



Gouvernance des algorithmes

Christine Balagué

Titulaire de la Chaire réseaux sociaux et objets connectés
2013-2015: Vice-présidente du Conseil National du Numérique
christine.balague@it-sudparis.eu



Petite histoire des algorithmes

- **III^{ème} siècle avant notre ère:**
 - Les babyloniens décrivent déjà des méthodes de calcul et des résolutions d'équation (calcul de l'impôt)
- **300 ans avant notre ère:**
 - Avec Euclide le premier véritable algorithme permet le calcul du « plus grand dénominateur commun »
- **IX^{ème} siècle après J.C.:**
 - Le mathématicien Muhammad Ali Khwarizmi, auquel on doit notre système de numération décimal, laisse son nom à 2 mots fondamentaux: algèbre et algorithme.
 - Le moine Adélard de Bath introduit le terme latin « algorismus »
- **1940:** Alan Turing décrit le premier modèle mathématique d'un algorithme.



La question de la gouvernance des algorithmes est un sujet d'actualité

'Partnership on AI' formed by Google, Facebook, Amazon, IBM and Microsoft

Two big Silicon Valley names are missing from the alliance, which aims to set societal and ethical best practice for artificial intelligence research



Partnership on Artificial Intelligence hopes to invite 'academics, non-profits and specialists in policy and ethics' to join. Photograph: Alamy

Going by the unwieldy name of the Partnership on Artificial Intelligence to Benefit People and Society, the alliance isn't a lobbying organisation (at least, it says it "does not intend" to lobby government bodies). Instead, it says it will "conduct research, recommend best practices, and publish research under an open license in areas such as ethics, fairness and inclusivity; transparency, privacy, and interoperability; collaboration between people and AI systems; and the trustworthiness, reliability and robustness of the technology".

theguardian

Wednesday 28 September 2016
22.00 BST

Trump Will Become President, Says Extremely Accurate Statistician

His statistical model has only ever been wrong once in 104 years.

By [Rachel Dicker](#) | Associate Editor, Social Media Feb. 26, 2016, at 11:00 a.m.



The odds are ever in Donald Trump's favor.

Helmut Norpoth, a professor of political science at [Stony Brook University](#), has developed a statistical model that predicts a 97 percent to 99 percent chance that Trump will win the 2016 presidential election if he wins the Republican nomination, [The Blaze](#) reports.

Norpoth's model has correctly predicted the outcome of every single presidential election since 1912 – save one, the election of 1960, which some [believe was rigged](#). It uses a candidate's performance in their party's primary coupled with electoral cycle patterns to determine the likely outcome of the general election.

Admission post-bac : le ministère poussé à dévoiler l'algorithme qui oriente les lycéens

LE MONDE | 19.09.2016 à 08h16 • Mis à jour le 20.09.2016 à 09h16 |

Par Séverin Gravelleau



- Un algorithme décliné différemment sur 2 logiciels (Affelnet et APB): 1 million d'élèves tous les ans
- Réaction association Droit des citoyens
- Objectif initial APB: rationaliser le système le rendre plus efficace et moins opaque
- Effets inverses:
 - Opacité sur l'impact des notes
 - Concentration de boursiers (25% district Est de Paris vs 10% en moyenne)
 - Sélection aléatoire des universités quand le nbre de candidats est très important pour une formation

24 heures avec les algorithmes

Révolution. Laisser les robots devancer vos désirs, gérer votre quotidien et même vos amours n'est plus de la science-fiction. Récit d'une journée « mathématisée » tout à fait ordinaire...

PAR MATHILDE HELLE



CAPTEURS DE SOMMEIL

7h45

Vous vous éveillez d'excellente humeur, car votre réveil intelligent, calqué sur vos phases d'endormissement, a sonné à l'instant idéal. Un miracle rendu possible grâce à des capteurs et à un algorithme surprenant qui analyse votre sommeil. Un peu de musique pour vous encourager à affronter la longue journée qui vous attend ? Vous faites confiance à Spotify et à son nouvel algorithme pour trouver le morceau qui convient.

PETIT DÉJEÛNÉ

8h15

Qu'allez-vous manger ce matin ? L'algorithme de WeCook vous aide à préparer un petit déjeuner équilibré qui prend en compte à la fois vos goûts, votre budget et vos besoins nutritionnels. Tout en buvant votre café, vous allez faire un tour sur les réseaux sociaux. Facebook vous recommande d'ajouter à votre cercle d'amis la jeune femme rencontrée au bar hier soir, Twitter vous propose de suivre le compte de l'équipe de France de football. Eh oui, Twitter sait que vous avez regardé le dernier

ENS-MOI QU'IL Y A

8h30

Comme tous les matins, vous prenez le métro, et comme tous les matins vous vous demandez si, aujourd'hui encore, « le trafic est ralenti sur la ligne 8 dans les deux sens ». Vous regardez alors l'application RATP qui vous propose l'itinéraire le plus rapide en fonction de l'heure et du trafic. Vous quittez votre appartement. Spotify toujours dans les oreilles. Une fois dans le métro, vous vous lancez dans une partie de « Candy Crush ». Vous ne le savez peut-être pas, mais la gestion des bonus de ce jeu dépend entièrement d'un algorithme.

RENCONTRE DU 3^e TYPE

10 heures

Au bureau, peinant à traduire un terme anglais, vous tentez Google Translate. Mais l'algorithme n'étant pas toujours fiable, vous vous résignez comme au bon vieux temps à consulter un dictionnaire. Après deux heures et demi de travail, vous vous accordez une pause café et en profitez pour faire un tour sur Tinder. Ne sachant que faire ce soir, vous cherchez quelqu'un pour vous tenir compagnie. Qui sait, vous trouverez peut-être l'amour.

LE NO DÉJEUNER

13 heures

Depuis quelques mois, vous vous êtes lancé un défi de taille : devenir végétarien. Difficile de tenir sur le long terme, notamment quand il s'agit de manger sur le pouce. Mais les algorithmes sont là pour vous aider à tenir le cap ! NotCo est capable de copier pour vous des produits alimentaires d'origine animale à l'aide de l'intelligence artificielle, de la biochimie et des plantes. Serein, vous dégustez alors votre NotHolidays végétarien et retournez travailler.

CYBER DOCTEUR

23 heures

Vous êtes frappé par un violent mal d'estomac mais ne savez quel médicament prendre dans votre pharmacie personnelle pour vous soulager. CitizenDoc vous aide à trouver celui qui calmera votre douleur. Vous tapez vos symptômes dans l'application, qui vous recommande le médicament adéquat. Vous faites ensuite un petit tour sur Météo France pour savoir le temps qu'il fera demain. À l'aube, c'est décidé, vous irez courir avec RunKeeper. Un ami, un voisin ? Non, encore un algorithme.

TELEPORTATION URBAINE

19 h 30

Vous avez enfin bouclé le dossier sur lequel vous travaillez depuis plusieurs semaines. Mais vous ne sortez jamais à l'heure et vous prenez les transports en commun. Heureusement, Uber est là... Fixant ses tarifs en fonction du trafic et du prix du marché, l'application vous trouve rapidement le chauffeur le plus proche grâce à son algorithme. Trois minutes plus tard, vous êtes déjà en route.

EN MUSIQUE

20 heures

Vous retrouver la jeune femme avec laquelle vous avez rendez-vous. Vous faites connaissance. Une musique que vous ne connaissez pas et qui vous plaît beaucoup résonne dans le bus. En quelques secondes, l'algorithme de l'application Shazam vous trouve son titre, que vous vous empresserez de télécharger chez vous.

NON NÉGOCIABLE

21 h 30

En gentleman, vous commandez un Uber à votre nouvelle amie afin qu'elle rentre chez elle en sécurité. Pas de chance, la course sera majoritairement à votre avantage à faire confiance aux algorithmes. Aucun moyen de négocier, le prix est fixé dès le départ par l'application. Quant à vous, vous vous contentez du bus, qui vous ramènera chez vous en quinze minutes, comme l'application RATP vous l'avait indiqué.

DE MOI CE QUE TU AIMES

16 h 30

Nouvelle pause. Vous faites un tour sur Amazon. Tous les articles que vous avez précédemment consultés apparaissent à l'écran, et la bannière « Vous aimerez peut-être aussi » attire votre attention. La tentation est grande, et vous finissez par passer commande. Avant de vous remettre au travail, vous regardez furtivement votre téléphone : la demoiselle à qui vous parlez depuis quelques jours sur Tinder vous propose d'aller boire un verre le soir même.

FILM SUR MESURE

22 heures

Après avoir préparé un dîner, toujours aidé par WeCook, vous vous affalez sur votre canapé. Rien d'intéressant à la télévision. Vous avez envie d'un bon film, mais ne savez lequel louer. Plus besoin d'aller sur Allociné, Netflix est alors votre allié. Son algorithme de recommandation vous propose, d'après vos goûts, une sélection de films.



Google



amazon



◀ Une planète
plus intelligente



Watson


tripadvisor®



Spotify

NETFLIX

 Withings®
Inspire health


GROUPE
SEB

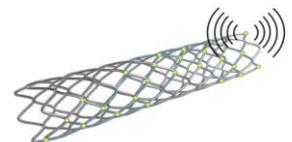
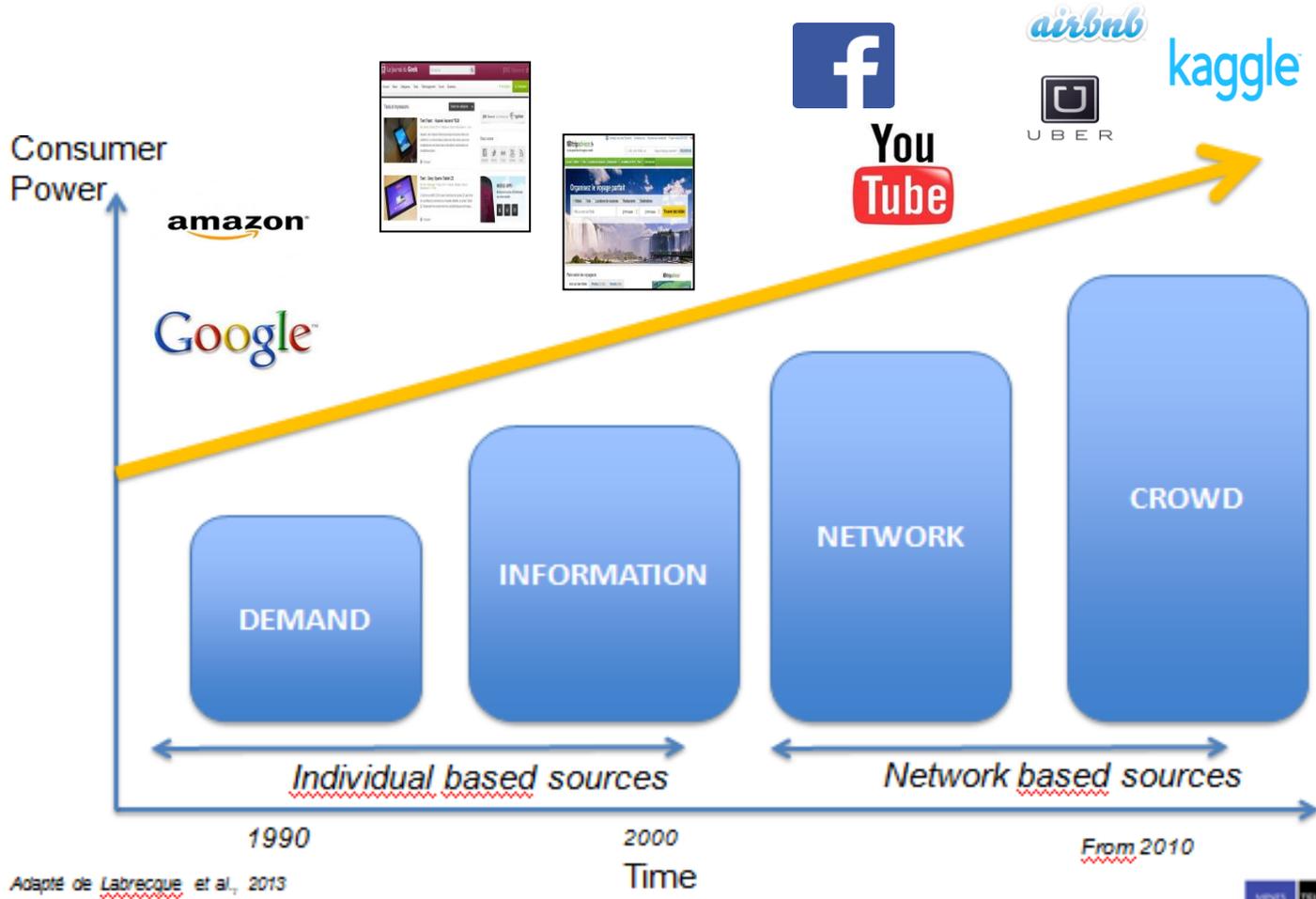


La question de la gouvernance des algorithmes est liée à l'évolution des technologies et de la société numérique

D'un processus de diffusion d'information à la production de données



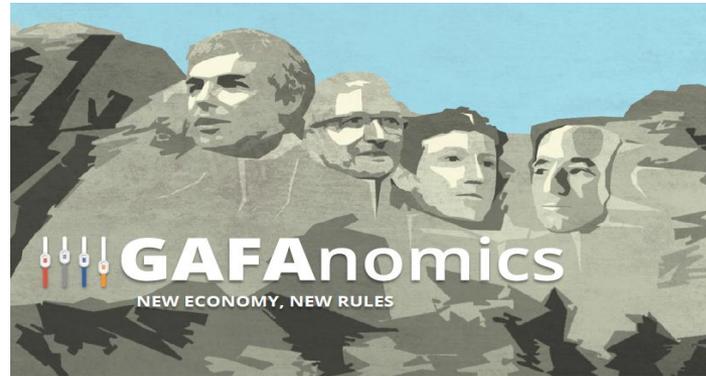
une encapacitation des individus à produire des données



Adapté de Labrecque et al., 2013

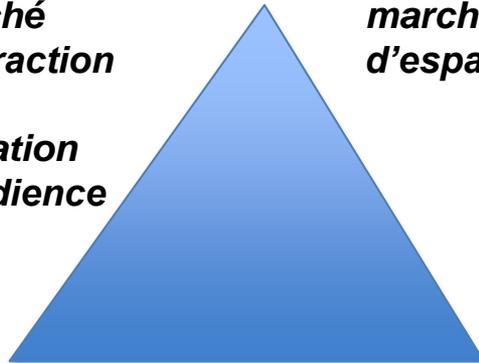


La dominance des GAFA, usines informationnelles et ALGOWORLD



Face 1:
marché
d'attraction
et de
captation
d'audience

Face 2:
marché de vente
d'espaces publicitaires



Face 3: propres services
concurrents des clients

	g	🍏	f	a
TELECOM & IT	Fiber	Apple Slim	WhatsApp	Cloud Drive
HEALTH	Calico	HealthKit	Move	amazon.com Marketplace
RETAIL	Shopping express	iBeacon	Facebook "Buy" Button	amazon fresh Grocery Delivery
ENERGY & UTILITIES	nest Smart home	Solar Power	Internet.org project	Fulfillment by Amazon
MEDIA & ENTERTAINMENT	Play	iTunes Radio	Oculus	twitch Gamers Video platform
FINANCIALS	Wallet	Apple Pay	Friend-To-Friend Payment ¹	amazonpayments Payments API
MOBILITY, TRAVEL & LEISURE	Car	CarPlay	Messenger + Uber Integration ¹	amazon cloud player + Media app for connected cars

¹ Rumoured to be launched soon

Le medium algorithmique selon Pierre Lévy

■ Société data-centrique

- Communication stigmergique
 - Les personnes communiquent entre elles en modifiant leur environnement commun: les données numériques en ligne
- Modèle data-centrique de la communication
 - Production (algorithmes / logiciels)
 - Acheminement vers les destinataires pertinents (algorithmes de recommandations)
 - Fouille (algorithme des moteurs de recherche, de sélection, de rangement)
 - Analyse (algorithmes prédictifs)

Les limites du medium algorithmique contemporain selon Pierre Lévy

■ Les limites cognitives

- Éducation numérique des individus

■ Les limites sémantiques

- Ex Google traduction n'a pas ouvert de communication translinguistique

■ Les limites du positivisme statistique

- Transformer les torrents de données en fleuves de la connaissance mais quels sont les données pertinentes, catégoriser les patterns
- Corrélacion ne signifie pas causalité
- Nécessité d'avoir de s'appuyer sur des théories pour comprendre

Stade de développement: où en est-on?

- **Les algorithmes sont opaques pour la majorité des internautes ou des individus**
- **Il es impossible de produire de la connaissance fiable au moyen de méthodes secrètes**
- **On est dans une phase transitoire du médium algorithmique: il faut explorer la manipulation automatique des symboles et les algorithmes**

La structure algorithmique

■ Une recette de cuisine



Données



*Codage, opérateurs,
conteneurs, instructions*



Résultat

- **Processus structuré**
- **Suite d'instructions dans un contexte donné**
- **Avancement pas à pas**
- **Finalité**

La structure algorithmique

■ Codage

- Codage des données de manière adéquate et uniforme (texte, son, image)
- Système de codage

■ Opérateurs

- Outils mécaniques, processeurs de données
- Adéquation entre le codage des données et les opérateurs

■ Conteneurs

- Entrepôts de données
- Système d'enregistrement logique: écriture, effacement, lecture

■ Instructions

- Ensemble de règles qui organise la circulation des données entre les conteneurs et les opérateurs
- Ensemble de tests (si...alors...)
- Résultat lorsqu'un test a déterminé que le traitement était terminé

Les algorithmes sont des acteurs

Les algorithmes utilisés régulièrement accomplissent des tâches:

- Calcul
- Codage
- Classement
- Classification
- Filtrage
- Optimisation
- Evaluation
- Routage
- Ordonnancement
- Recherche
- Tri
- Stockage
- Vérification

Table 1. – adapted from Diakopoulos (2014) and Latzer et al (2015).

Function	Type	Examples
Prioritization	General search engines	Google, Bing, Baidu
	Special search engines	Genealogy; image search; Shutterstock; social media; Info.com
	Meta search engines	Yummly
	Semantic search engines	Quora, Ask.com
	Questions & answers services	
Classification	Reputation systems	Ebay, Uber, Airbnb
	News scoring	Reddit, Digg
	Credit scoring	CreditKarma
	Social scoring	Klout
Association	Predictive policing	PredPol
	Predicting developments and trends	ScoreAhit, Music Xray, Google Flu Trends
Filtering	Spam filter	Norton
	Child protection filter	Net Nanny
	Recommender systems	Spotify, Netflix
	News aggregators	Facebook News Feed

Public good algorithms

Table 2. Public goods algorithms

Public good	Example	Function(s)	Purpose
Justice	Recidivism models for sentencing and parole	Classification Association	To assess possibility and associated risks of recidivism relative to justice
Public safety	Predictive policing	Prioritization Association	To locate areas of ongoing and burgeoning criminal activity in a given place
Fair access to finance	Credit and loan access	Classification Association	To provide fair access to finance options, particularly for vulnerable populations
Fair employment	Employee hiring	Prioritization Filtering	To promote fair employment process and prevent discriminatory bias

Quels sont les questions éthiques posées par les algorithmes?

Les données

- « Objets d'un design humain » (Crawford)
- Qualité
- Biaisées
- Incomplètes
- Vulnérables

Traitement:

les algorithmes ont des jugements de valeur

- 2 personnes ayant des jugements de valeur différents ont une raison rationnelle de construire les algorithmes différemment

- Exemple:

2 algorithmes classifiant une cellule malade vs non malade

Algorithmes de prise de décisions

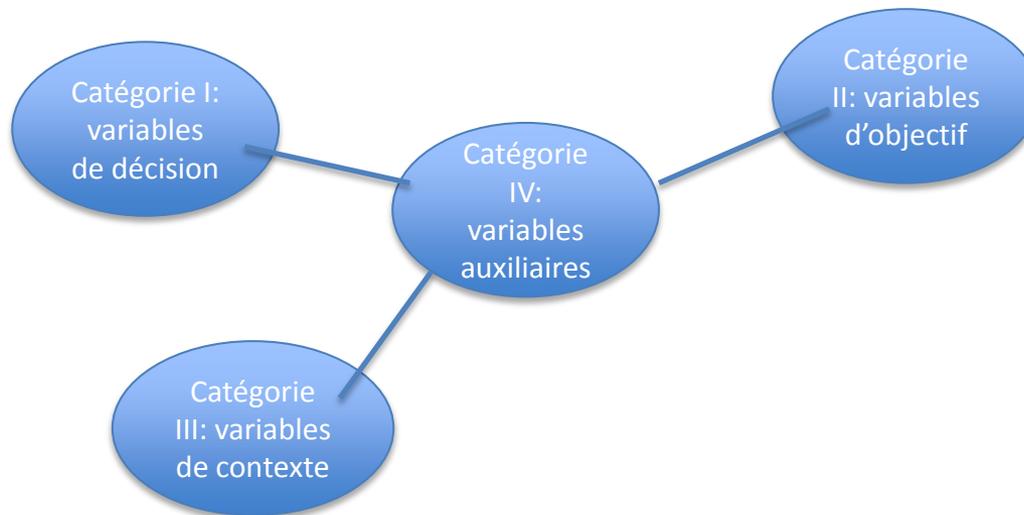
- Les algorithmes comprennent explicitement ou implicitement des jugements de valeur

« Value Laden Algorithms » (1)

Kraemer, van Overveld, Peterson (2011)

■ Design activity:

- Variable de décision (catégorie I): décider qu'une variable devrait avoir une certaine valeur
 - Ex: roue avant du vélo aura 60 cm de diamètre
- Variable objective (Catégorie II): valeur de la variable définissant un objectif atteint
 - Ex: poids de la roue si l'objectif est un vélo léger
- Variables de contexte (catégorie III): densité du carbone
- Variables auxiliaires (catégorie IV): composition de variables (poids est la somme des poids de chaque roue plus du guidon)

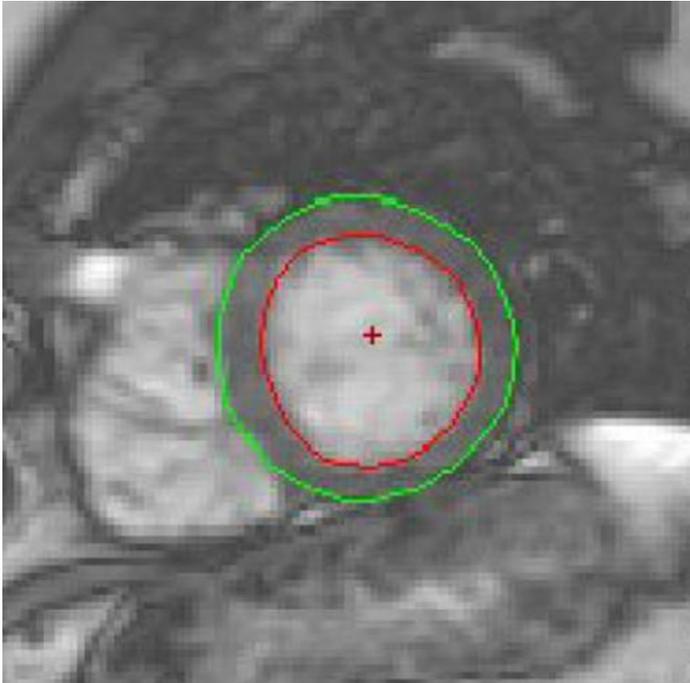


« Value Laden Algorithms » (2)

Kraemer, van Overveld, Peterson (2011)

■ Exemple: technologies imagerie médicale

- Représentation de structures biologiques par un ordinateur le plus précisément possible
- Objectif: diagnostiquer des maladies
- Risque des algorithmes:
 - Faux Positifs: symptôme de maladie alors que ce n'est pas vrai
 - Faux Négatifs: pas d'identification d'un symptôme de maladie réel



Trade off entre le nombre de faux positifs et le nombre de faux négatifs

Jugement de valeur sur le nombre de faux positifs tolérable vs le nombre de faux négatifs.

Les valeurs des variables de catégorie I impactent les variables de catégories II

Comment intégrer les questions d'éthique dans ce cas?

■ Principe de précaution:

- Pour des médecins, il est préférable d'avoir des faux positifs que des faux négatifs,
- Pour les scientifiques, c'est l'inverse, car les connaissances scientifiques sont cumulatives.

■ L'algorithme doit donc être transparent pour les utilisateurs, en particulier sur les choix éthiques effectués

- Proposer à l'utilisateur du scanner de spécifier sa préférence entre faux positifs et faux négatifs
- Former les médecins et les utilisateurs

Le paradoxe des faux positifs dans le domaine du renseignement (*source: Inria*)

- **Objectif: recherche des terroristes dans une population**
- **Tout algorithme a une marge d'erreur. Si la marge d'erreur est de 1% (très faible), l'algorithme identifiera:**
 - 600 000 personnes de faux positifs (sans intention terroriste mais identifiés comme tels) sur une population totale de 60 millions de personnes
 - Si le nombre de terroristes est par exemple de 60, ces vrais terroristes ne représenteront que 0,01% de la population identifiée
- **Phénomène scientifique connu : identification statistique d'évènements rares**
- **Débat lors du projet de loi renseignement en France**

La non transparence des algorithmes

Cas Google Page Rank



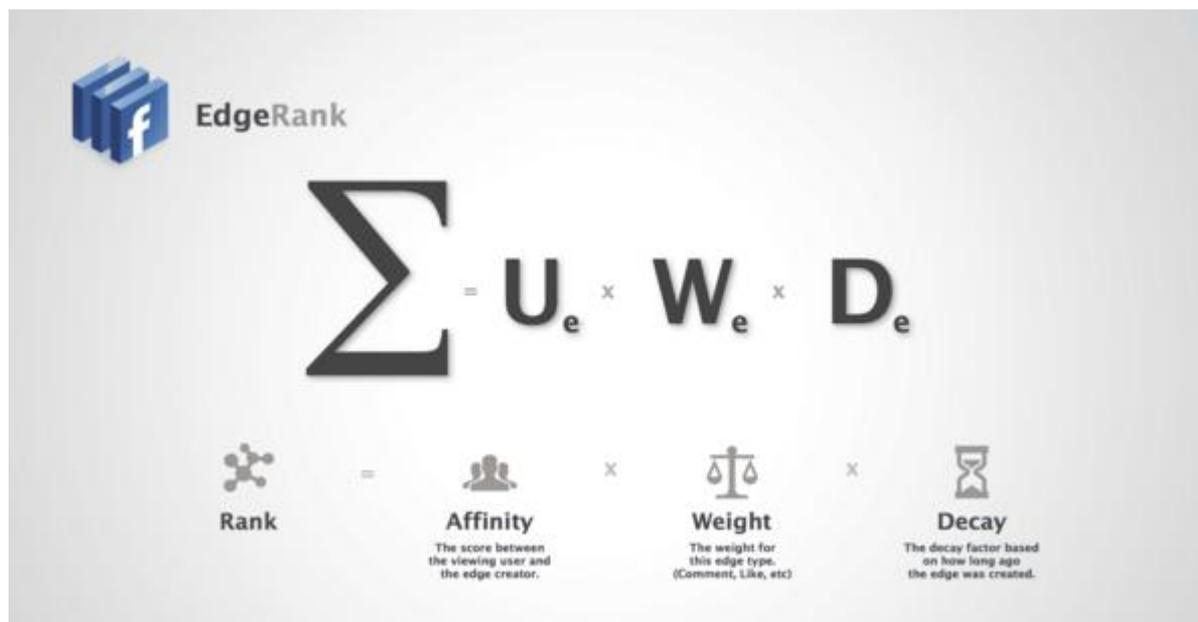
-Publication en 1998 à Stanford de l'article:
The Anatomy of a Large-Scale Hypertextual Web Search Engine, de Brin & Page

-Aujourd'hui, l'algorithme est secret: plus de 200 critères évoluant régulièrement

- Critères liés à l'utilisateur

La non transparence des algorithmes

Cas de Facebook: le Edgerank



-Comment filtrer plus de 200 contenus vus par utilisateur chaque jour ?

-Boîte noire (poids des critères, évolution)

- Deep Learning
Yann LeCun, Director of AI Research, Facebook
« I don't see my work when people talk about algos »

Les risques des algorithmes



JUN 28, 2014 | TECHNOLOGY

Everything We Know About Facebook's Secret Mood Manipulation Experiment

It was probably legal. But was it ethical?



For one week in January 2012, data scientists skewed what almost 700,000 Facebook users saw when they logged into its service.

Some people were shown content with a preponderance of happy and positive words; some were shown content analyzed as sadder than average.

And when the week was over, these manipulated users were more likely to post either especially positive or negative words themselves.

Proceedings of the National Academy of Sciences of the United States of America

CURRENT ISSUE // ARCHIVE // NEWS & MULTIMEDIA // AUTHORS // ABOUT // COLLECTED ARTICLES // BROWSE BY TOP

↑ > Current Issue > vol. 111 no. 24 > Adam D. I. Kramer, 8788–8790, doi: 10.1073/pnas.1320040111



Experimental evidence of massive-scale emotional contagion through social networks

Adam D. I. Kramer^{a,1}, Jamie E. Guillory^{b,2}, and Jeffrey T. Hancock^{b,c}

Author Affiliations ↗

Edited by Susan T. Fiske, Princeton University, Princeton, NJ, and approved March 25, 2014 (received for review October 23, 2013)

Crowd et algorithmes

Le cas TripAdvisor (1)



Vue d'ensemble Avis (714) Questions et réponses (7) Position

Toutes les photos de visiteurs (553)

L'auteur TripAdvisor a aimé Lire les 714 avis

Évaluation de visiteur

Excellent	554
Très bon	91
Moyen	44
Médiocre	16
Horrible	9

"Très bon restaurant"

Très bonne table de Bora Bora. restaurant à faire lors de votre séjour. l'ambiance est typique et le service très bon.

Noemie T Avis publié: il y a 3 semaines

Résumé des notes

Cuisine		Service	
Rapport qualité-prix		Ambiance	

Conseillé pour: Romantique Pour les grandes occasions

■ TripAdvisor:

- 125 millions d'opinions et de reviews
- Algorithme: Popularity Index

■ Classement de tous les hôtels d'une zone géographique

■ TripAdvisor ne donne pas les détails de l'algorithme mais précise:

« this proprietary algorithm incorporates traveler ratings to determine overall traveler satisfaction »

Crowd et algorithmes

Le cas TripAdvisor (2)



Your TripAdvisor popularity ranking is based on traveler reviews and reflects your standing against other businesses within your geographical location. Popularity is important because the closer you are to #1, the more likely it is that travelers will see your property when they search your area. Here are the three primary factors that have an impact on your popularity, and how you can improve:

Quality of Reviews

Provide a stellar guest experience, use prior feedback to continually improve and keep those positive reviews rolling in!

Quantity of Reviews

Encourage guests to write reviews and use the free tools in your [Management Center](#) – flyers, Review Express and widgets – to remind them. Asking is a great idea, but offering incentives is against our policy.

Age of Reviews

Recent reviews carry more weight in your ranking, while older reviews count for less over time.

- Les algorithmes supposent que les items comparés soient commensurables, standardisés avec des propriétés définies et stables
- Or, les reviews des utilisateurs sur TripAdvisor sont à la fois subjectives, idiosyncratiques, reflétant des expériences personnelles
- L'Algorithme d'Index de Popularité
 - ignore la subjectivité des reviews, se concentrant sur le score quantitatif donné par les utilisateurs sur plusieurs catégories dont le sens n'est pas donné
 - porte une attention sur l'interdépendance des critères
 - l'ordre des critères et le temps de l'évaluation traduit un certain point de vue
 - éradique les idiosyncrasies, les conflits de reviews entre les utilisateurs
- Biais liés aux votes de type likes ou étoiles

Algorithme de prix et discrimination

WHAT THEY KNOW

Websites Vary Prices, Deals Based on Users' Information

By JENNIFER VALENTINO-DEVRIES, JEREMY SINGER-VINE and
ASHKAN SOLTANI

December 24, 2012

It was the same Swingline stapler, on the same Staples.com website. But for Kim Wamble, the price was \$15.79, while the price on Trude Frizzell's screen, just a few miles away, was \$14.29.

A key difference: where Staples seemed to think they were located.

A Wall Street Journal investigation found that the Staples Inc. website displays different prices to people after estimating their locations. More than that, Staples appeared to consider the person's distance from a rival brick-and-mortar store, either [OfficeMax Inc.](http://OfficeMax.com) or [Office Depot Inc.](http://OfficeDepot.com) If rival stores were within 20 miles or so, Staples.com usually showed a discounted price.

WHAT PRICE WOULD YOU SEE?



Les algorithmes formatent nos représentations (Dominique Cardon)

■ Popularité, mesures d'audience

- Dénombrer les clics des internautes et ordonner la popularité des sites
- *Représentation traditionnelle* du public, des médias ou des électeurs

■ Famille des mesures issue du PageRank

- Classement de l'information, au-dessus du Web afin de hiérarchiser l'autorité des sites avec des liens hypertextes qu'ils s'échangent
- Mesure *méritocratique*, qui prétend isoler les excellents des médiocres à partir du jugement des autres

■ Mesures de réputation (sites de notations, réseaux sociaux)

- Valoriser la réputation de personnes ou de produits, à l'intérieur du Web
- Statistiques typiques du *benchmark*

Les questions éthiques posées par les algorithmes (1)

- Les algorithmes font des choix sur l'information qu'ils utilisent, qu'ils affichent ou qu'ils cachent, ce qui les rend très puissants
- Ces choix ne sont jamais effectués dans le vide et reflètent des hypothèses et des idées à la fois conscientes et subconscientes de leurs créateurs.
- Les algorithmes ont des points de vue
- Les algorithmes sont opaques (black box algorithmic culture) , ce qui génère une perception de faible objectivité. Pourtant la puissance de ces algorithmes est leur capacité à se présenter comme des faits objectifs. Il se positionnent comme représentant la vérité (Slogan de TripAdvisor: « Get the truth,Then Go »)

Les questions éthiques posées par les algorithmes (2)

- **Est-ce que ces technologies instrumentalisent l'homme plutôt que les aident?**
 - Pour échapper à l'instrumentation, faut-il nous reposer la question du sens de la vie ?

Ethique à Nicomaque d'Aristote:

 - Eudémonisme: béatitude, bien être, « bonne vie » etc..
 - Cela nécessite des vertus: développer ses compétences, ses pratiques, sa capacité à agir délibérément...
 - Quelle vision les algorithmes conçus par les scientifiques proposent de la « bonne vie »?
- **Technologies du bon comportement**
 - Leur grand récit global est de dire que tout est de la faute de l'individu
 - Si on mange mieux, si on conduit de manière responsable etc... tout irait mieux
 - Lorsqu'on fait reposer le problème sur les individus, nous détournons l'attention des solutions systémiques

Les questions éthiques posées par les algorithmes (3)

- Enfermement des individus:
 - « Si les personnes ont des comportements monotones, si elles ont des amis qui ont les mêmes idées et les mêmes goûts, si elles suivent le même trajet, les calculateurs les enferment dans leur régularité » (D. Cardon)
 - Ex: Facebook
- Le mythe de l'Etat Parfait:
 - Par de bons calculs, sur de bonnes données, les gouvernements seraient plus justes, prendraient de meilleures décisions
 - Ex: discours dénoncé par Evgeny Morozov dans son ouvrage *To Save Everything, Click Here*



```
// Boucle optimisée (FirstStep / NextStep / IsFree / GetSource)
while (Index < 64 && !Save2Do) {
  y = (pcaseNB[x])->FirstStep(Src);
  if (pcaseNB[y]->IsFree(Index+1)) { // L'index attendu
    Src = x;
    x = y;
    Index += 1;
  } else {
    while (1) {
      while ((y = pcaseNB[x]->NextStep()) != -1) {
        if (pcaseNB[y]->IsFree(Index+1)) {
          Src = x;
          x = y;
          Index += 1;
          break;
        }
      }
      if (y == -1) {
        Index -= 1;
        x = pcaseNB[x]->GetSource();
      } else {
        break;
      }
    }
  }
}
}
```



QUE FAIRE?

Privacy enhancing technologies & systems

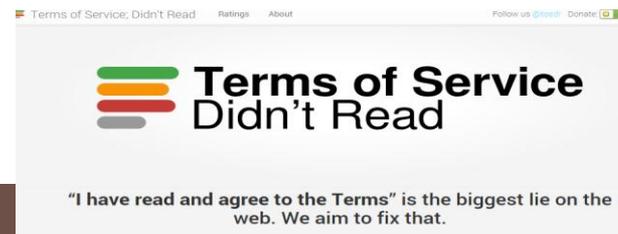


Open algorithms: A new paradigm for using private data for social good

By Thomas Roca, Emmanuel Letouzé | 18 July 2016

OANSO (Open Algorithms for National Statistical Offices) project: Open Software Platform and Open source Algorithms to help NSOs access to and/or collect better data and create new ways of understanding it. It would enable them to collect indicators based on data collected by private companies (ex: mobile phones or banks) in developing countries, in a safe, private and secure way.

Privacy-enhancing models:
Differential privacy,
private information retrieval,
Syntactic anonymity,
PIMS,



Développement de logiciels de tests



Home » News & Events » Events Calendar » PrivacyCon

PrivacyCon



JAN 14, 2016

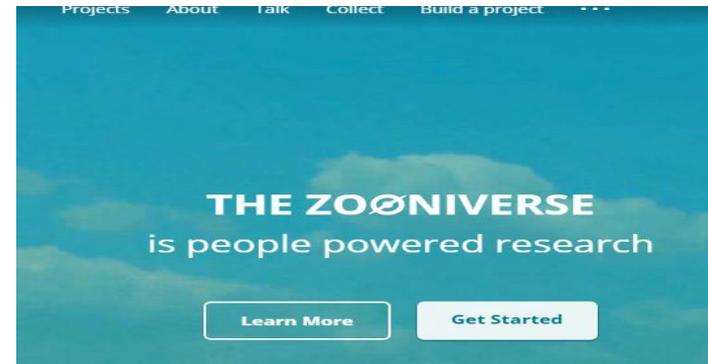
-**AdFisher** (Universités Berkeley et Carnegie Mellon) est un système qui détermine les informations récoltées par les logiciels "traqueurs en ligne" et leurs corrélations avec les annonces publicitaires proposés, lors de la navigation des Internauts

-**Sunlight** (Université Columbia) est un service en ligne qui étudie le ciblage des annonces publicitaires à partir d'informations des boîtes de courriel Gmail des Internauts

-**FairTest** (Université Columbia) est une boîte à outils pour les développeurs de logiciel, qui permet de tester la qualité d'un algorithme, en détectant des associations non souhaités.

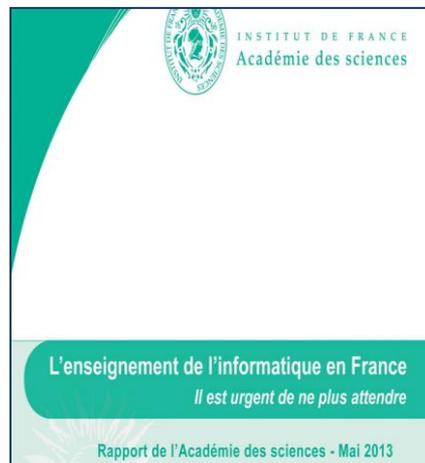
- **Mobilitics** (Inria + Cnil)

Plateformes collaboratives de sciences citoyennes



Data literacy

■ Littératie numérique:



■ Data pop alliance:

- MIT media lab, Harvard humanitarian initiative, ... / promote a people centered big data revolution
- Reports:
 - Beyond data literacy: reinventing community engagement and empowerment in the age of data (Oct 2015)
 - Leveraging Algorithms for Positive Disruption: On data, democracy, society and statistics (dec 2015)

Repenser les cas d'usage



Penser expérience utilisateur et non cas d'usage

Penser la valeur, l'intelligence collective

Le cas du Smart Home

- **IBM en 2010 a défini 4 fonctions pour le smart Home:**

- 1) entertainment and convenience
- 2) energy management
- 3) safety and security
- 4) health and welfare

Déclinaison en bénéfices marché « confort et praticité »,
« dommage et danger » etc...

En 6 ans, rien n'a changé...faible appropriation des utilisateurs

- **Mais...**

Est-ce-que les cas d'usage définis par les scientifiques ou l'industrie comme “confort et praticité”, “dommage et danger” représentent les besoins sous-jacents correspondant à la valeur pour le consommateur?

Cas d'usage / d'utilisation

- En génie logiciel et en ingénierie des systèmes, un cas d'utilisation, défini simplement, permet de décrire **une séquence d'événements** qui, pris tous ensemble, définissent un système **faisant quelque chose d'utile** .

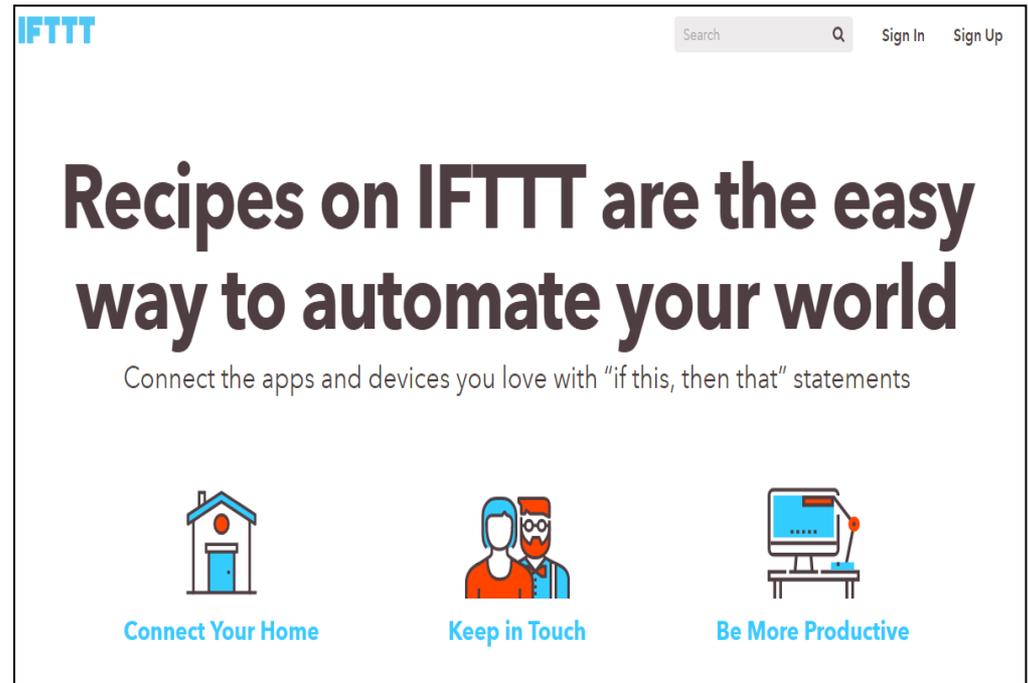
Chaque cas d'utilisation contient un ou plusieurs scénarios qui **définissent comment le système devrait interagir avec les utilisateurs (appelés acteurs) pour atteindre un but ou une fonction spécifique d'un travail**. Un acteur d'un cas d'utilisation peut être un humain ou un autre système externe à celui que l'on tente de définir.

https://fr.wikipedia.org/wiki/Cas_d%27utilisation

- Il est raisonnable de se demander si une approche top down ou bottom up a le plus de sens pour les utilisateurs. Est-ce que les consommateurs veulent prendre du temps pour créer leur propre cas d'usage?

IFTTT.com case (If This Then That) (1)

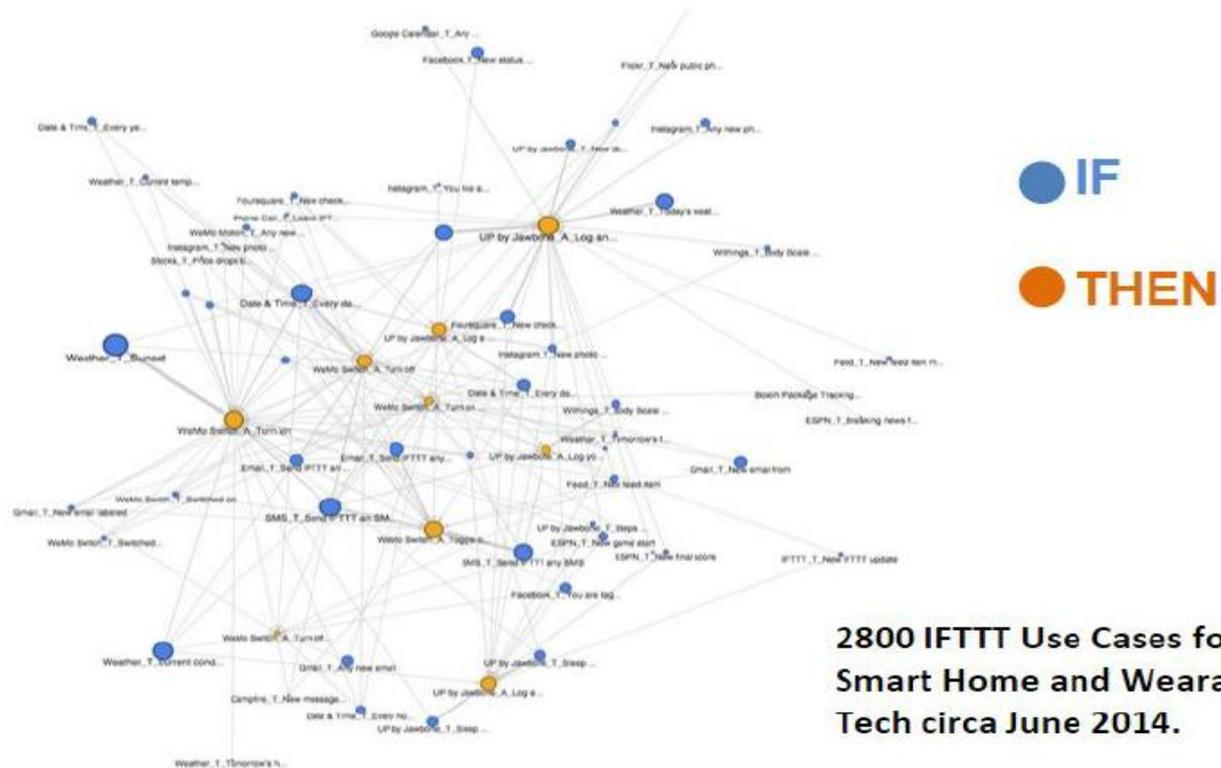
- Fondé en 2010 par Linden Tibbets
- Permet aux utilisateurs de créer des chaînes d'instructions conditionnelles, appelées "recipes"
- En 2012, un million de tâches ont été créées
- 364 services
- Fin 2014, IFTTT est l'une des plus fortes valorisations mondiales d'entreprises IoT (170 Millions de \$)



L'expérience dépend des interactions

IFTTT.com case (2)

Intégrer l'utilisateur dans la définition des séquences d'évènements sur le Smart Home



Passer des cas d'usage à la complexité de l'expérience

- Les cas d'usage supposent que les utilisateurs ont un objectif clair, ce qui n'est pas toujours le cas:
 - La plupart des cas d'usage Smart home sont perçus comme vague et ennuyeux (Ex: « safety & security »)
- Les développeurs ne peuvent prédire à l'avance les buts de chaque utilisateur dans l'usage de leur smart home
- Si on se focalise sur les cas d'usage, on passe à côté des usages idiosyncratiques que les consommateurs peuvent avoir.
- Il est préférable d'intégrer la complexité de l'expérience plutôt que des cas d'usage

Réfléchir à une gouvernance?

Les algorithmes sont écrits par des individus, donc on a le pouvoir de les faire évoluer

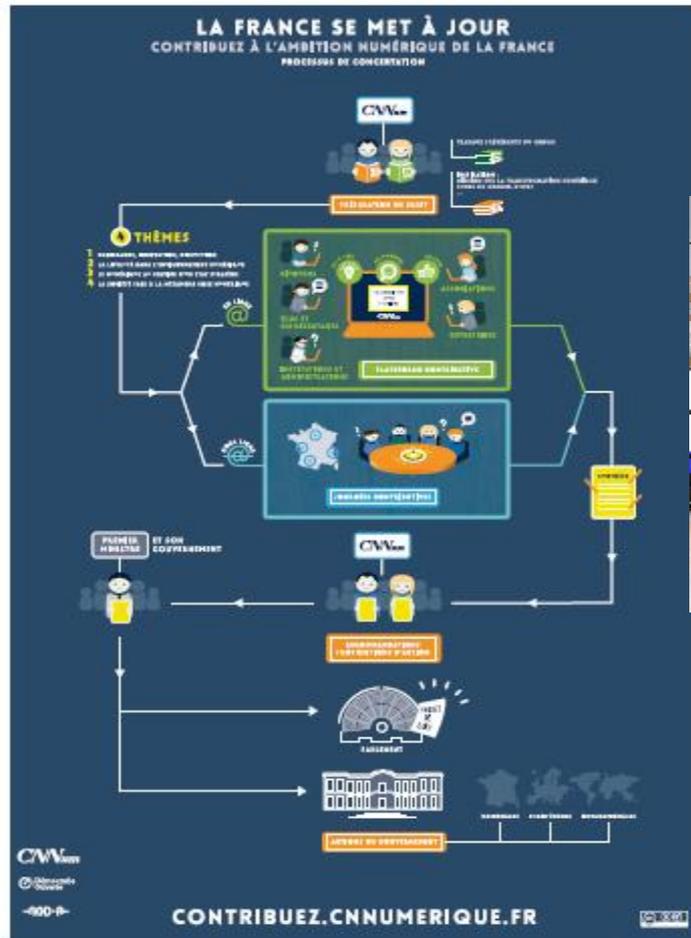
On peut les utiliser là où ils sont transformateurs positifs de la société, il peut y avoir des usages citoyens

Il existe des logiciels libres, le problème porte sur l'accès aux données massives (données concentrées chez un nombre limité d'acteurs privés)

Rien ne nous interdit de fixer des règles

*Serge Abitboul, Les algorithmes régissent ils leurs vies?
France culture, 27/09/ 2016*

La loyauté dans les travaux du Conseil National du Numérique



La France se met à jour

AMBIITION NUMÉRIQUE
#contribuez

CONCERTATION NATIONALE
SUR LE NUMÉRIQUE
Présentation

contribuez.cnumerique.fr



Pourquoi introduire la loyauté des algorithmes? (1)

- **Réintroduction du profilage par recoupement de données**
 - Discriminations prohibées par la loi dans les domaines de l'emploi, du logement, de l'éducation, de l'accès aux biens et aux services: origine, sexe, situation de famille, grossesse, apparence physique, patronyme, état de santé, handicap, caractéristiques génétiques, mœurs, orientation sexuelle, âge, opinions politiques, activités syndicales, appartenance ou non-appartenance à une ethnie, nation, race ou religion déterminée
 - Avec le big data et la fouille de données, on peut effectuer ce profilage par l'analyse d'autres variables. Cas simples: sexe ou religion (via recoupements avec des données d'achats, de commentaires sur des sites de cuisine)
- **Risques d'enfermement des internautes dans une personnalisation des services en fonction de leurs goûts ou dans des sphères d'opinions supposées.**

Pourquoi introduire la loyauté des algorithmes? (2)

- **Risque d'une confiance excessive dans les choix préconisés par les algorithmes**
 - Les algorithmes orientent les choix des individus sans que ceux-ci en aient véritablement conscience
 - Tout postulat est un construit social, donc discutable
- **Risque « solutionniste »**
 - Société qui encourage le recours systématique à des solutions algorithmiques masquant la complexité des enjeux socio-économiques qui requièrent d'autres types d'interventions.



Introduire un droit à la transparence et à la conformité aux engagements de la plateforme, des critères de personnalisation, de classement et de référencement

- **Reconnaître un droit à l'information sur les critères utilisés par les algorithmes**
- **Imposer la séparation claire des suggestions organiques, qui font le cœur du service rendu, des suggestions sponsorisées, fruits d'accords commerciaux**

Obtenir de la part des acteurs contre l'utilisation discriminante des données dans les politiques de prix

- **Se doter de capacités d'observations renforcées**
 - Détecter l'utilisation de critères discriminants dans la fixation des prix

- **Clarifier les critères de distinction entre critères licites et illicites de différenciation du prix**

- **La tendance croissante à la personnalisation de services de mutuelles, assurances, organismes de crédit, met en cause la mutualisation des risques et le modèle social français. Il faut mener une étude d'impact avec les acteurs concernés**

Garantir l'accès à une information plurielle

- **Demander aux grandes plateformes de respecter des engagements de pluralisme de l'information délivrée et d'offrir la possibilité de désactiver la personnalisation des résultats de leur service**
- **Introduire dans la formation à la littératie numérique (à l'école, comme dans l'enseignement supérieur ou la formation professionnelle) une compréhension des fonctions d'orientation des algorithmes**

Normaliser l'accès aux ressources et espaces de visibilité clefs des plateformes incontournables

- **Instaurer une obligation d'information préalable dans des délais raisonnables en cas de modifications majeures (accès aux API, critères de classement dans les algorithmes)**
- **Appliquer le principe de non-discrimination dans le référencement, sauf en cas de considérations légitimes, vérifiables par des tiers et conformes à l'intérêt des internautes (personnalisation, pertinence, qualité, etc...)**

Une agence européenne de notation de la loyauté qui s'appuie sur un réseau de contributeurs

■ Double mission:

- Rendre accessible via un point d'entrée unique toute une série d'informations déjà rassemblées par les observateurs et outils existants
 - Remontées d'information (assoc. consommateurs, acteurs de l'internet citoyen, multitude, etc...)
 - Outils de tracking de circulation de données, de comparaison et lisibilité de CGU ..
- Ouvrir un espace de signalement de pratiques contraires à la loyauté et à l'autodétermination des individus

■ 2 leviers d'action

- Réputation via des labels de plateforme ou algorithmes loyaux
- Prise en compte par les investisseurs privés et publics

■ Mobilisation de la rétro-ingénierie

Un corps d'experts en algorithmes (algorithmistes) mobilisable sur demande exclusive d'une autorité de régulation

- **Création d'un cursus de formation à la croisée entre droit et science**
- **Expérimentation, avec l'aide d'instituts de recherche, de ce que pourrait être techniquement le contrôle d'un algorithme, afin d'en mesurer la portée et les limites**
- **Créer un poste *d'administrateur général des données* au sein des conseils d'administration des entreprises afin de veiller aux impacts éthiques des algorithmes mobilisés par les services**

Rapports et missions gouvernementales

Big Data: A Report on Algorithmic Systems, Opportunity, and Civil Rights

Executive Office of the President

May 2016



SECRETARIAT D'ETAT CHARGE DU NUMERIQUE

Paris, le

La Secrétaire d'Etat

à **27 OCT. 2016**

Monsieur le Vice-président du Conseil général
de l'Economie, de l'Industrie, de l'Energie et
des Technologies

Mission relative aux modalités de régulation des algorithmes de traitement des
contenus.



Merci de votre attention

**christine.balague@it-sudparis.eu
[@balague](https://www.linkedin.com/company/balague)**



deux écoles filiales et deux partenaires stratégiques