

## Chaire “Law, Accountability and Social Trust in A.I.”



**Céline CASTETS-RENARD**

*PR1 UT1/IUF* – Droit du numérique, aspects éthiques et légaux de l’I.A.



**Sylvain CUSSAT-BLANC**

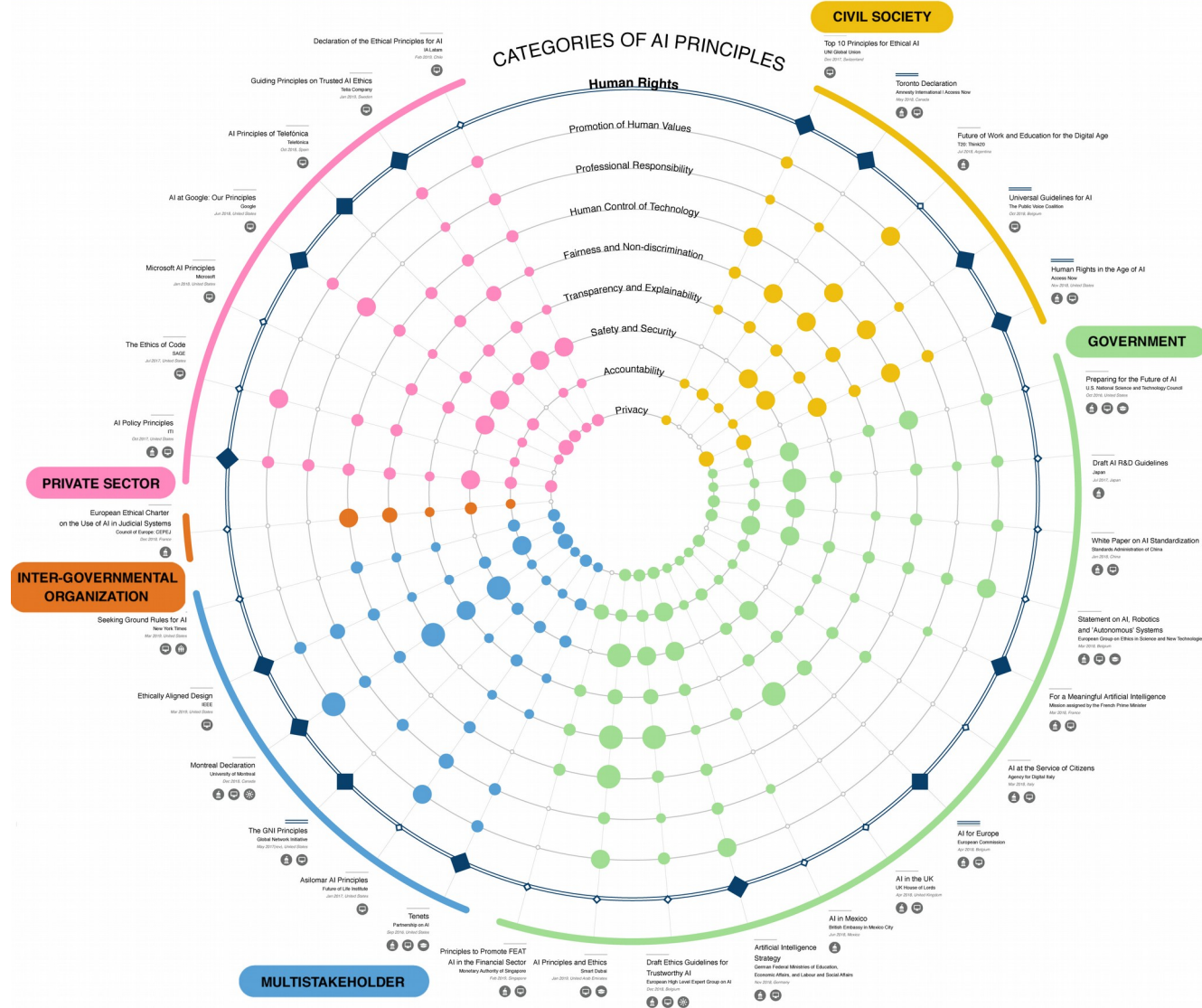
*McF IRIT* – Algorithmes Evolutionnaires, I.A. bio-inspirée



**Laurent RISSER**

*IR1-HDR IMT* – Interprétabilité en I.A., régularisation, applications

# Ethical Principles



<https://ai-hr.cyber.harvard.edu/primp-viz.html>

- Ethical principles related to fundamental rights
- A Map of Ethical and Rights-Based Approaches (Harvard, July 2019)

# Legal Issues

- Algorithmic decision-making: unbiased, non-discriminant, equitable and fair, transparent, explicable, accountable
- Ex. EU (GDPR, art. 22):
  - *The right not to be subject to a decision solely based on automated processing, including profiling*
- Ex. Fr (LIL, art. 10)
  - *The right to obtain a human intervention, the right to contest the decision*
  - *The right to know the rules of the processing and its main characteristics*
- Conseil Constitutionnel (May 2018):
  - *the data controller has to control the algorithmic processing and its developments to explain to the data subject the way of the processing was applied to its situation*
- Ex. NYC Bill (Dec. 2017): A local law in relation to automated decision systems used by agencies (task force)
- Sectorial regulation on autonomous vehicles, smart city, facial recognition...

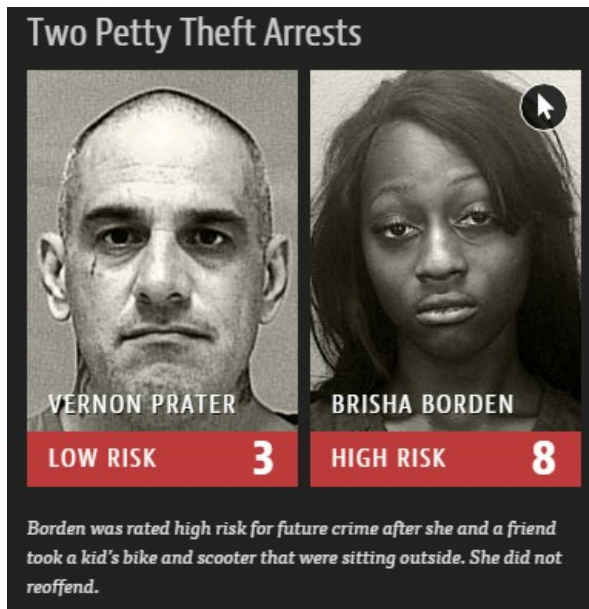
# Qu'est ce que la discrimination ?

## Article 225-1 du code pénal :

- « Constitue une **discrimination** toute distinction opérée entre les personnes physiques sur le fondement de leur origine, de leur sexe, de leur situation de famille, de leur grossesse, de leur apparence physique, [...] de leur état de santé, de leur perte d'autonomie, de leur handicap, de leurs caractéristiques génétiques, de leurs mœurs, de leur orientation sexuelle, de leur identité de genre, de leur âge, de leurs opinions politiques, de leurs activités syndicales, de leurs capacités à s'exprimer dans une langue autre que le français, de leur appartenance ou de leur non-appartenance, vraie ou supposée, à une ethnie, une Nation, une prétendue race ou une religion déterminée.
- Constitue une **discrimination indirecte** une disposition, un critère ou une pratique **neutre en apparence**, mais susceptible d'entraîner, pour l'un des motifs mentionnés au premier alinéa, un **désavantage particulier pour des personnes par rapport à d'autres personnes**, à moins que cette disposition, ce critère ou cette pratique ne soit objectivement justifiée par un but légitime et que les moyens pour réaliser ce but ne soient nécessaires et appropriés. »

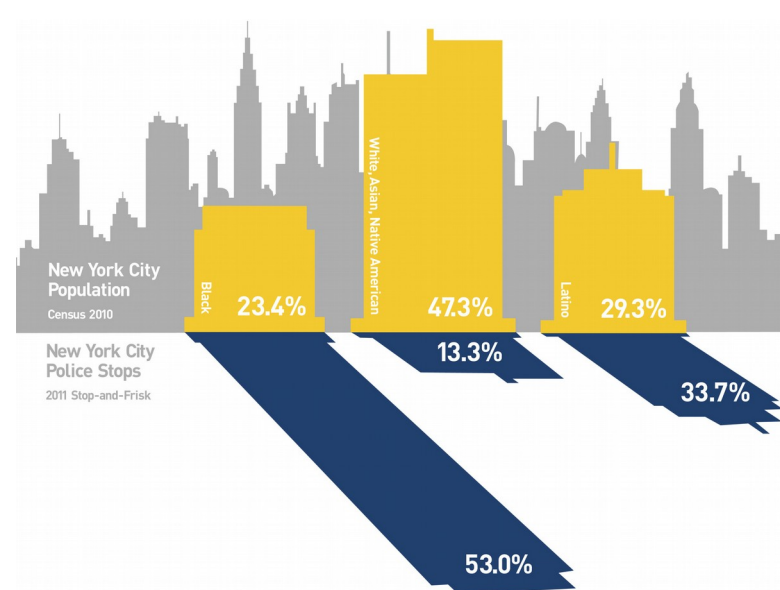
# Discrimination et *Machine learning* : les biais algorithmiques

## COMPAS



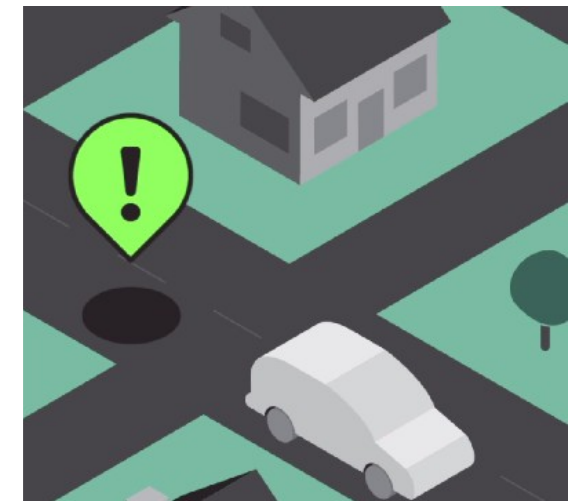
Source : Propublica

## STOP AND FRISK



Source : Civil Rights Litigation ClearingHouse

## STREET BUMP



Source : Boston.gov

**Un algorithme, aussi extraordinaire soit-il, demeure une fabrication humaine. Les erreurs et biais présents lors de sa conception se retrouveront inévitablement dans les résultats produits.**

# La Loi Informatique et Libertés (LIL) : l'interdiction conditionnelle des décisions automatisées

## Article 47 LIL :

« 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 »

### Secteurs concernés



Public



Privé

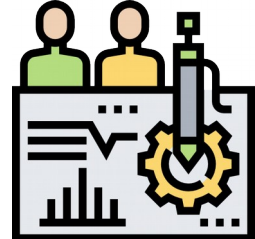
### Criticité de la décision



### Acteurs présents



### Données traitées



# 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. »

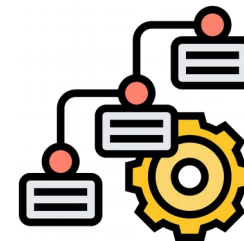
**Moment de la communication**



**Acteurs concernés**



**Contenu de l'information**



Logique de l'algorithme



Conséquences pour l'individu

# La Loi pour une République Numérique (LRN) : le cas des administrations publiques (1/2)

- Obligation de publication :

« [...] les administrations [...] **publient** en ligne les **règles définissant les principaux traitements algorithmiques** utilisés dans l'accomplissement de leurs missions **lorsqu'ils fondent des décisions individuelles.** »

Article L. 312-1-3 CRPA

- Obligation de communication à l'intéressé :

« [...] 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.** »

Article L. 311-3-1 CRPA

## Secteurs



Administrations

## Acteurs présents



Toutes les décisions automatisées

## Criticité



Indifférente

## Données



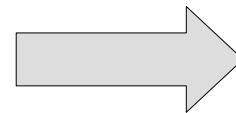
Données personnelles  
seulement



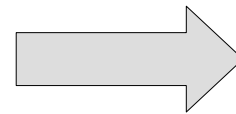
# La Loi pour une République Numérique (LRN) : le cas des administrations publiques (2/2)

Communiquées « sous une forme **intelligible** [...] :

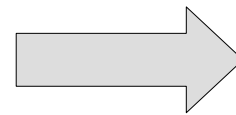
1. Le **degré et le mode de contribution** du traitement algorithmique à la prise de décision ;
2. Les **données traitées** et leurs **sources** ;
3. Les **paramètres de traitement** et, le cas échéant, leur **pondération, appliqués à la situation de l'intéressé** ;
4. Les **opérations effectuées** par le traitement »



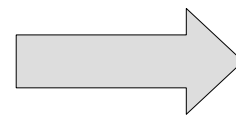
Relation humain / machine au cours du processus de décision



Données entrées (*inputs*) et leurs origines



Variables en jeu et leur poids respectifs



Étapes du processus de traitement

Article R. 311-3-1-2 CRPA

# Les projets à venir : l'exemple de la loi bioéthique

## Article 11 du projet de loi :

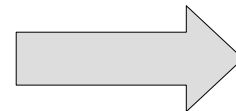
Préserver une garantie humaine  
dans les résultats :

- 1) **Bonne information du patient** lorsqu'un traitement par IA est utilisée à l'occasion d'un acte de soins
- 2) **Garantie d'une intervention humaine** dans le processus
- 3) **Traçabilité des actions** et des **données** utilisées pour le traitements

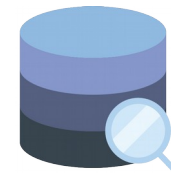
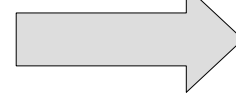
Article L. 4001-3. -I du code de la santé publique



- Quelles informations ?
- Quel format ?



- Qui est-ce ?
- Quel rôle joue-t-il ?
- Que garantie-t-il ?



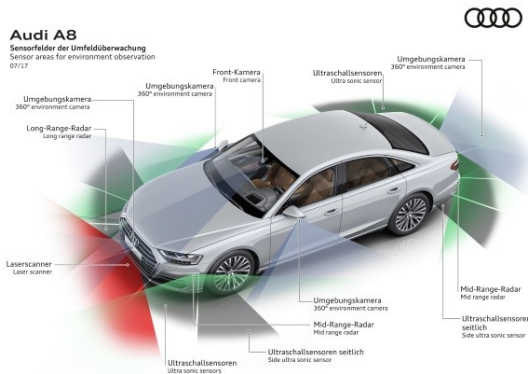
- Comment ?
- Quel objectif ?
- Quelle protection ?

# Et à l'international ?



En Allemagne :

**German Road Traffic Act Amendment Regulating the Use for « Motors Vehicles with Highly or Fully Automated Driving Function » - July 17, 2017**



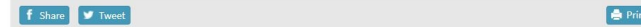
Source : Audi



Aux États-Unis :

**Projet de loi fédérale « Algorithm Accountability Act » pour lutter contre la discrimination algorithmique**

Booker, Wyden, Clarke Introduce Bill Requiring Companies To Target Bias In Corporate Algorithms  
April 10, 2019

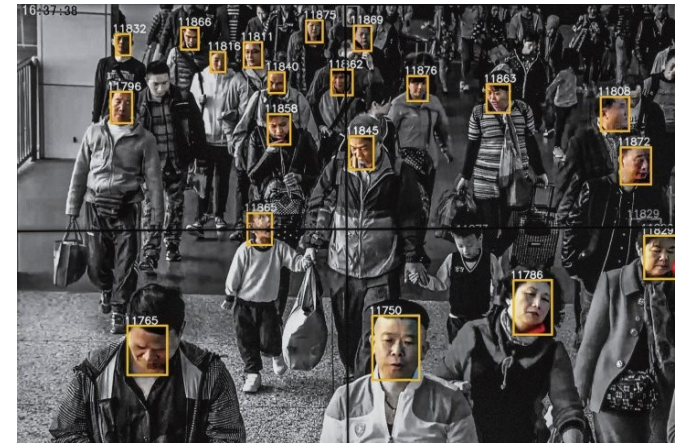


**WASHINGTON, D.C.** -- U.S. Senators Cory Booker (D-NJ) and Ron Wyden (D-OR), along with Rep. Yvette D. Clarke (D-NY) today introduced the [Algorithmic Accountability Act](#), which requires companies to study and fix flawed computer algorithms that result in inaccurate, unfair, biased or discriminatory decisions impacting Americans.

Source : [booker.senate.gov](http://booker.senate.gov)



En Chine :



**How China Is Using “Social Credit Scores” to Reward and Punish Its Citizens**

By Charlie Campbell / Chengdu



**Y**i Tingyue has a lot going for her. She won a scholarship to China’s prestigious Sichuan University, where she graduated with a master’s in graphic design. She drives an Audi A4 and owns a penthouse

Source : *Time magazine*

# Proactivity in Terms of Norms

- To contribute to the development of an ethical and legal framework to regulate the use of A.I., taking special account of bias, discrimination, equity and social justice (fairness)
- To have tools to deal with the "accountability" of actions and decisions resulting from the A.I. (transparency, accountability, auditability, "certifiability", liability) for a social trust
- To define effective A.I. strategies for future rules in the industry (risk limitation)

# Proactivity in Terms of Tools

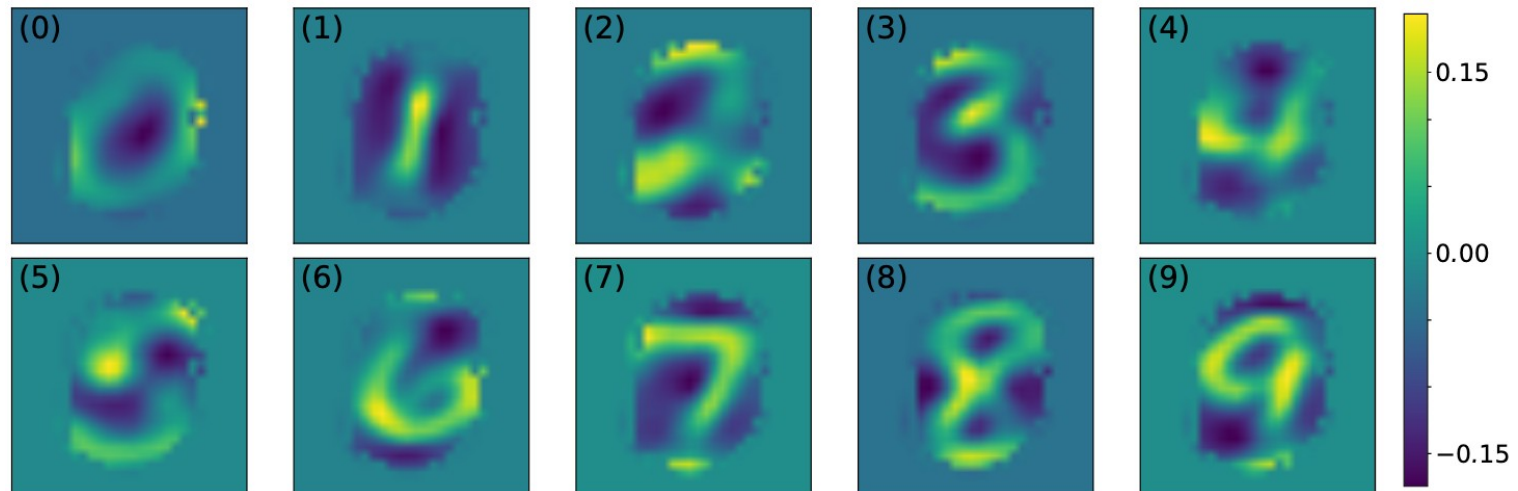
- Developing interpretability solutions in AI
- Application to automatic decision systems (Autonomous transportations and Smart city)
- Studying accreditation and certification methods for AI based systems

# Proactivity in Terms of Tools

- Genetic Programming (GP):
  - Exploring usability of whitebox approaches in industrial contextes
  - Use of evolutionary computation to build:
    - Programs
    - Symbolic regression
  - Possible applications to
    - Autonomous aircraft/cars
    - Genetic improvement (GP for debugging, code optimization, etc.)
    - Data modelling
    - Etc.

# Proactivity in Terms of Tools

- Interpretability of black-box decision rules



Bachoc, Gamboa, Halford, Loubes, Risser (2019) — <https://arxiv.org/pdf/1810.07924.pdf>

- Regulariza  $\hat{\theta} = \arg \min_{\theta} R(\theta) + \lambda W_2^2(\mu_{\theta,0}^n, \mu_{\theta,1}^n)$  es

Risser, Vincenot, Couellan, Loubes (2019) — <https://arxiv.org/pdf/1908.05783.pdf>

### **Within ANITI**

IP1 (Fair Representative Data For Artificial Intelligence)  
Relationships with the Chair of J.-M. Loubes on Fairness and IP2

### **Partnerships**

University of Ottawa  
PhD CIFRE, Toulouse Metropole. Ronan Pons, recruited  
Ongoing procedure: PhD CIFRE, Orange Lab  
PhD, Claire Boine, Researcher, Boston University

