Éthique et Santé
Loyauté des Algorithmiques

Philippe Besse

Université de Toulouse – INSA
Institut de Mathématiques – UMR CNRS 5219
CIMI – Projet AOC

éthique et numérique
les algorithmes en débat

Quelles questions éthiques ?

- *La technologie elle elle neutre ?* : qui y a accès ?
- Pour quel usage ? Cas de *Pharmakon*
- Entraves à la concurrence, *algorithmic pricing*, comparateurs *Virtual Competition* (Ezrachi et Stucke, 2016)
- *Open data* (2013-16), anonymisation, fin du consentement libre et éclairé
  - Projet `care.data` NHS et Royaume Uni : (*Social License*)
  - Projet *Data Science Initiative* : $\mathcal{X}$ et CNAM, base Sniiram
- ...

Acceptabilité et *Loyauté*

- *Trustworthiness* : Mériter la confiance : fiabilité, crédibilité, non discriminatoire
- *Accountability* : responsabilité, capacité à rendre compte

Algorithmes, Recherche Scientifique, Santé, *Loyauté*

1. *Production* scientifique : flux de données et algorithmes
   - *Reproductibility in Science* (Begley et Ioannidis, 2015)
   - *Why Most Clinical Research Is Not Useful* (Ioannidis, 2016)
2. *Utilisation* & Décision algorithmique
   - *Trustworthiness* Médecine personnalisée
   - *Accountability* de la relation médecin — patient
3. Algorithmes loyaux par construction

Déontologie scientifique, Statistique, Grosses Data

- Démarche hypothético-déductive
- Statistique : usages, abus, fraudes...
  - 1930 ko Planification expérimentale et Test d'hypothèse
  - 1990 Mo Données préalables, fouille ou *data mining*
  - 2000 Go Données omiques avec $p >> n$: indétermination
    Reproductibilité des résultats scientifiques
    Gènes de la dépression (Hek et al. 2013)
  - 2010 To Grosses data : n très grand, Tests *vs.* Prévisions
- Algorithmes de *Machine Learning* (*e.g.* biomarqueurs)


M PIXELS CHRONIQUES DES (R)ÉVOLUTIONS NUMÉRIQUES

Après IBM et Google, Microsoft s'intéresse de près au cancer

LE MONDE | 21.09.2016 à 15h47

Éthique, Épistémologie et Science des (grosses) Données

- Fin de la théorie et obsolescence de la démarche scientifique (Anderson, 2008)
- Hypothèses déduites des données pas d'une théorie
- Validation de la recherche : cas de la base Sniiram
  1. Test d'une hypothèse *a priori* : effets indésirables et risque d'un médicament
  2. Fouille systématique : corrélation, co-occurrence, motifs
     Validation Retour à (1) Sinon risque d'artefact (*data snooping*)
- Évolution méthodologique pas épistémologique (Keppler, 1609)
- Reproductibilité : diffusion des données et des codes d'analyse
- Science ouverte

Décision algorithmique et flux de données

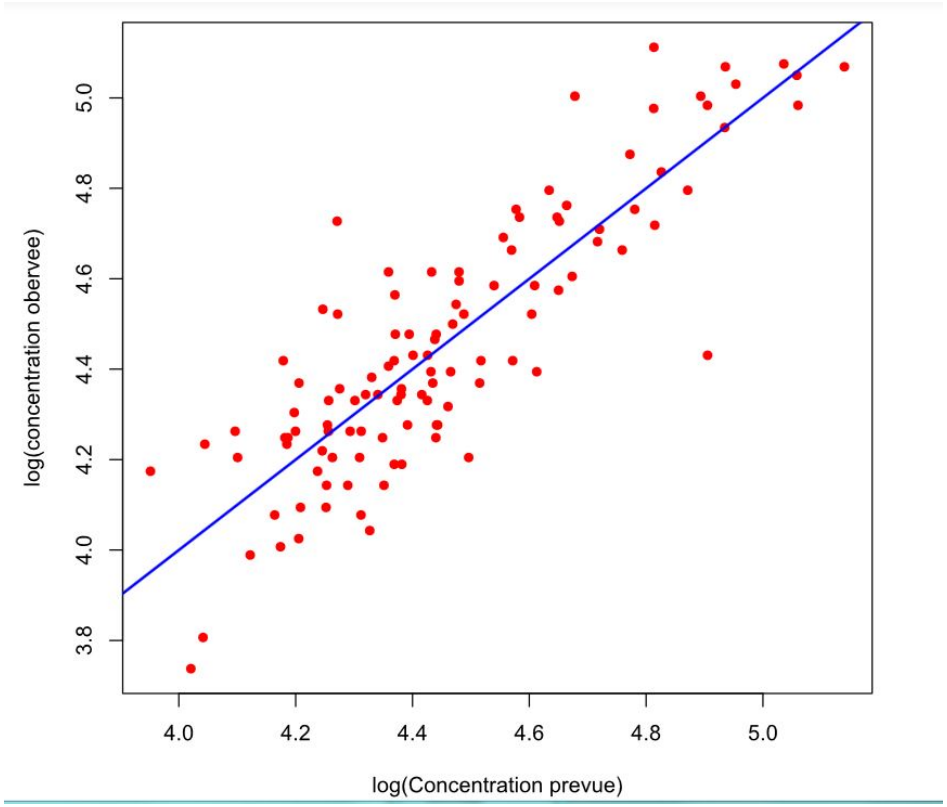
- Décision issue d'un traitement automatisé
- Algorithme procédural type APB (admission post-bac)
- Algorithme par Apprentissage Machine ou Statistique
  - Choix de :
    Diagnostic, traitement, action commerciale, maintenance préventive, accord de crédit, mise sous surveillance, d'un produit...
  - Prévision (appris des données) d'une probabilité ou risque de :
    Déclencher une maladie, départ d'un client, défaillance d'un système, défaut de paiement, radicalisation, d'appétence...
  - Décision découle d'un Modèle ou Algorithme :
    - Estimé ou appris sur un *échantillon d'apprentissage*
    - Optimisé (compromis biais-variance) par *validation croisée*
    - Évalué sur un *échantillon test* indépendant

Loyauté des Algorithmes

- *Accountability* et *Trustworthiness*
- Se traduisent et s'évaluent par leur :
  - Explicabilité et transparence
  - Qualité de prévision et justesse de décision
  - Biais et discrimination
- Contraintes juridiques *vs.* caractéristiques techniques
- Zone de non droit ou *disruption*
- Quelle éthique ?
  - Explicabilité : décret loi Rép. Num.
  - Qualité : rien
  - Biais : discrimination
    - Individuelle ou collective
    - Intentionnelle ou non
- Enjeu : Acceptabilité d'une nouvelle technologie

Explicabilité : modèle linéaire du ”siècle dernier”

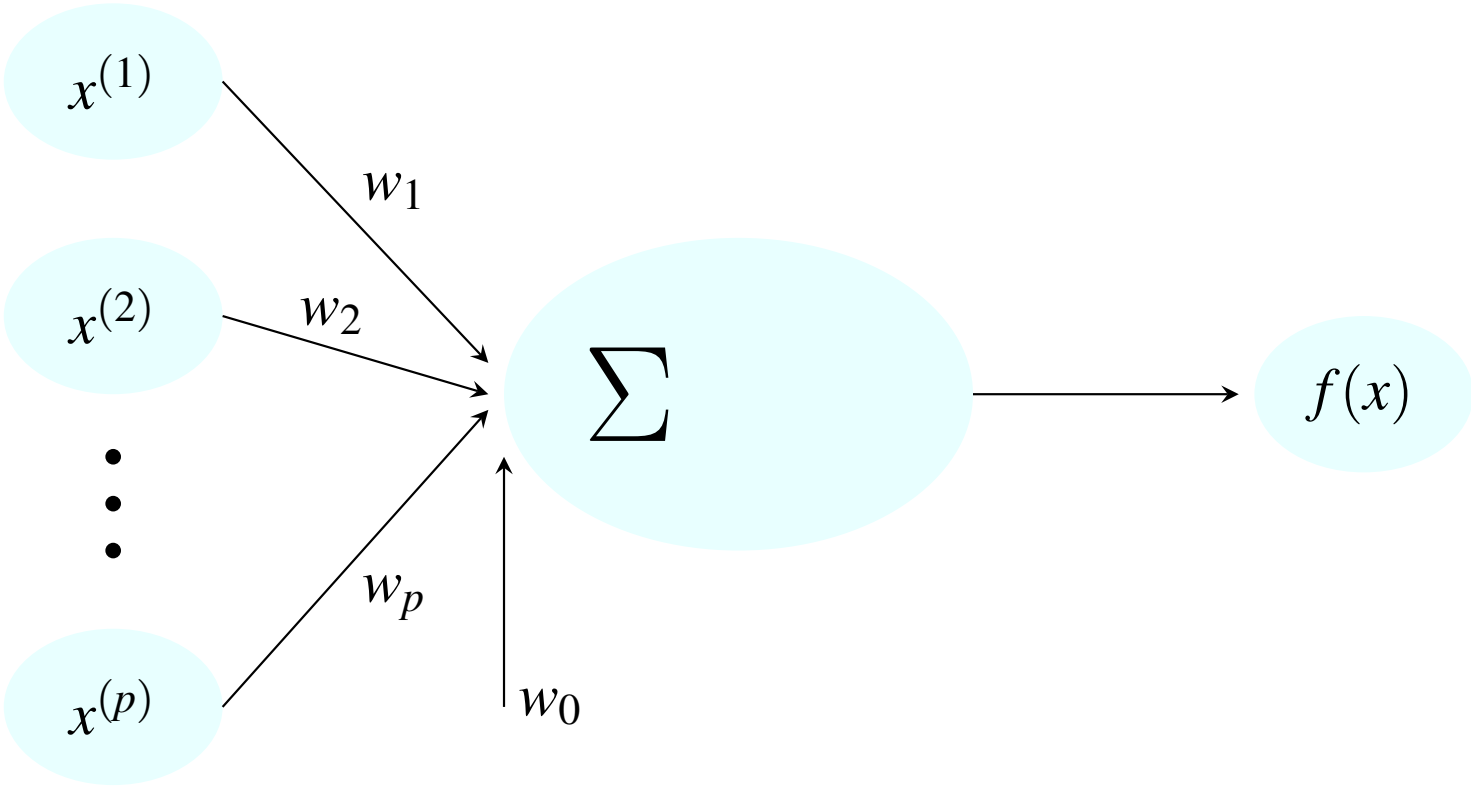
Prévoir la Concentration en Ozone



$$\begin{aligned} \log(\text{ConcODemain}) &= 2,4 + 0,35 \times \log(\text{ConcOJour}) + 0,05 \times \text{Sec} + \\ &+ 0,03 \times \text{T12} - 0,03 \times \text{Ne9} + 0.1 \times \text{Vx9} \end{aligned}$$

Modèle / Neurone Linéaire

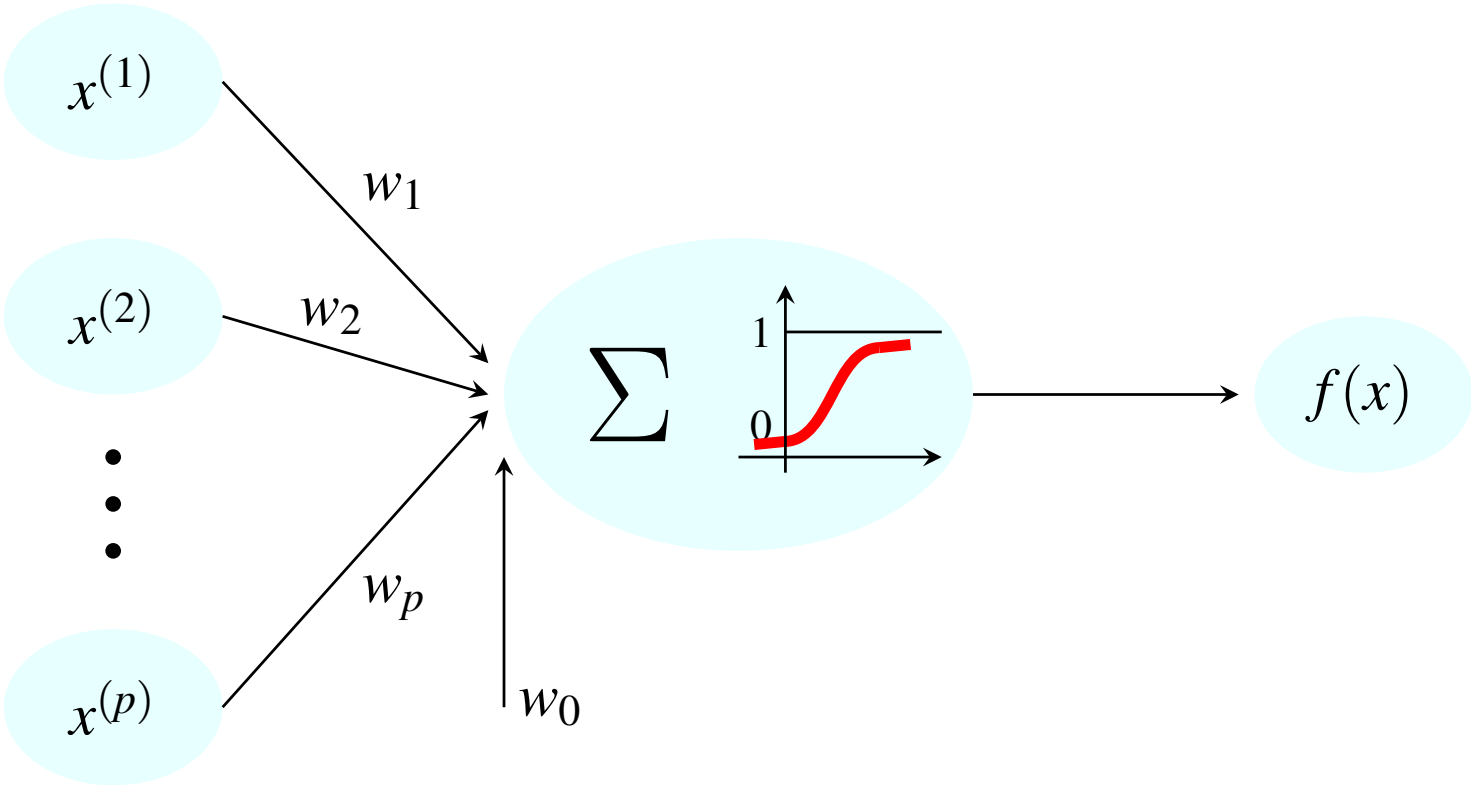
Modéliser / prévoir une variable quantitative



$$f(x) = w_0 + w_1 \times x^{(1)} + w_2 \times x^{(2)} + \cdots + w_p \times x^{(p)}$$

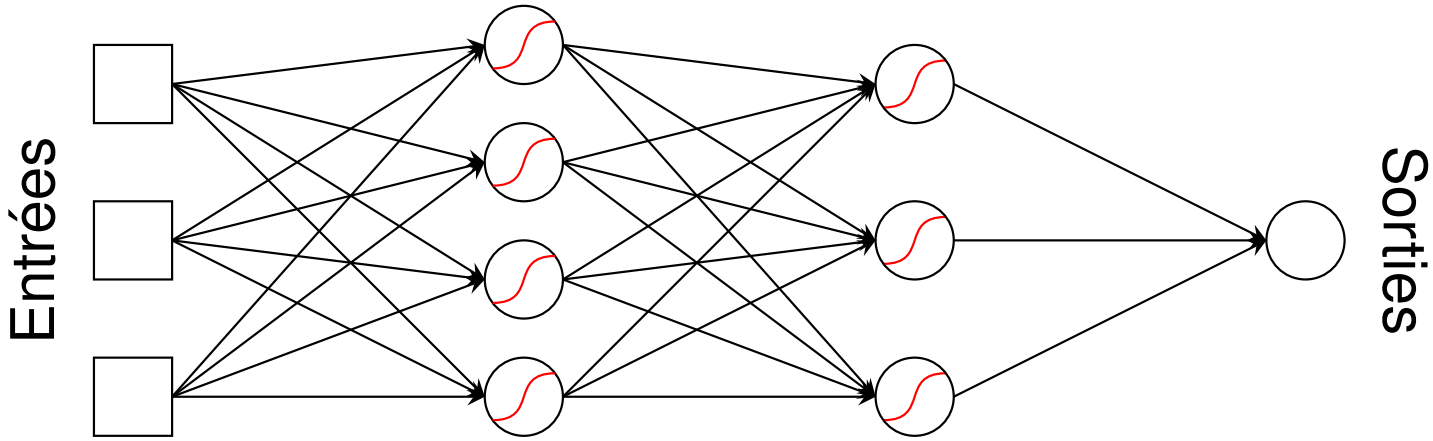
Modèle / **Neurone** *logistique*

Variable binaire : Maladie, Panne, Départ, Faillite...



Exemple en épidémiologie : évaluer les facteurs de risque

Explicabilité : réseau de neurones (Perceptron)



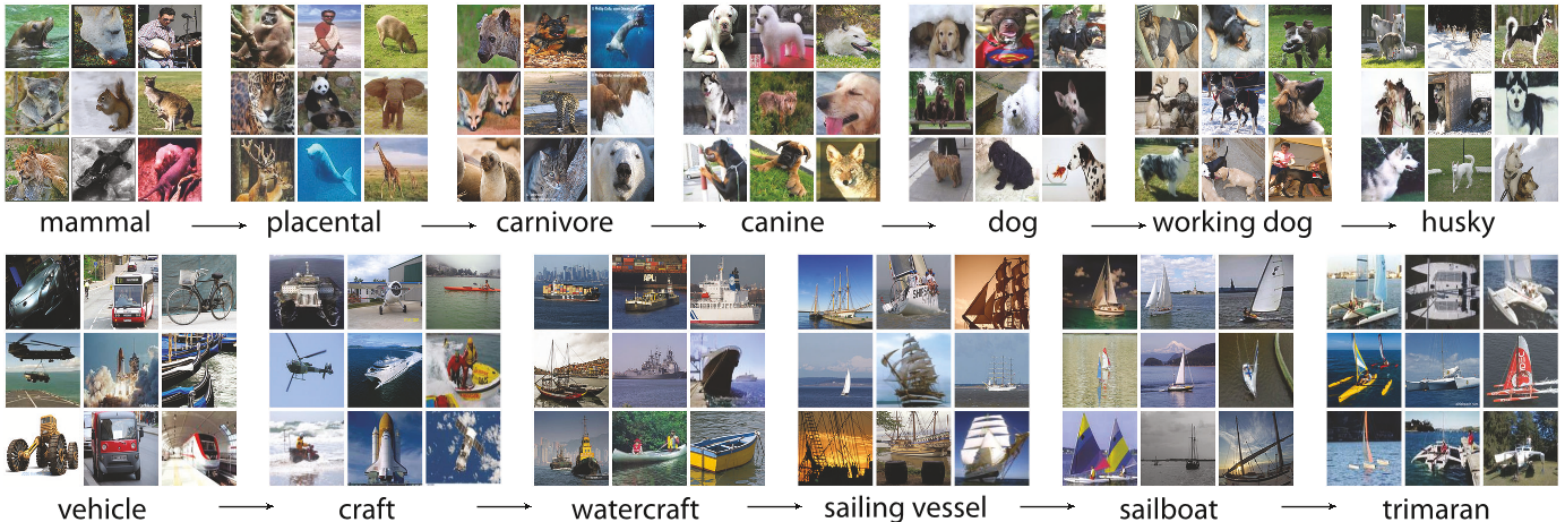
$x = (x^{(1)}, \ldots, x^{(p)})$ *Couche 1* *Couche 2* $y = F(x)$

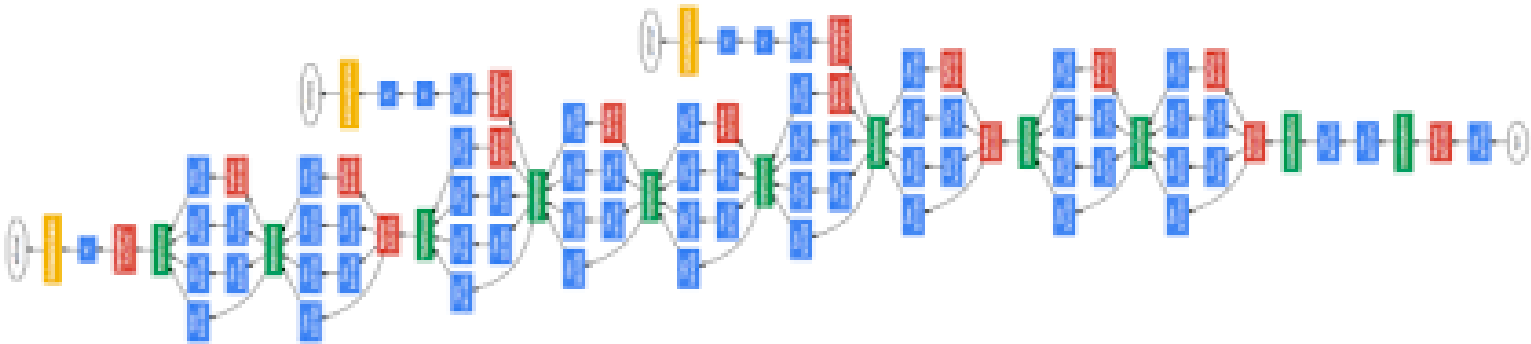
- Explication impossible : *Boîte Noire*
- Aides à l'interprétation
- Idem pour k-p.p.v., SVM, *boosting, random forest*...

Explicabilité : Deep Learning & ImageNet

15 millions d'images, 22000 catégories

2016 : 152 couches et mieux que l'expert humain

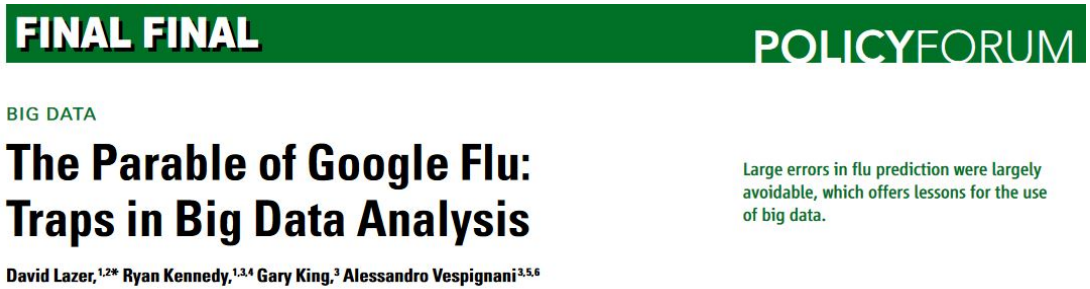




Grosses données et qualité de prévision

- Plus de données entraine-t-il une meilleure prévision ?
- *L'efficacité prédictive sera d'autant plus grande qu'elle sera le fruit de l'agrégation de données massives*
  in *La Gouvernementalité Algorithmique* (Rouvroy et Berns, 2013)
- **Vrai** et **Faux**
- Ne pas confondre estimation / prévision d'une moyenne
  (*loi des grands nombres*)
  et celle d'un comportement individuel

Fiabilité des algorithmes



FINAL FINAL POLICYFORUM

BIG DATA

The Parable of Google Flu: Traps in Big Data Analysis

Large errors in flu prediction were largely avoidable, which offers lessons for the use of big data.

David Lazer,[1,2*] Ryan Kennedy,[1,3,4] Gary King,[3] Alessandro Vespignani[3,5,6]

Google flu trend de 2008 à 2015

Journal of oncology practice

The Authoritative Resource for Practicing Oncology

Enter search words / phrases / DOI / authors / keywords / etc.

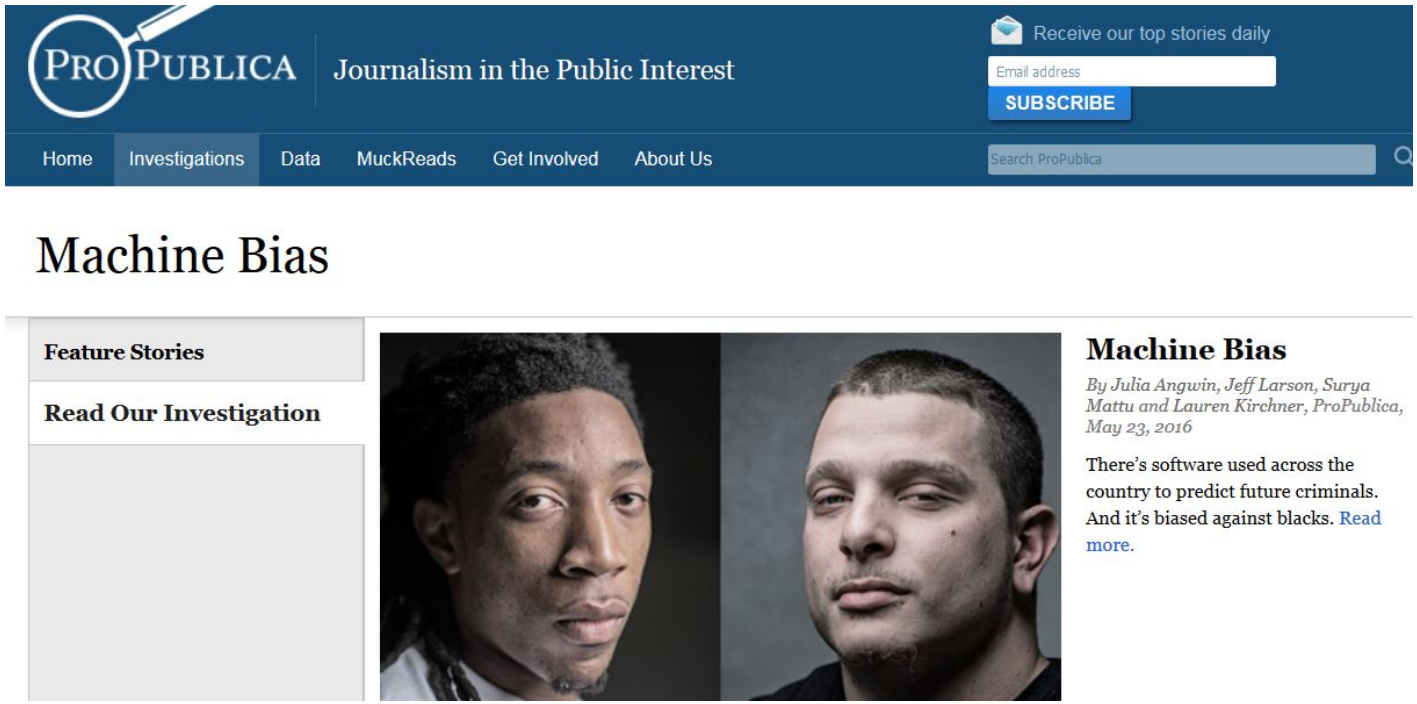
Newest Articles | Issues | Browse By Topic | Special Content | Authors | Su

ORIGINAL CONTRIBUTIONS | FOCUS ON QUALITY

Screening for Pancreatic Adenocarcinoma Using Signals From Web Search Logs: Feasibility Study and Results

Taux de faux positif : 10^{-4}*, taux de vrais positifs : 5 à 15%* Pharmakon ?

Justice prédictives : ProPublica *vs.* Equivant (NorthPointe Inc.)



Angwin et al. (2016)

ProPublica *vs.* Equivant (NorthPointe Inc.)

- Absence de discrimination selon NorthPointe Inc.
  - Distributions des scores (m_1 et m_2) similaires
  - Taux d'erreur ($FN + FP/n$) similaires
- Discrimination selon ProPublica

Matrice de confusion

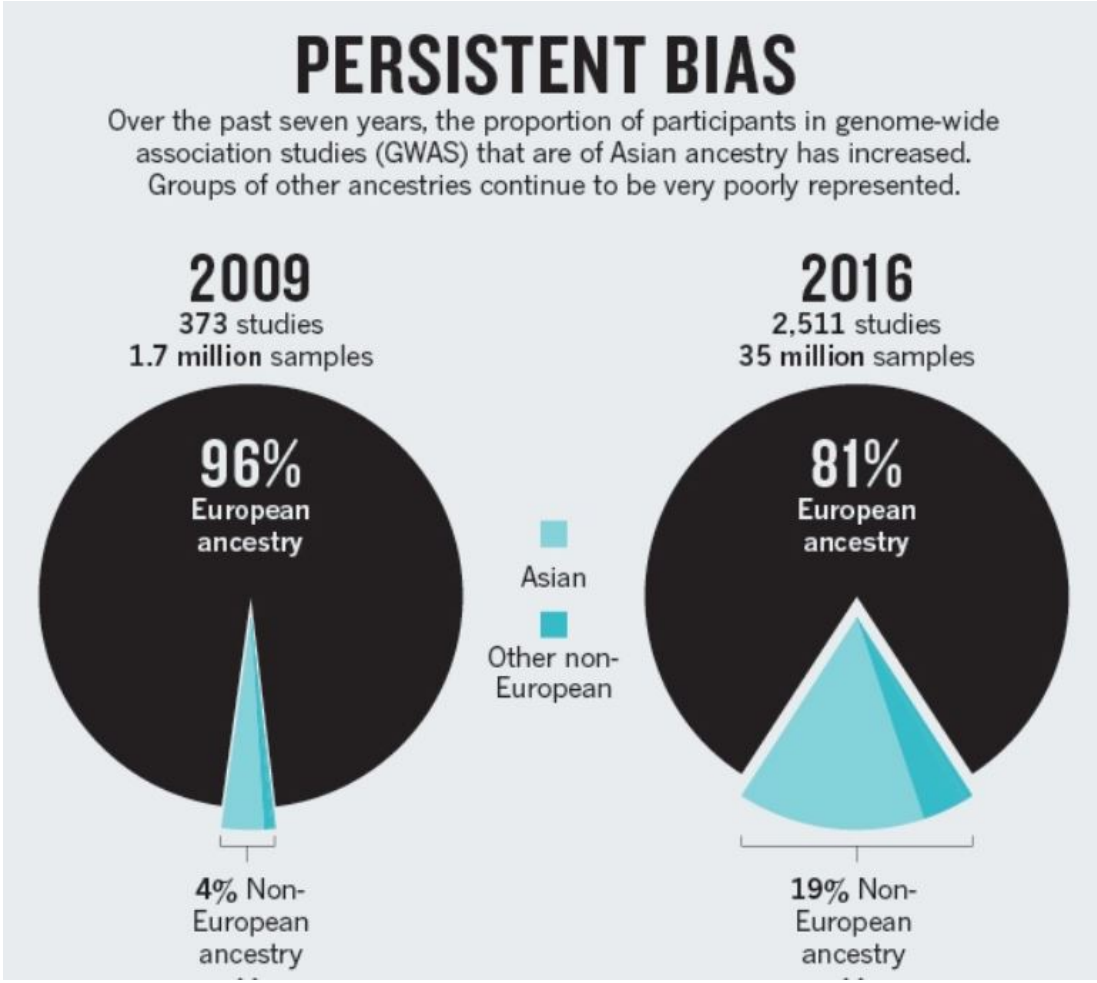
Observation / Récidive	Score: Faible	Score: Élevé	
Oui	FN	VP	q_1
Non	VN	FP	q_2
	m_1	m_2	n

  - Taux de faux positifs$= FP/q_2$
    afro-américain (45%) vs. caucasiens (25%)
- Taux de récidive afro-américain plus élevé (Chouldechova, 2016)
- Taux d'erreur très élevé (40%)

Biais en Santé : exemple

- Médecine de précision *vs.* de population
- Traitement personnalisé : sommes nous tous égaux ?
- *GWAS Genome Wide Association Studies*
  Bases d'associations pangénomiques
- Liaisons entre variants génétiques (SNPs) et traits phénotypiques
- Biais
  - Ethnique : population d'ascendance blanche européenne
    Genomics is failing on diversity (Popejoy et Fullerton, 2016)
  - Âge et environnement : bases transversales et pas longitudinales
  - Genre : Chang et al. (2014), Pulit et al. (2017)

GWAS : Biais ethnique



Popejoy et Fullerton (2016)

Conclusion

- Loyauté des algorithmes d'apprentissage
  - Explicabilité
  - Qualité de prévision, de décision (cf. sondages)
  - Absence de biais (*testing*)
- Cadre juridique flou

1. Déontologie scientifique
   - *Science des (grosses) données* & Algorithmes d'apprentissage
   - Rigueur, *data snooping* & reproductibilité
2. Éthique et usages des algorithmes : Acceptabilité ou Rejet
   - Relation médecin / patient
   - Usage de l'usage : Assurance et asymétrie d'information *vs. GINA*
   - Que penser de `23andme.com`
3. Algorithmes loyaux par construction (*Accountability by design*)
   - Explicabilité *vs.* qualité
   - Débiaiser si la donnée sensible est connue

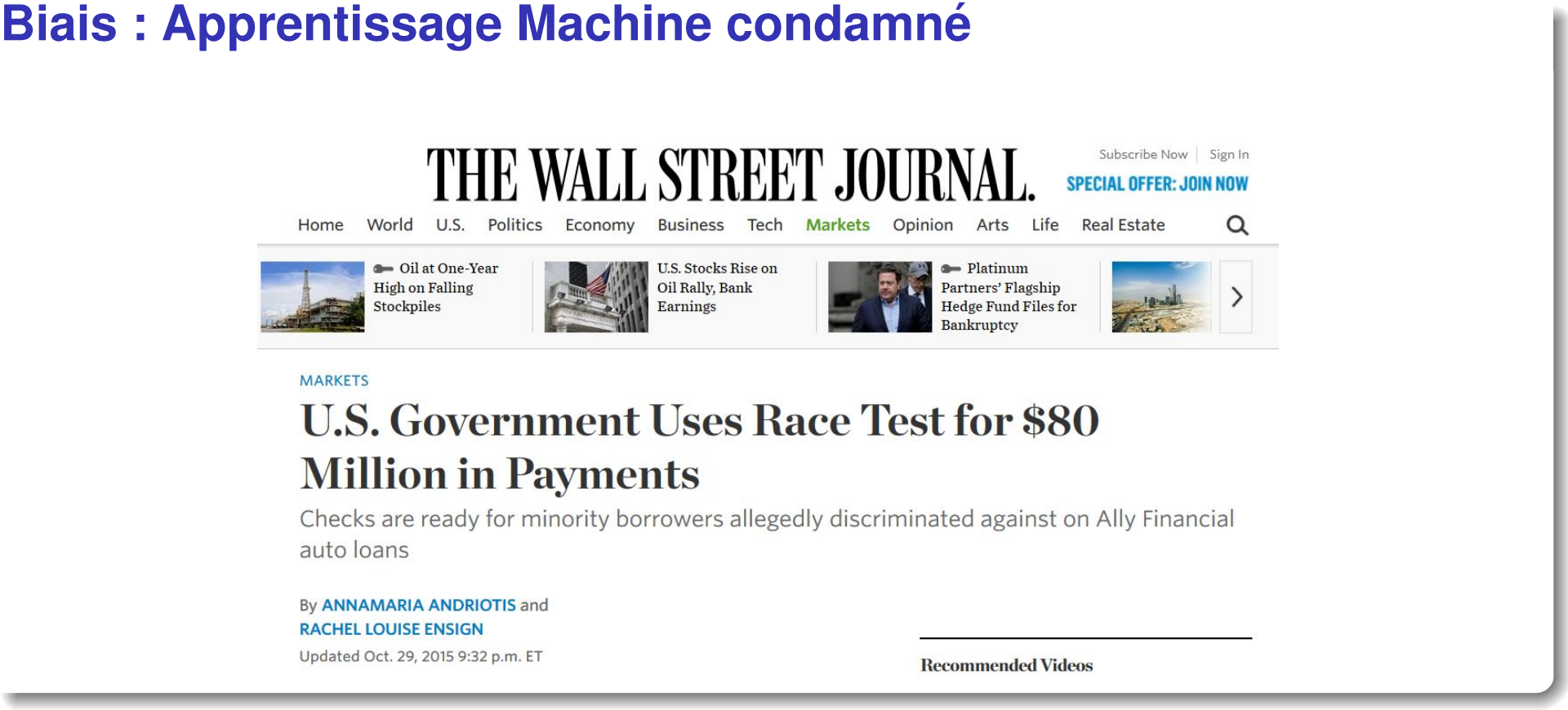
Références

- Angwin J., Larson J., Mattu S., Kirchner L. (2016). How we analyzed the compas recidivism algorithm. ProPublica, en ligne consulté le 28/04/2017.
- Begley G., Ioannidis J. (2015), Reproducibility in Science Improving the Standard for Basic and Preclinical Research, *Circulation Research*, 116(1), 116-126.
- Chouldechova A. (2016). Fair prediction with disparate impact: A study of bias in recidivism prediction instruments, arXiv pre-print.
- Datta A., Sen S., Zick Y. (2016). Algorithmic Transparency via Quantitative Input Influence : Theory and Experiments with Learning Systems, in IEEE Symposium on Security and Privacy.
- Ezrachi A., Stucke M. (2016). *Virtual Competition The promise and perils of algorithmic-driven economy*, Harvard University Press.
- Feldman M., Friedler S., Moeller J., Scheidegger C., Venkatasubramanian S. (2015). Certifying and removing disparate impact, arXiv-preprint.
- Hajian S., Domingo-Ferrer J., Farràs O. (2014). Generalization-based Privacy Preservation and Discrimination Prevention in Data Publishing and Mining, Data Mining and Knowledge Discovery 28 (5-6), 1158-1188.
- Ioannidis J. (2015). Why Most Published Research Findings Are False, *PLOS Medecine*, 2(8).
- Ioannidis J. (2016). Why Most Clinical Research Is Not Useful, *PLOS Medecine*, 13(6).
- Kamiran F., Calders T. (2011). Data Pre-Processing Techniques for Classification without Discrimination, Knowledge and Information Systems 33(1).
- Kamiran F., Calders T, Pechenizkiy M. (2010). Discrimination Aware Decision Tree Learning in ICDM, 869-874.

Références – suite

- Pedreschi D., Ruggieri S., Turini F. (2008). Discrimination-Aware Data Mining. In *KDD*, pp. 560-568.
- Popejoy A., Fullerton S. (2016). Genomics is failing on diversity, *Nature*, 538, 161 ?164.
- Rouvroy A., Berns T. (2013). Gouvernementalité algorithmique et perspectives d'émancipation, *Réseaux*, 177, 163-196.
- Ruggieri S. (2014). Using t-closeness anonymity to control for non-discrimination, Transaction on Data Privacy, 7, 99-129.
- Ruggieri S., Pedreschi D., Turini F. (2010). Data mining for discrimination discovery. In TKDD 4(2).
- Zafar M., Valera I., Rodriguez M., Gummadi K. (2017). Fairness Constraints : Mechanisms for Fair Classication in International Conference on Articial Intelligence and Statistics (AISTATS), vol. 5.
- Wachter S., Mittelstadt B., Floridi L. (2017). Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation, *International Data Privacy Law*, à paraître.
- Zafar M., Valera I., Rodriguez M., Gummadi K. (2017). Fairness Constraints: Mechanisms for Fair Classication in International Conference on Articial Intelligence and Statistics (AISTATS), vol. 5.
- Zemel R., Wu Y., Swersky K., Pitassi T., Dwork C. (2013). Learning Fair Representations in JMLR W&CP 28(3), 325 ?333.
- Zeng J., Ustuny B., Rudin C. (2016). Interpretable Classification Models for Recidivism Prediction, arXiv pre-print.
- Zliobaité I. (2015). A survey on measuring indirect discrimination in machine learning. arXiv pre-print.

Biais : Apprentissage Machine condamné



THE WALL STREET JOURNAL.

Subscribe Now | Sign In

SPECIAL OFFER: JOIN NOW

Home World U.S. Politics Economy Business Tech Markets Opinion Arts Life Real Estate

Oil at One-Year High on Falling Stockpiles

U.S. Stocks Rise on Oil Rally, Bank Earnings

Platinum Partners' Flagship Hedge Fund Files for Bankruptcy

MARKETS

U.S. Government Uses Race Test for $80 Million in Payments

Checks are ready for minority borrowers allegedly discriminated against on Ally Financial auto loans

By ANNAMARIA ANDRIOTIS and RACHEL LOUISE ENSIGN

Updated Oct. 29, 2015 9:32 p.m. ET

Recommended Videos

Règlement 2016/679/EU sur la protection des données personnelles

Article 22 Décision individuelle automatisée, y compris le profilage

1 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...

3 Le responsable du traitement met en œuvre des mesures appropriées pour la sauvegarde des droits et libertés et des intérêts légitimes de la personne concernée, au moins du droit de la personne concernée d'obtenir une intervention humaine de la part du responsable du traitement, d'exprimer son point de vue et de contester la décision.

4 Les décisions visées ne peuvent être fondées sur les catégories particulières de données à caractère personnel (cf. article 9 : biométriques, génétiques, de santé, ethniques ; orientation politique, syndicale, sexuelle, religieuse, philosophique).

Effectif en mai 2018

Loi n°2016-1321 du 7/10/2016 pour une République Numérique

Article 4 Une décision individuelle prise sur le fondement 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 par l'administration à l'intéressé s'il en fait la demande

Article 6 Sous réserve des secrets protégés, 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 50 Les opérateurs de plateformes en ligne dont l'activité dépasse un seuil de nombre de connexions défini par décret élaborent et diffusent aux consommateurs des bonnes pratiques visant à renforcer les obligations de clarté, de transparence et de loyauté.

Décret du 16/03/2017 Art. R. 311-3-1-2.

L'administration communique à la personne faisant l'objet d'une décision individuelle prise sur le fondement d'un traitement algorithmique, à la demande de celle-ci, sous une forme intelligible et sous réserve de ne pas porter atteinte à des secrets protégés par la loi, les informations suivantes :

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.