

LOYAUTÉ ET INTERPRÉTABILITÉ DES ALGORITHMES

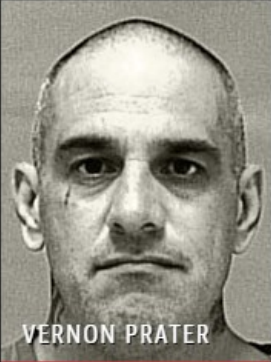

Paris – 11 Novembre 2019
Conférence « Justice et Numérique »

Ronan PONS – Doctorant de la chaire *Law, Accountability and Fairness*
au sein d'ANITI (3IA Toulouse)

Discrimination et *Machine learning* : Le problème des biais algorithmiques

COMPAS

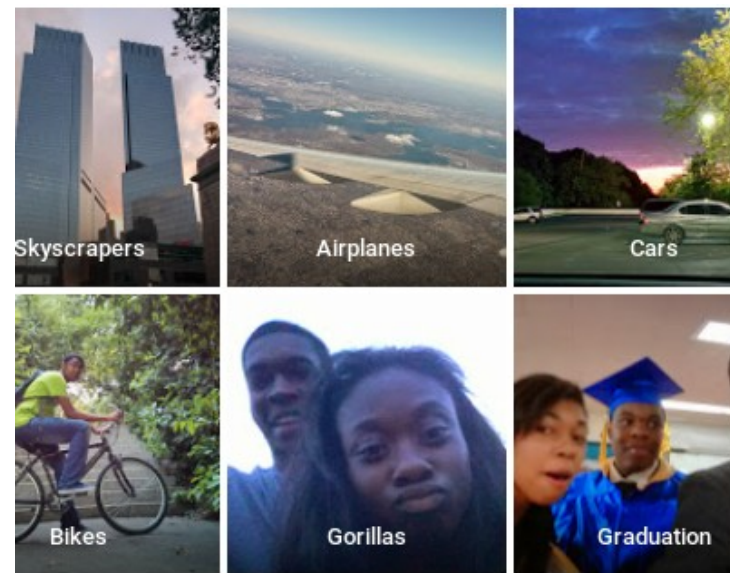
Two Petty Theft Arrests

 VERNON PRATER	 BRISHA BORDEN
LOW RISK 3	HIGH RISK 8

Borden was rated high risk for future crime after she and a friend took a kid's bike and scooter that were sitting outside. She did not reoffend.

Source : *Propublica*

GOOGLE PHOTOS



Source : *The Verge*

I. L'encadrement réglementaire des décisions automatisées / algorithmiques



Le Règlement Général à la Protection des Données (RGPD) : le droit de s'opposer

- Article 22 RGPD : Le principe

« La personne concernée a le **droit de ne pas faire l'objet d'une décision fondée exclusivement sur un traitement automatisé**, y compris le profilage, produisant des **effets juridiques** la concernant ou l'affectant de **manière significative de façon similaire.** »

- Conditions d'application :

Secteurs concernés



Public



Privé

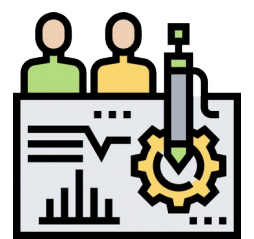
Criticité



Acteurs présents



Données traitées





La Loi Informatique et Libertés (LIL) :

l'interdiction conditionnelle des décisions automatisées

- Article 47 al. 2 LIL : le principe

« Aucune décision produisant des **effets juridiques** à l'égard d'une personne ou l'affectant de **manière significative** ne peut être prise **sur le seul fondement** d'un **traitement automatisé de données à caractère personnel**, y compris le **profilage** »

- Les exceptions (soumises à garantie)

- 1 Le consentement de l'individu
- 2 La nécessité contractuelle
- 3 Les exceptions nationales

La Loi pour une République Numérique : l'exception des décisions administratives

- Article 47 al. 2. 2° LIL : les exceptions (soumises à garanties)

« Des **décisions administratives individuelles** prises dans le respect de l'article L. 311-3-1 et chapitre 1^{er} du titre Ier du livre IV du CRPA, à condition que le traitement ne porte pas sur des données mentionnées au I de l'article 6 de la présente loi. Ces décisions comportent, à peine de nullité, la mention explicite prévue à l'article L. 311-3-1 du CRPA. »

- Conditions d'application :

Secteur concerné

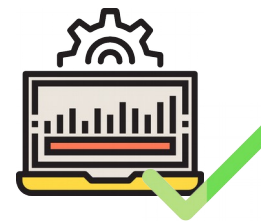


Administrations

Criticité



**Acteurs
présents**



**Données
concernées**



II.L'arsenal réglementaire actuel



Le Règlement Général à la Protection des Données personnelles (RGPD) : Obligation d'information et Droit d'accès

- Articles 13.2.f), 14.2.g) et 15.1.h)

« [...] l'existence d'une prise de décision automatisée, [...] et, des **informations utiles** concernant la **logique sous-jacente**, ainsi que l'importance et les **conséquences prévues de ce traitement pour la personne concernée.** »

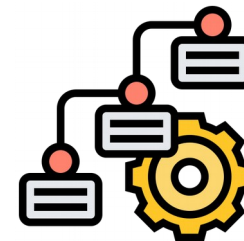
Moment de la communication



Acteurs concernés




Contenu de l'information



Logique de l'algorithme



Conséquences pour l'individu



La Loi pour une République Numérique : La publication du fonctionnement global


- Pour les administrations : Le code-source

```
11     def BuildGrid(self):
12         """ This method builds the grid with the values of 'row' and 'col' as dimensions """
13         for line in range(self.row):
14             self.grid.append([]) # each line is a list
15             for column in range(self.col):
16                 self.grid[line].append((line, column)) # fills the grid with cell coordinates
17
```

Article L. 300-2 CRPA

- Pour les plateformes en ligne : Les règles
 - Les conditions de référencement et déréférencement
 - Les critères de classement des contenus et leurs paramètres
 - La présence d'un lien capitalistique

Article L. 111-7 et D. 111-7 du code de la consommation



La Loi pour une République Numérique : L'explication d'une décision individuelle

- Article L. 311-3-1 CRPA :

« [...] une décision prise sur le fondement d'un traitement d'un traitement algorithmique comporte une **mention explicite** en informant l'intéressé.

Les **règles** définissant ce traitement ainsi que les **principales caractéristiques** de sa mise en œuvre sont **communiquées à l'intéressé s'il en fait la demande.** »

- Principales caractéristiques :

1 Degré et mode de contribution à la décision

2 Données traitées et leurs sources

3 Paramètres de traitement et leur pondération

4 Opérations effectuées dans le traitement



Projet de loi relatif à la bioéthique : la protection par l'intervention humaine

Article L. 4001-3 du Code de la santé publique :

« [...] II. L'**adaptation des paramètres** d'un traitement mentionné au I pour des actions à visée préventive, diagnostique ou thérapeutique concernant une personne est réalisée avec **l'intervention d'un professionnel de santé** et peut être **modifiée** par celui-ci. [...] »



- Qui est-ce ?
- Quel rôle joue-t-il ?
- Quelles compétences techniques possède-t-il ?
- Quelle efficacité ?

III. Challenges, limites et pistes de solutions

La question de la preuve de la discrimination

- Définition pénale de la discrimination :

« **Toute distinction** opérée entre des personnes physiques ou des personnes morales (en raison de leurs membres) sur le fondement de leur origine, de leur sexe, de leur état de santé, de leur appartenance religieuse, de leur orientation sexuelle [...] »

Article 225-1 du code pénal

- Preuve de la discrimination indirecte :

- Inégalité sur l'ensemble des résultats ?
- Inégalité de performance ?
 - Comparaison des vrais positifs ? et/ou
 - Comparaison des faux positifs ?



Limite statistique

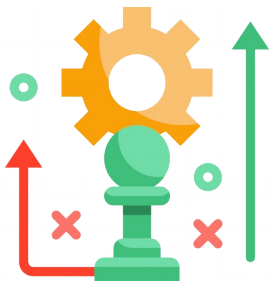
L'intérêt de la transparence remis en question



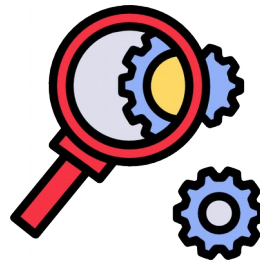
Impact sur la
compétitivité



Divulgence de
données personnelles



Risque de
« triche »



Opacité inhérente à
certains algorithmes



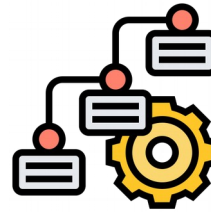
Interprétabilité
de l'information

Des pistes de solutions prometteuses

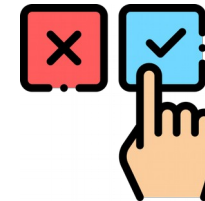
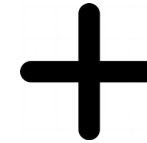
- Comprendre sans ouvrir la « boîte noire » :



Par le test



Fonctionnement
global simplifié

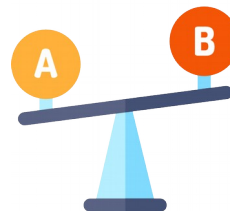
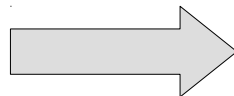


Raisons déterminantes
d'une décision

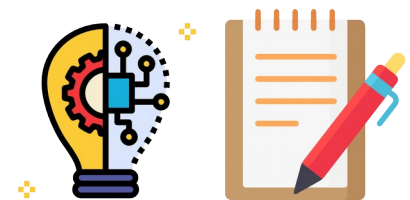
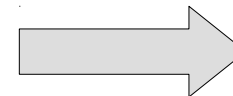
- Une conformité / accountability crédible :



Identifier les
exigences éthiques
et juridiques



Mise en balance
avec les possibilités
techniques



Proposer des
solutions
interdisciplinaires

Merci de votre attention !