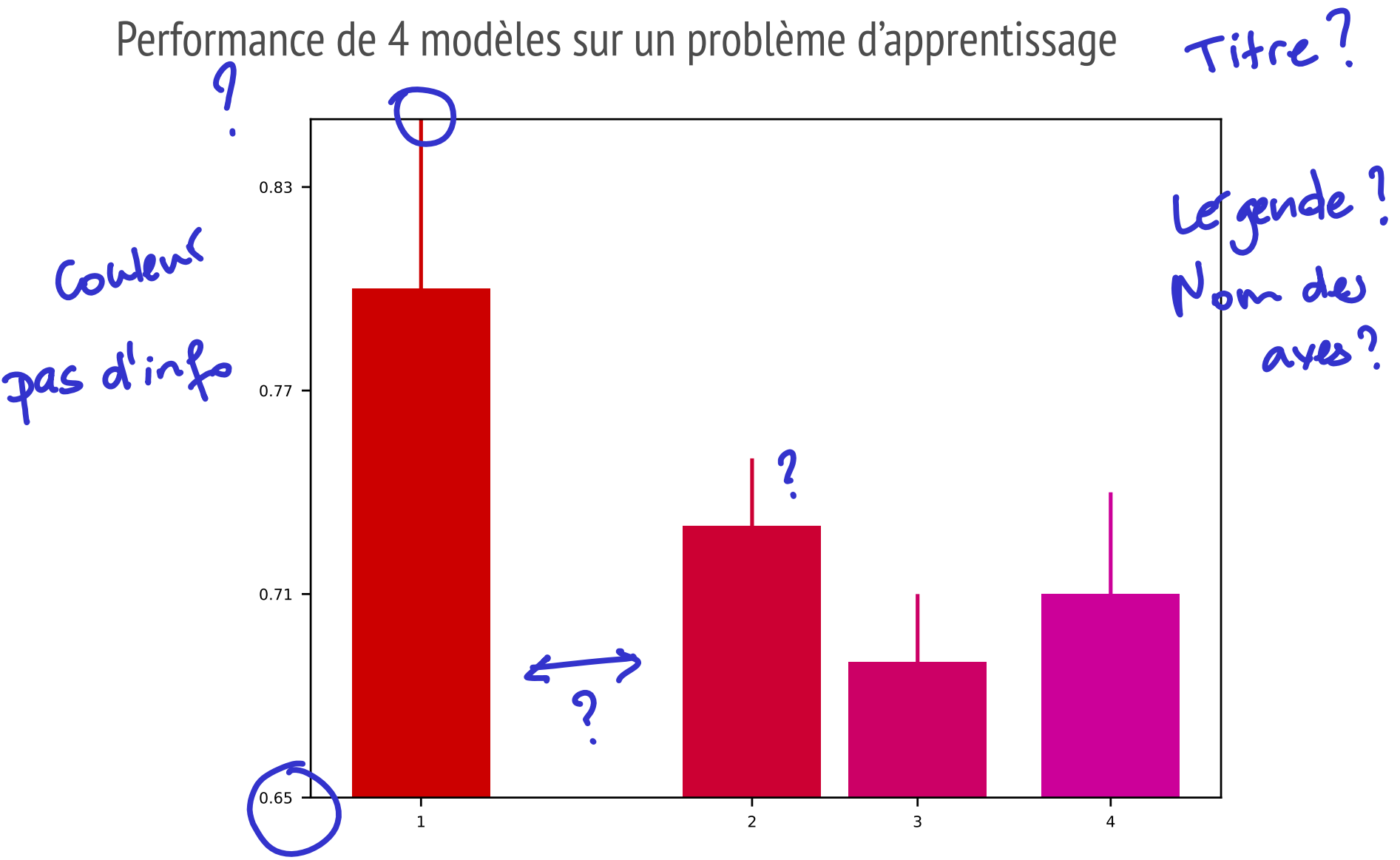
1. Exemple 1


INTENTION DE VOTE AUX ÉLECTIONS LÉGISLATIVES 2024
DIRECT
LCI
16:59
SOURCE : IFOP-FIDUCIAL POUR ICI
33 %
28 %
18 %
5 %
4 %
3 %
3 %
2 %
1,5 %
DIVERS GAUCHE
DIVERS DROITE
NPA/LO
NFP, MAJORITÉ, RN : NOTRE SONDAGE EXCLUSIF
LÉGISLATIVES Sondage IFOP pour LCI : La participation au scrutin est estimée à 62%, contre 47,51% au premier tour en 2022
18.00
24H PUJADAS


RN
NFP
Ensemble !
LR
LR-RN
Reconquête
DG
DD
NPA/LO
33%
28%
18%
5%
4%
3%
3%
2%
1.5%
0
10
20
30
40

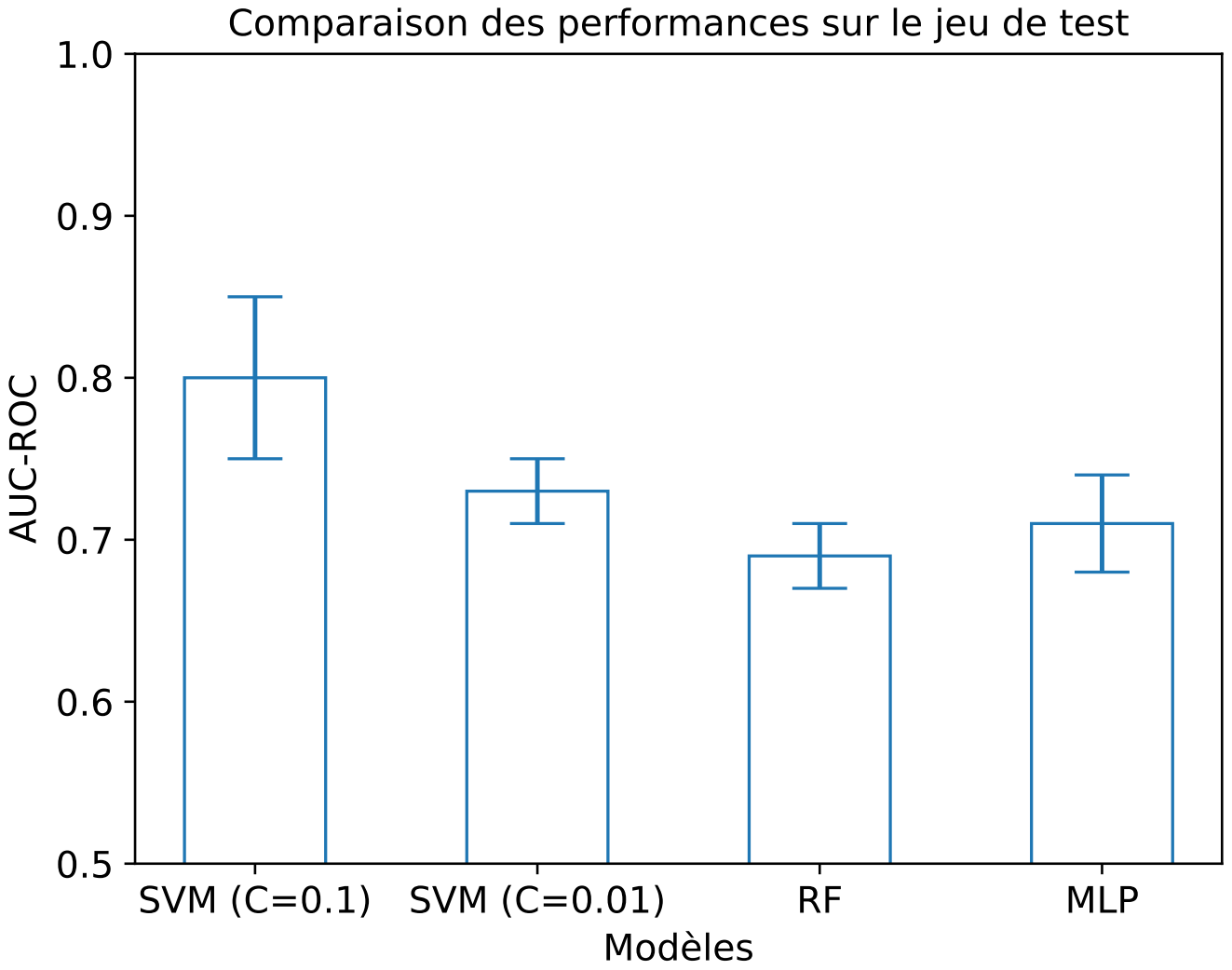
2. Exemple 2

Performance de 4 modèles sur un problème d'apprentissage

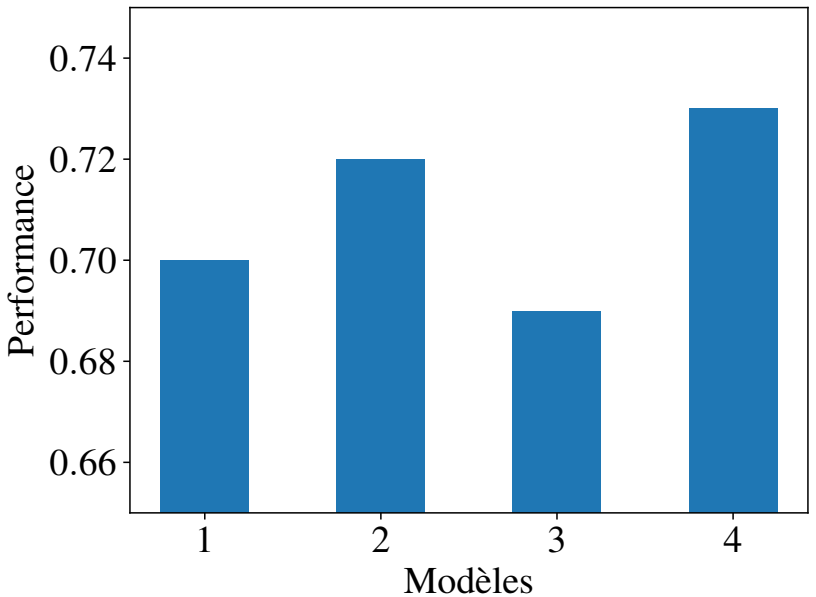


2. Exemple 2

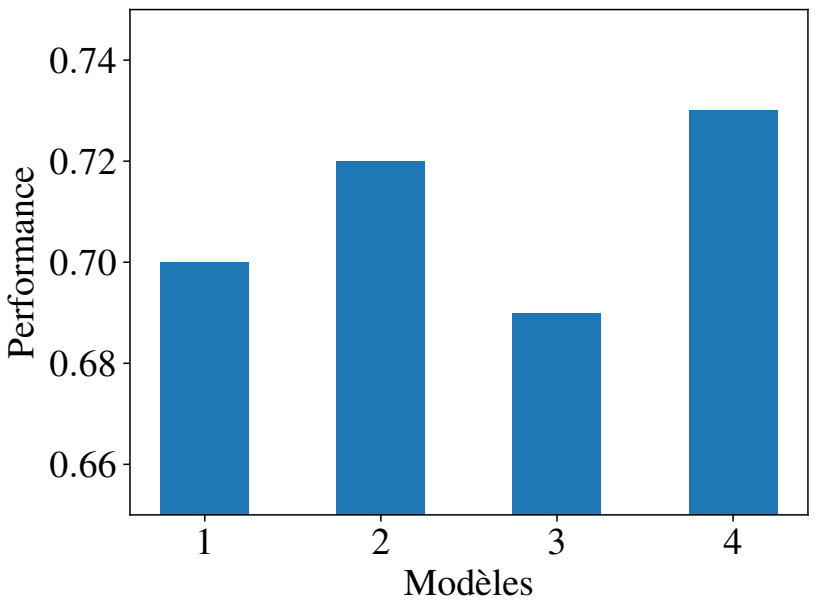
Performance de 4 modèles sur un problème d'apprentissage

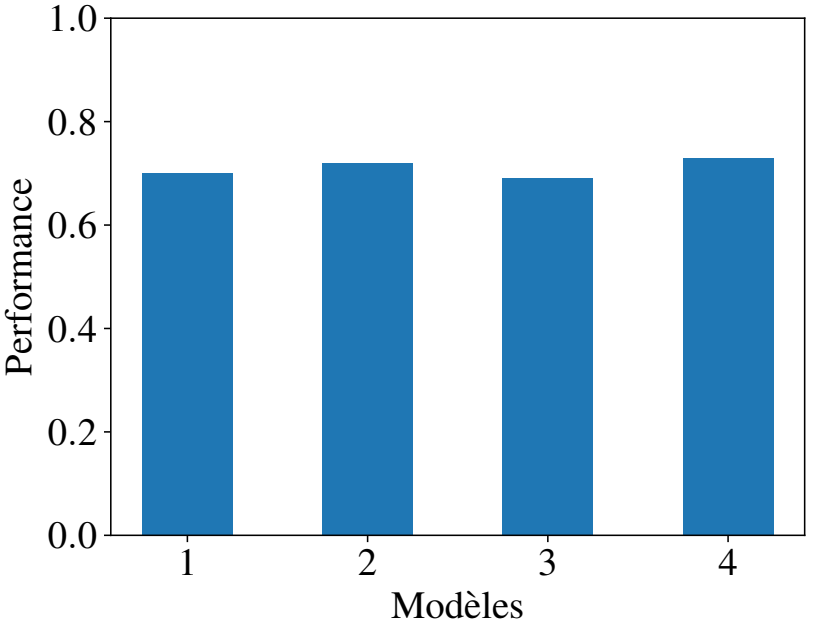


3. Choix des axes (1)

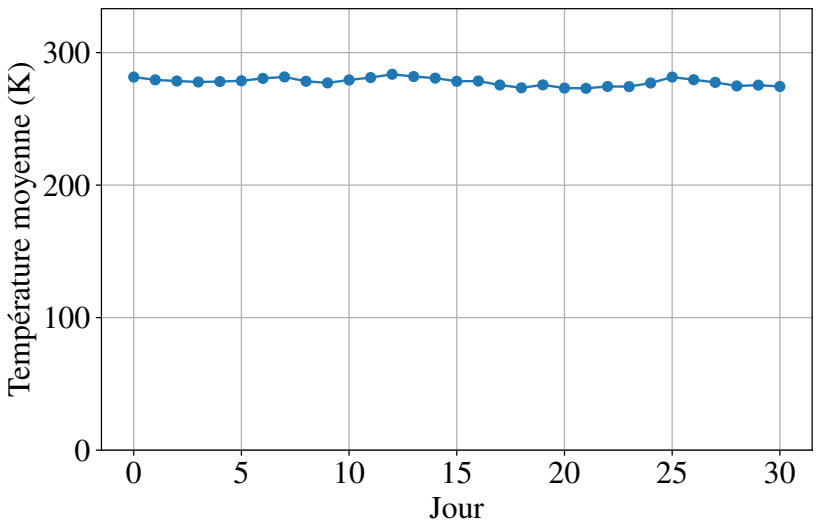


3. Choix des axes (1)

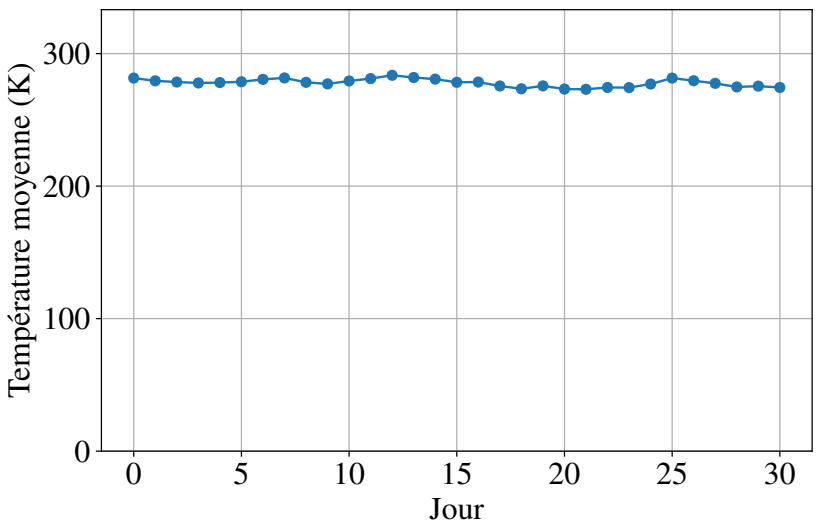


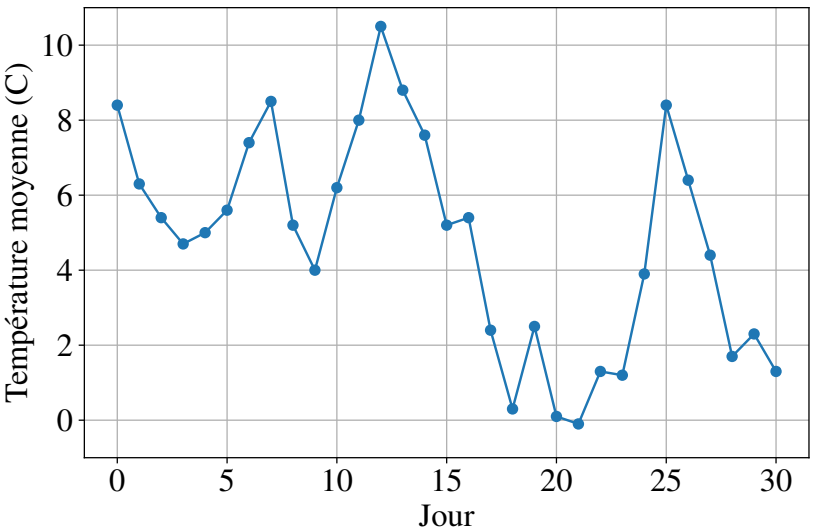


3. Choix des axes (2)

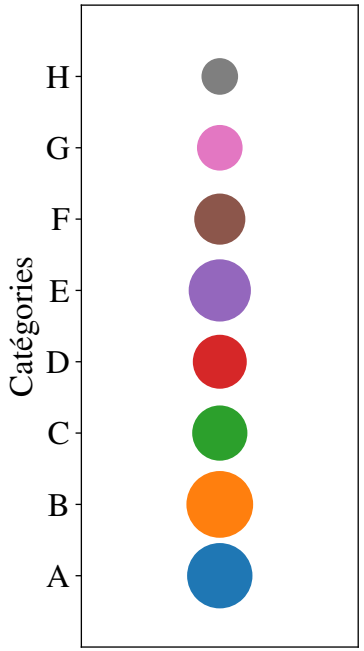


3. Choix des axes (2)

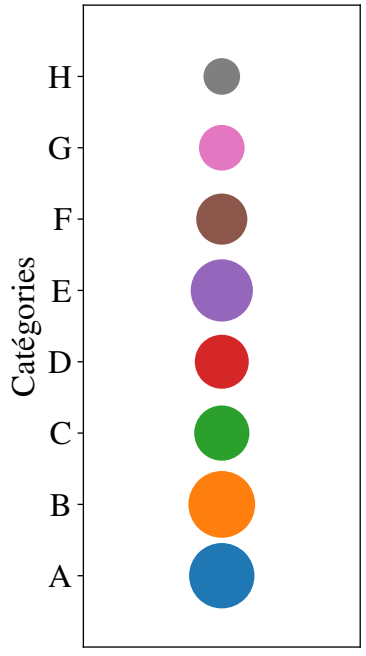


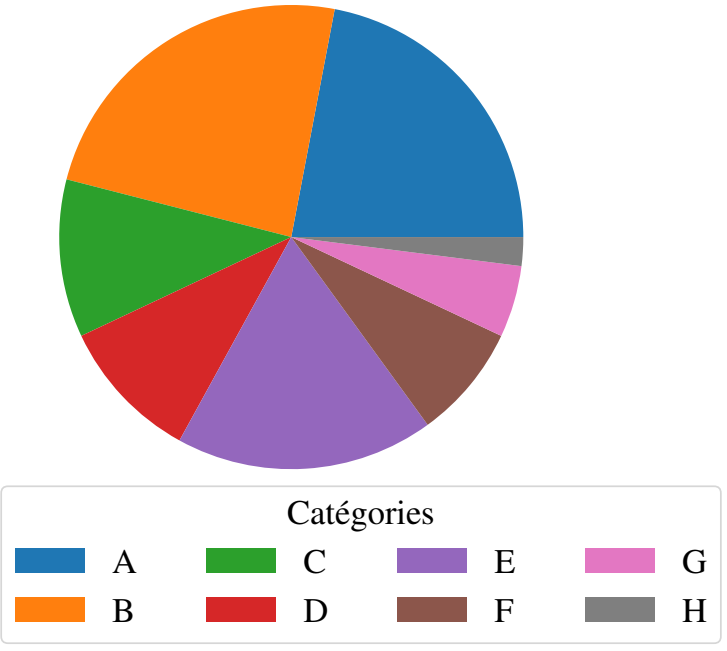


4. Proportional ink

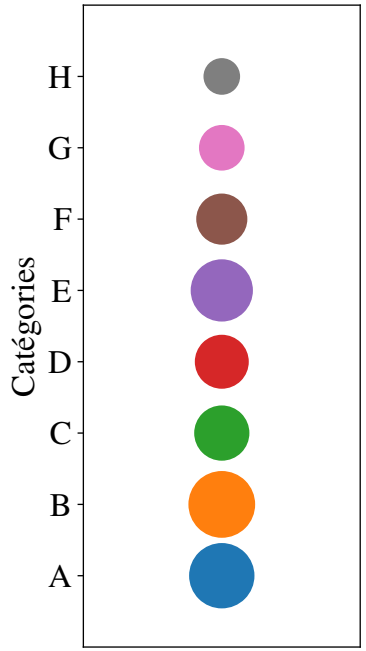


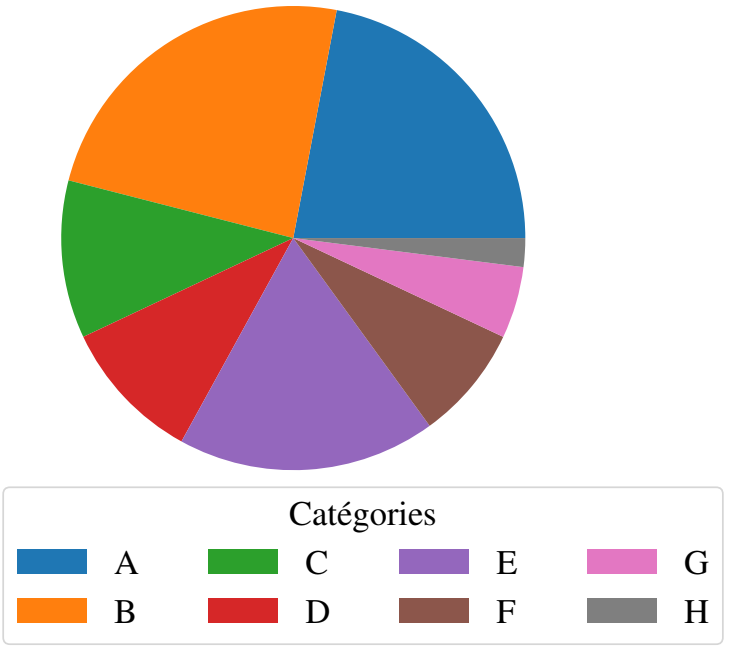
4. Proportional ink

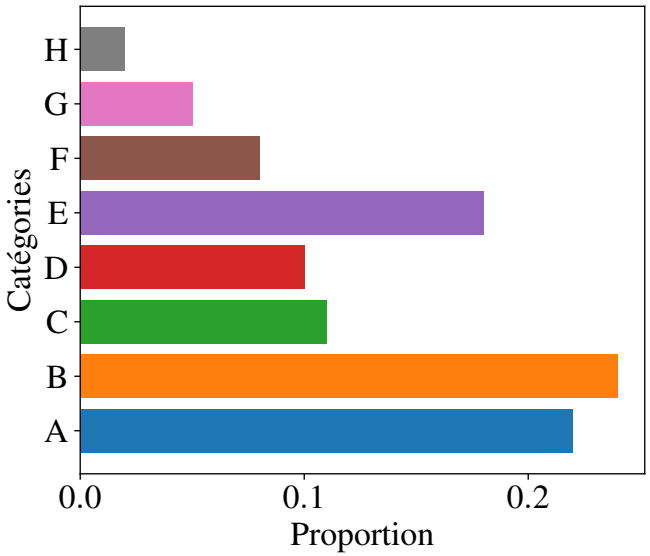




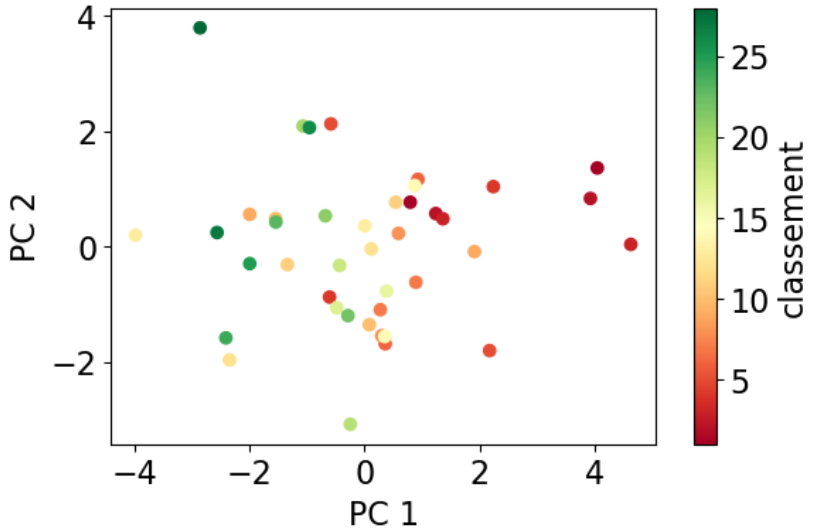
4. Proportional ink





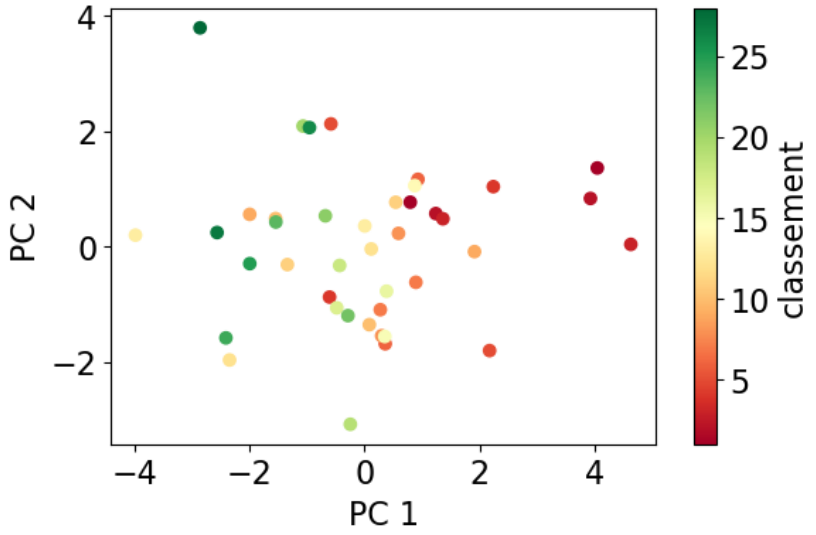


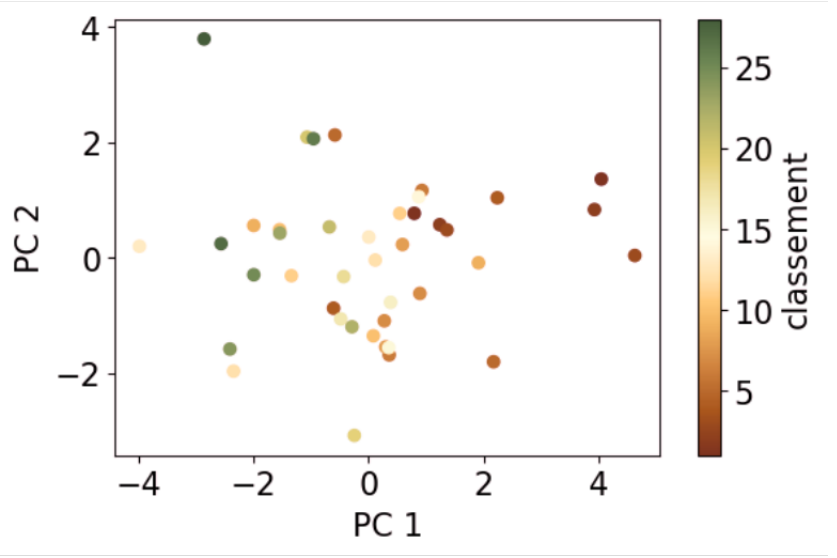
5. Dyschromatopie




```
plt.scatter(...cmap='RdYlGn')
```

5. Dyschromatopie



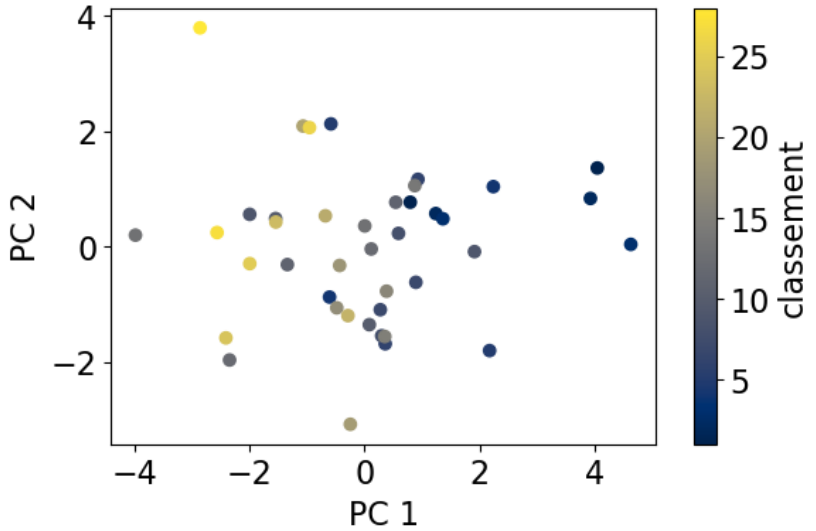



```
plt.scatter(...cmap='RdYlGn')
```

Simulation de deutéranopie par CoBliS

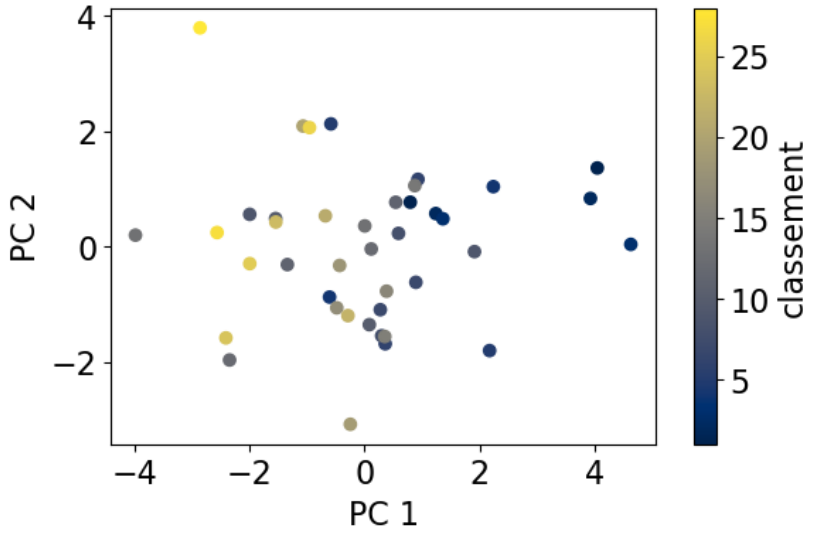
[lien vers CoBliS]

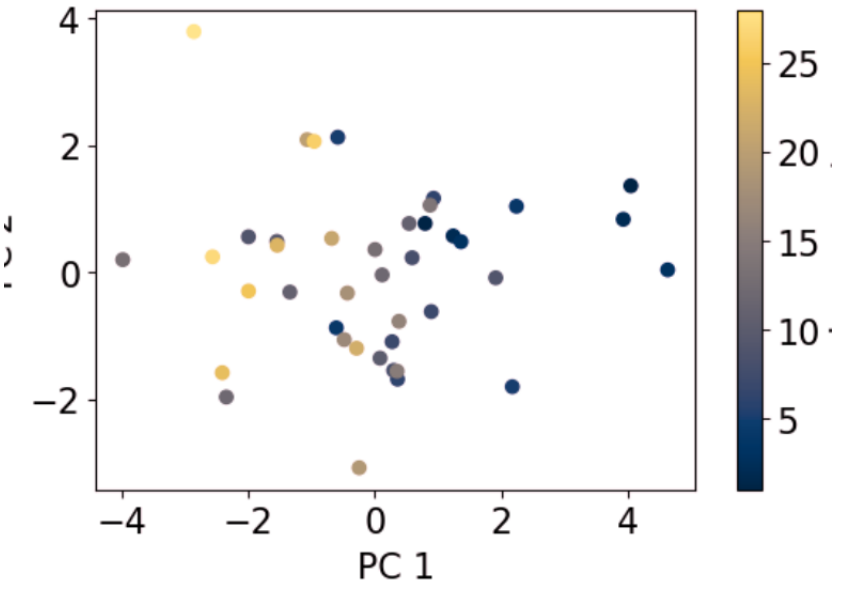
5. Dyschromatopie




```
plt.scatter(...cmap='cividis)
```

5. Dyschromatopie





`plt.scatter(...cmap='cividis)`

Simulation de deutéranopie par CoBliS

[lien vers CoBliS]

2. Questionnements autour de l'utilisation du ML

① Veut-on vraiment résoudre ce problème?

AI Gaydar

2.1 Quel problème ?

② Le ML est-il la meilleure approche?

- des outils existants savent résoudre le pb

- peu de données, peu de connaissance métier...

Exemple : Détection de criminels

- Article sur arxiv : *Automated Inference on Criminality using Face Images*, Xiaolin Wu & Xi Zhang (2017)
- Motivation : "Unlike a human examiner/judge, a computer vision classifier has absolutely no subjective baggage, having no emotions, no biases whatsoever due to past experience, race, religion, political doctrine, gender, age, etc."

Biais / Représentativité

2.2 Quelles donnéees ?

Exemple : Détection de criminels

– Article sur arxiv : *Automated Inference on Criminality using Face Images*, Xiaolin Wu & Xi Zhang (2017)

(a) Three samples in criminal ID photo set S_c.

(b) Three samples in non-criminal ID photo set S_n

Figure 1. Sample ID photos in our data set.

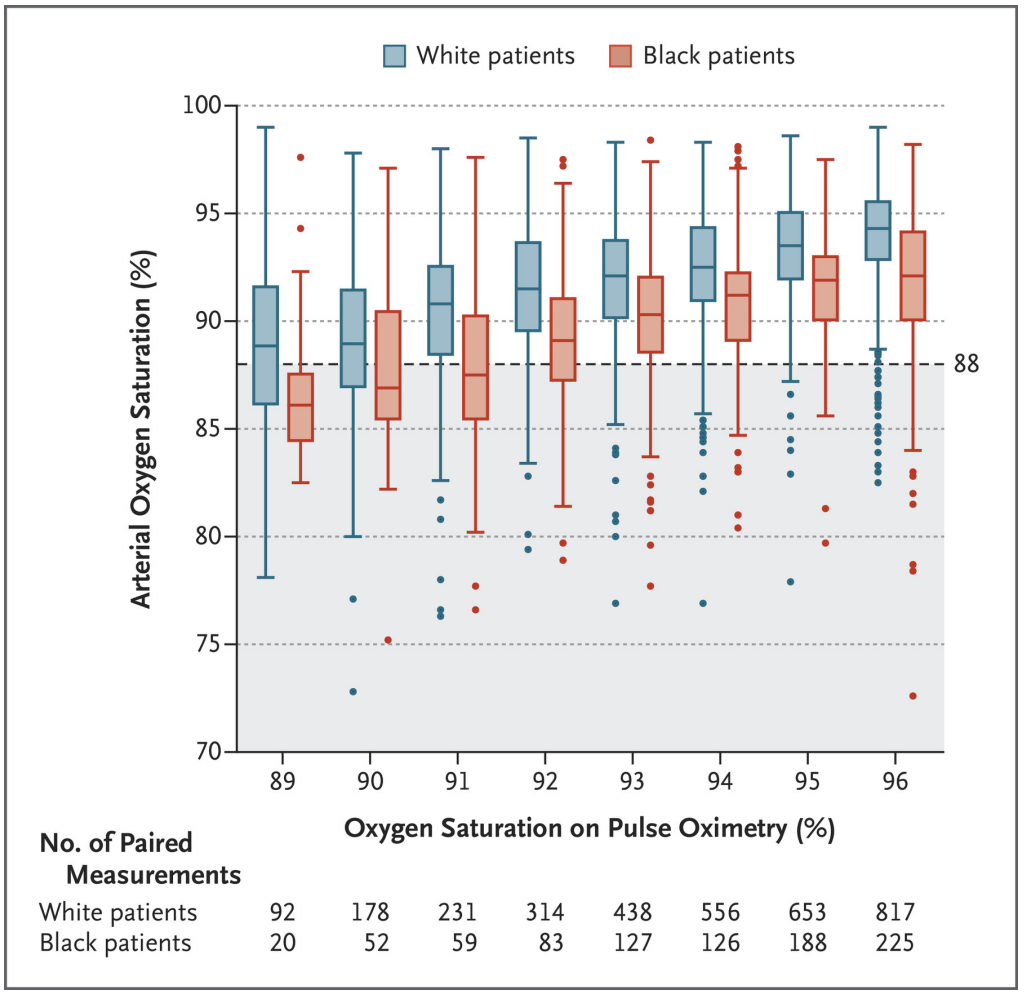
Exemple : Recrutement automatisé

source: Reuters [lien]

- Système fortement biaisé en faveur des CV déposés par des hommes
- Pourtant cette information ne faisait pas partie des variables utilisées

It's not just AI : oxymètres de pouls

– *Racial Bias in Pulse Oximetry Measurement*, Sjoding et al., New England Journal of Medicine, 2020; 383:2477-2478 [lien]



Accuracy of Pulse Oximetry in Measuring Arterial Oxygen Saturation, According to Race.

Acquisition des données

Consentement ?

Confidentialité ? Déidentification algorithmique

technique
- Pseudonymisation
- Differential privacy
- Sécurité des bases de données

réglementaires : protection des données personnelles (ex: RGPD)

"Travail du clic"

Vérifiabilité : garantir qu'un système/algorithme a le comportement attendu.

(preuves formelles)

très peu en ML

spécification

2.3

~~4~~. Fiabilité

Explicabilité cf PC4

→ explainable AI XAI

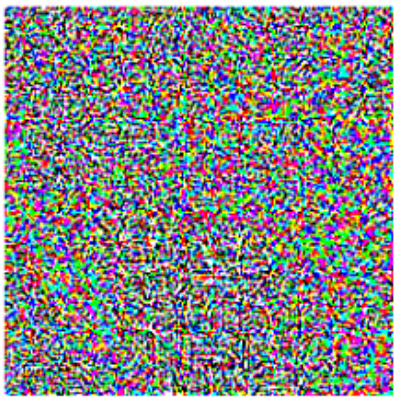
Robustesse

Attaque (bruit gaussien)

$\boldsymbol{x}$

"panda"
57.7% confidence

$+ .007 \times$



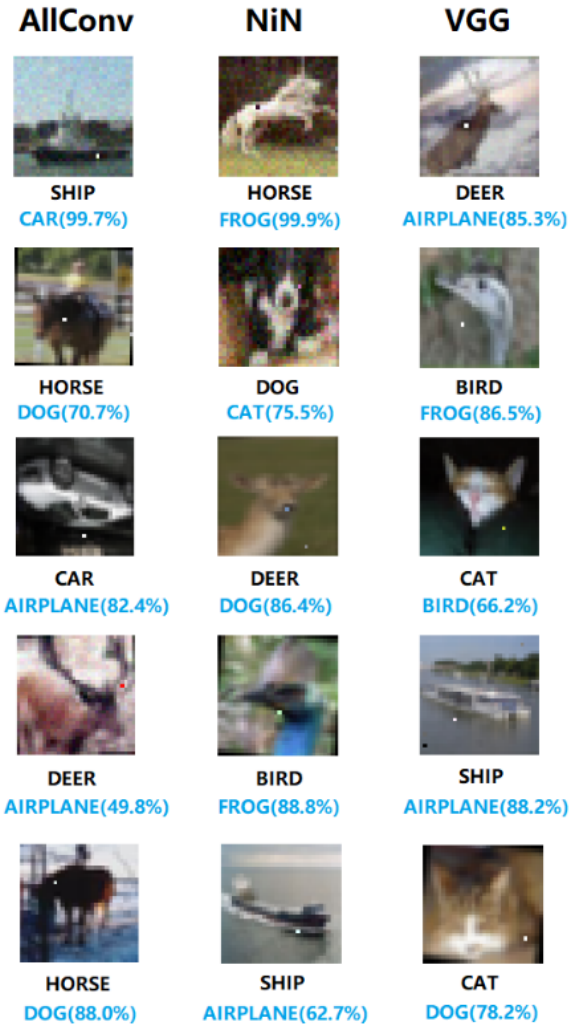
$\text{sign}(\nabla_{\boldsymbol{x}} J(\boldsymbol{\theta}, \boldsymbol{x}, y))$

$=$

$\boldsymbol{x} + \epsilon\text{sign}(\nabla_{\boldsymbol{x}} J(\boldsymbol{\theta}, \boldsymbol{x}, y))$
"gibbon"
99.3 % confidence

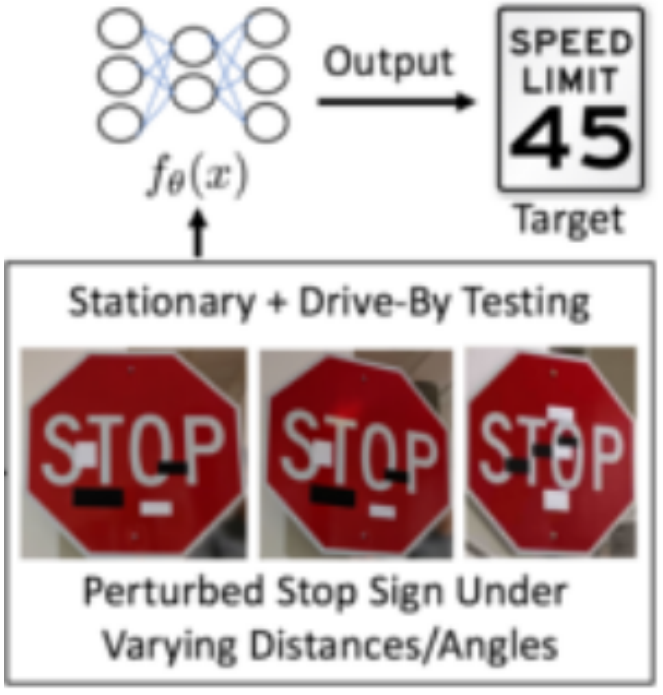
Goodfellow, Shlens & Szegedy (ICLR 2015)

Attaque (1-pixel)



Su, Vargas & Kouichi (IEEE Transactions on Evolutionary Computation 2019)

Attaque (monde réel)



Eykholt et al. (CBPR 2018)

EchoLeak (juin 2025)

source: AIM Labs (lien)

- **Injection:** Attacker sends innocuous-looking email that includes LLM scope violation exploit
- **Action:** User asks Copilot a question
- **Scope Violation:** Copilot mixes attacked input with sensitive data
- **Retrieval:** Copilot leaks sensitive data to attacker via SharePoint URLs

Empreinte écologique du numérique : 4.4% empreinte carbone (2024)

Strubell et al. (2020)
213×10^6 paramètres
entraînement : $\simeq$ 300 t eq CO_2

5. Ressources

Luccioni (2023) BLOOM
176×10^9 paramètres $\simeq$ 50 t eq CO_2
(mix énergétique moins carboné).