Institut Mines-Télécom
Business School

ENJEUX ETHIQUES DE L'INTELLIGENCE ARTIFICIELLE

Christine Balagué
Professeur
Titulaire de la Chaire Good in Tech
Membre du bureau de DATAIA
Ex Vice-Présidente du Conseil National du Numérique et membre de la Cerna

LES ENTREPRISES LES PLUS INNOVANTES EN 2018

EXHIBIT 1

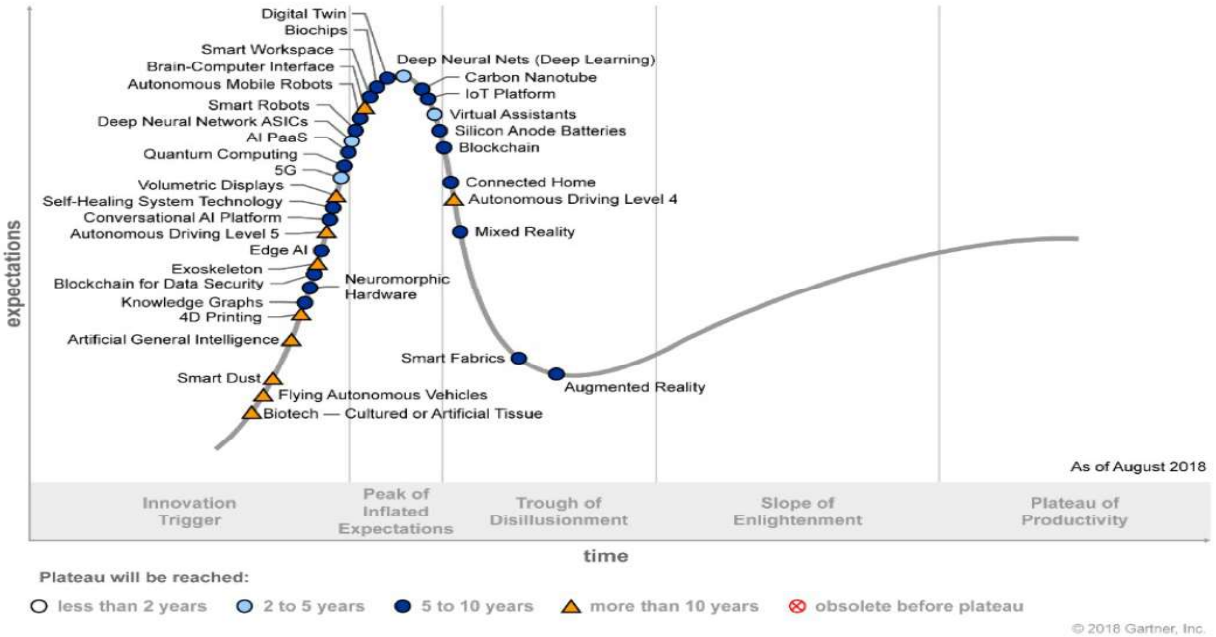
The Most Innovative Companies of 2018

1 Apple	11 Airbnb	21 Siemens	31 Intel	41 3M
2 Google	12 SpaceX	22 Unilever	32 NTT Docomo	42 SAP
3 Microsoft[1]	13 Netflix	23 BASF	33 Daimler[3]	43 DuPont
4 Amazon	14 Tencent	24 Expedia	34 AXA	44 InterContinental Hotels Group
5 Samsung[2]	15 Hewlett-Packard	25 Johnson & Johnson	35 Adidas	45 Disney
6 Tesla	16 Cisco Systems	26 JPMorgan Chase	36 BMW	46 Huawei
7 Facebook	17 Toyota	27 Bayer	37 Nissan	47 Procter & Gamble
8 IBM	18 General Electric	28 Dow Chemical	38 Pfizer	48 Verizon
9 Uber	19 Orange	29 AT&T	39 Time Warner	49 Philips
10 Alibaba	20 Marriott	30 Allianz	40 Renault	50 Nestlé

Source: 2017 BCG global innovation survey.
[1]Includes Nokia. [2]Includes all Samsung business groups (electronics and heavy industry). [3]Includes Mercedes-Benz.

BCG report: The Most Innovative Companies 2018: Innovators Go All In On Digital.

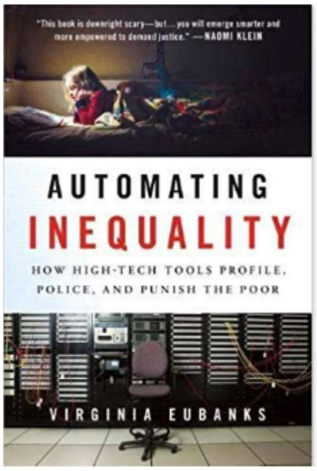
GARTNER HYPE CYCLE 2018



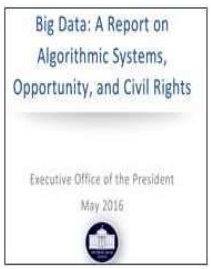
UNE POLARISATION DES DEBATS SUR IA ET SOCIETE

Vision positive	*Vision négative*
Accès à un nombre gigantesque de connaissances grâce aux moteurs de recherche comme Google	Modèles bi-faces à collecte massive de données en temps réel
Bénéfices des réseaux sociaux: communautés, lien social, réaction aux évènements, mobilisation	Fake news, manipulation de l'opinion, déstabilisation des démocraties
Simplification de la vie à la maison (sécurité, énergie, santé, etc...) via le Smart Home ou l'Intelligence Artificielle	Atteinte à la privacy, à l'intimité, à la dignité
Autonomie et capacité d'agir accrues des individus (empowerment)	Surveillance généralisée
Nouveaux métiers (Data scientist, community manager etc...)	Remplacement de l'emploi humain par des machines
Libération des tâches et emplois répétitifs, relation flexible au travail	Précarité et micro-tâches (Digital Labor, Gig Economy)
Progrès liés aux algorithmes (prédiction en santé, optimisation de la gestion de l'énergie, orientation des élèves, systèmes de recommandation...)	Opacité des algorithmes, biais, discrimination
Plus de liberté d'expression et d'innovation	Création d'inégalités (Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor, 2017, Virginia Eubanks)

IMPACT SUR LA SOCIETE: UNE AUTOMATISATION DES INEGALITES



L'IA RESPONSABLE ET ETHIQUE: UN ENJEU POUR LES ETATS ET POUR L'EUROPE



QU'EST-CE QUE L'ETHIQUE?

LES VALEURS FONDAMENTALES DE LA RÉFLEXION ÉTHIQUE

Macro-éthique

Principia : « ce qui vient en premier, ce qui est à la source » ou « ce qui fait autorité »

Principe de justice

Principe d'autonomie

Principe de bienfaisance

Principe de non-malfaisance

Respect de la vie

Utilité
Responsabilité

Proportionnalité
Précaution
Incertitude

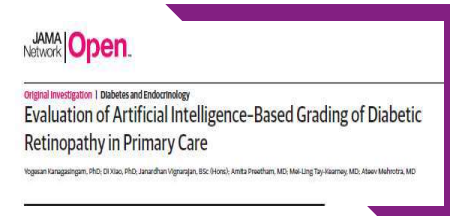
Micro-éthique

Principles of Biomedical Ethics by Tom L. Beauchamp & James F. Childress (2001)

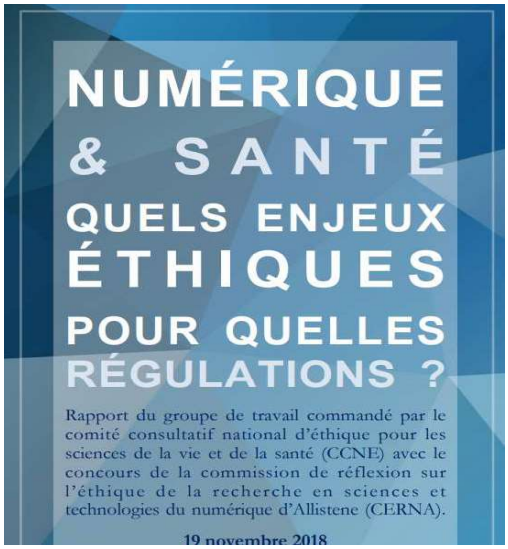
INTELLIGENCE ARTIFICIELLE RESPONSABLE ET ENJEUX ETHIQUES

MEDECINE DATA DRIVEN: QUELLE PLACE POUR L'HUMAIN?

Algorithm permission in the US predicting patients' death in hospitals



RAPPORT CCNE-CERNA NUMERIQUE & SANTE



Enjeux éthiques spécifiques:

- Consentement patient (maladies rares), responsabilité des médecins, enjeux du health data hub, objets connectés)
- Risques des algorithmes: biais, discrimination, exclusion, source d'inégalités
- Médecine algorithmique et prise de décision
- Quelle place pour le patient?
- Anonymisation & re-identification

Quelques recommandations:

- Loyauté, vigilance et réflexivité (rapport CNIL)
- Evolution du CCNE incluant sciences et technologies du numérique, des usages et de l'innovation
- Recherches sur la régulation de la santé numérique
- Robustesse de l'IA en santé
- Garantie humaine

QUELS SONT LES ENJEUX ÉTHIQUES ET DE RESPONSABILITE POSÉES PAR L'IA?

Biais: comment rendre les algorithmes plus justes, moins discriminants?

Explicabilité: comment rendre les algorithmes moins opaques et plus interprétables?

Enfermement vs autonomie: comment laisser l'individu libre de penser et d'agir?

Prise de décision: comment identifier les opinions encapsulées dans les algorithmes, quelles conséquences?

Data, biais et discrimination vs équité, justice

LES CHALLENGES ETHIQUES SUR LES BASES DE DONNEES

Mauvaise qualité des données

Données biaisées

Données incomplètes

Données vulnérables

Sociologie:

Digital Labor et micro travail

MACHINE LEARNING & DISCRIMINATION RACIALE

DISCRIMINATION SELON LE GENRE PAR LES ALGORITHMES

- Performing Google Search, Sweeney (2013) shows that black-sound names receive more displays of an ad about criminal record compared to white-sound names. Datta et al. (2014) confirm gender bias.
- Lambrecht et Tucker (2018) show gender discrimination in STEM jobs ad explained by eyeballs and spillovers.
- Cecere G., Jean C., Le Guel F., and Manant M. (2018) STEM and teens: An algorithm bias on social media: algorithms are discriminatory against girls

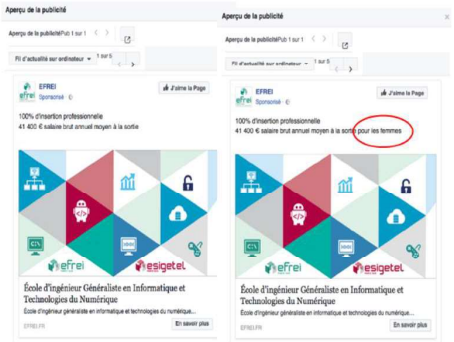


Figure 4: Distribution of Impressions

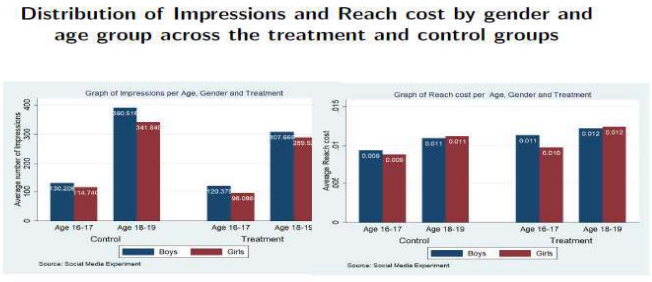


Figure 5: Distribution of Reach costs

LA RECONNAISSANCE FACIALE: UNE REALITE

BuzzFeed News The US Government Will Be Scanning Your Face At 20 Top Airports, Documents Show

TECH

The US Government Will Be Scanning Your Face At 20 Top Airports, Documents Show

"This is opening the door to an extraordinarily more intrusive and granular level of government control."

Davey Alba
BuzzFeed News Reporter

Posted on March 11, 2019, at 9:27 a.m. ET

NTECH LAB

Our algorithms

RIFICATION | FACE IDENTIFICATION | FACE DETECTION | AGE, GENDER | EMOTIONS

Emotion recognition algorithm

Detects 7 primary and 50 compound emotions of a person. Captures and interprets emotional data to deliver actionable insights.

Emotion detection can be adopted across a wide range of use cases and markets including entertainment and media, client satisfaction measurement.

Algorithm's quality is proven by 1st place at EmotioNet Challenge 2017.

EMOTIONS
ASTONISHMENT

LE FIGARO PREMIUM

Politique Société Vox International Economie Culture Figaro Live

Actualité › International

Moscou devient la vitrine de la reconnaissance faciale

Par Pierre Avril | Mis à jour le 16/01/2019 à 18:56 | Publié le 16/01/2019 à 18:42

LES CHALLENGES ETHIQUES DE LA RECONNAISSANCE FACIALE

- Le visage comme matière première, une donnée comme les autres?
- Consentement: perte totale d'anonymat dans l'espace public
- Rôle de l'Etat: quelle finalité? possibilité d'influence et de coercition émanant de pouvoir politique ou économique
- Surveillance et menace démocratique
- Impact des biais des algorithmes et de la mauvaise interprétation des résultats

Opacité et manipulation vs explicabilité et interprétabilité

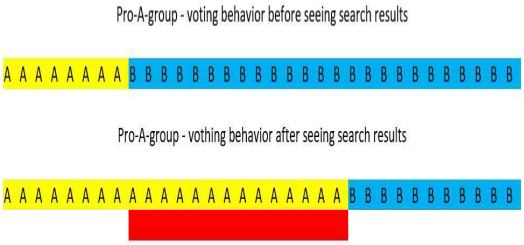
Linky : la CNIL met en demeure Direct Energie à propos des données personnelles

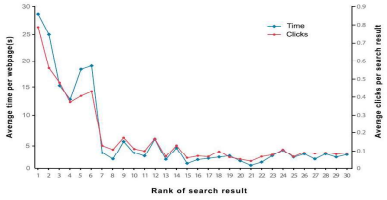
L'entreprise ne respecterait pas la loi imposant d'obtenir le consentement « libre, éclairé et spécifique » avant de collecter des données. Le fournisseur s'en défend.

Par Nabil Wakim et Morgane Tual · Publié le 27 mars 2018 à 10h45 - Mis à jour le 30 mars 2018 à 16h02

EFFET DE MANIPULATION DES MOTEURS DE RECHERCHE

Ranking biaisés des moteurs de recherche peuvent impacter
les préférences de vote de 20% des votants indécis





R. Epstein & R.E. Robertson "The search engine manipulation effect (SEME) and its possible impact on the outcome of elections", PNAS, 112, E4512-21, 2015

THE FAKE NEWS MACHINE

We can manipulate an election with 400 000 $!

Figure 1. The Fake News Triangle

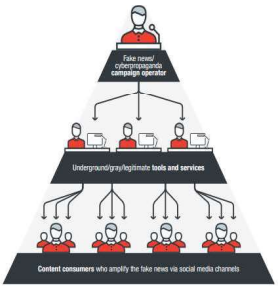
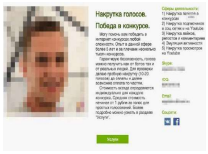
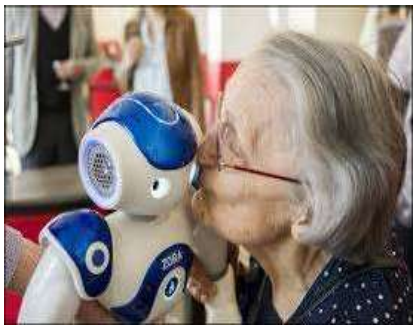


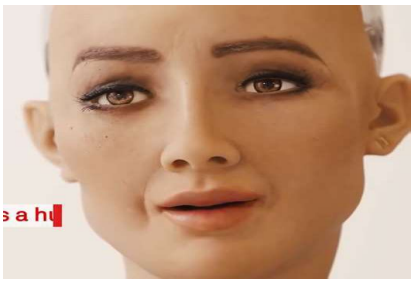
Figure 2. How an operator employs or abuses underground, gray, and legitimate marketplaces to disseminate fake news

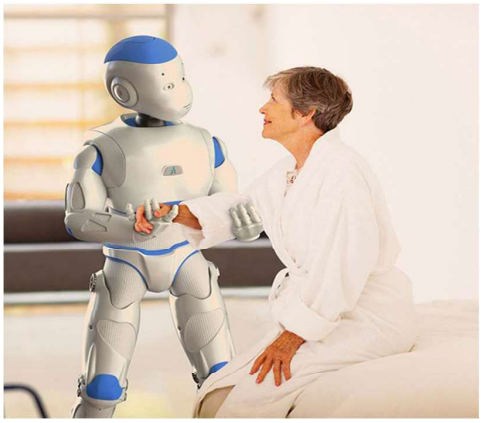


Lion Gu, Vladimir Kropotov, and Fyodor Yarochkin (2017) The Fake News Machine How Propagandists Abuse the Internet and Manipulate the Public, Trendlabs research paper, Trend Micro

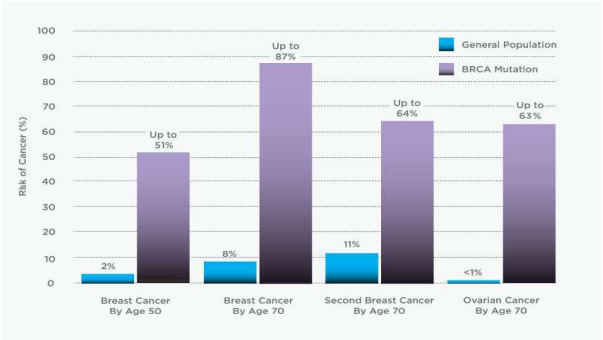
MANIPULATION AFFECTIVE

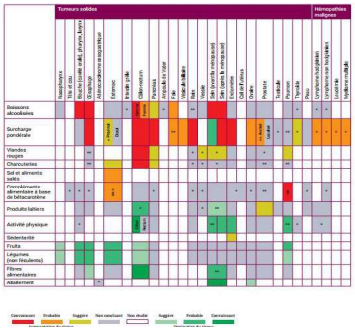






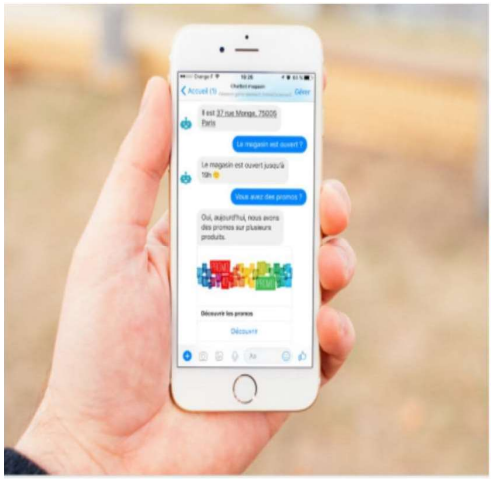
OPACITE DES ALGORITHMES PREDICTIFS DE CANCER: COMMENT COMMUNIQUER LES RESULTATS AUX PATIENTS?



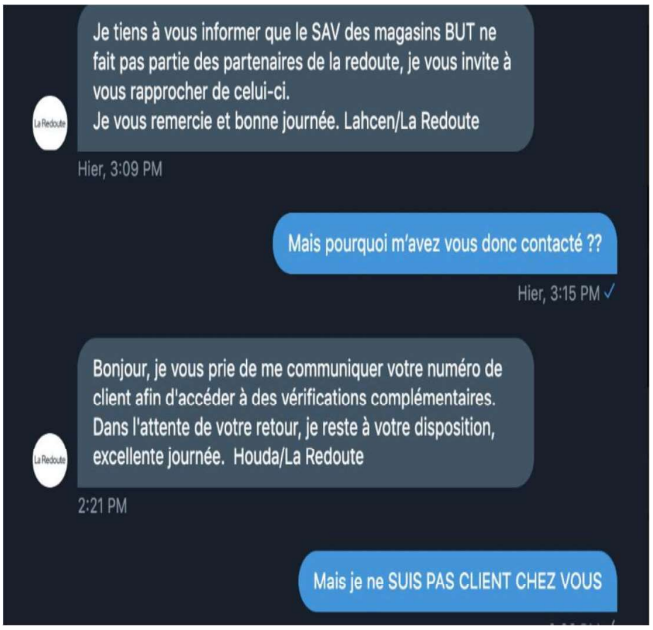


Annual report: june 2017
Institut national du cancer

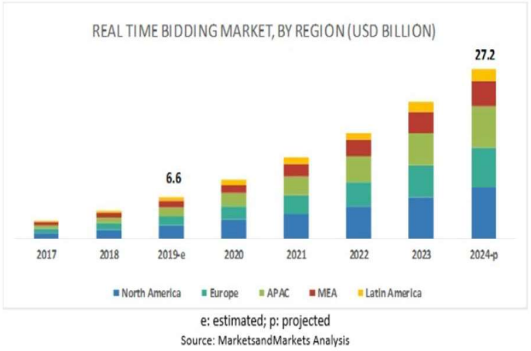
CHATBOT POUR CRM ET PRIVACY



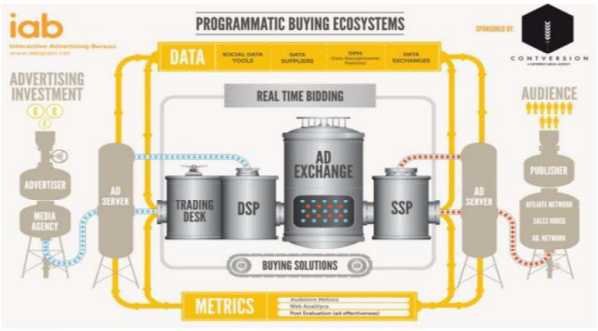
OPACITE DU CHATBOT



LA PUBLICITÉ PROGRAMMATIQUE RTB



e: estimated; p: projected
Source: MarketsandMarkets Analysis



Enfermement des individus, automatisation vs autonomie

Bulles filtrantes et chambre d'echo limitent l'accés à la diversité des opinions

Bakshy E., Messing S., Adamic L.A. (2015), Exposure to ideologically diverse news and opinion on Facebook, Science, vol. 348, issue 239

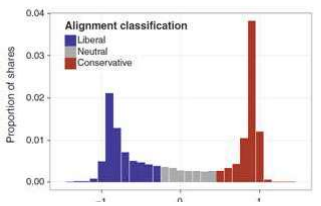
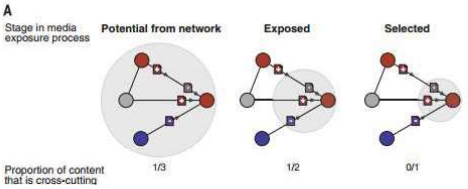


Fig. 1. Distribution of ideological alignment of content shared on Facebook measured as the average affiliation of sharers weighted by the total number of shares. Content was delineated as liberal, conservative, or neutral on the basis of the distribution of alignment scores (details are available in the supplementary materials).



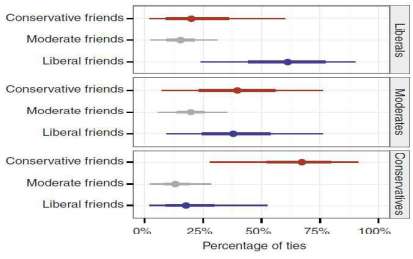
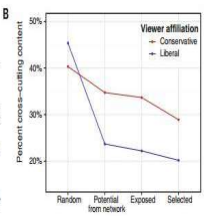
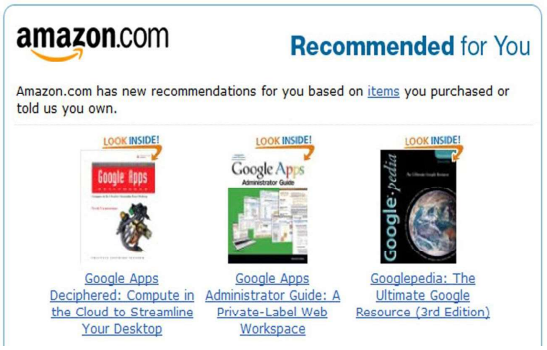


Fig. 2. Homophily in self-reported ideological affiliation. Proportion of links to friends of different ideological affiliations for liberal, moderate, and conservative users. Points indicate medians, thick lines indicate interquartile ranges, and thin lines represent 10th to 90th percentile ranges.

Fig. 3. Cross-cutting content at each stage in the diffusion process. (A) Illustration of how algorithmic ranking and individual choice affect the proportion of ideologically cross-cutting content that individuals encounter. Gray circles illustrate the content present at each stage in the media exposure process. Red circles indicate conservatives, and blue circles indicate liberals. **(B)** Average ideological diversity of content (i) shared by random others (random), (ii) shared by friends (potential from network), (iii) actually appeared in users' News Feeds (exposed), and (iv) users clicked on (selected).

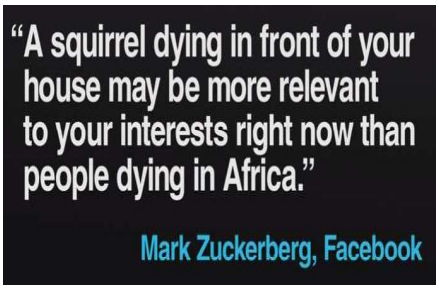


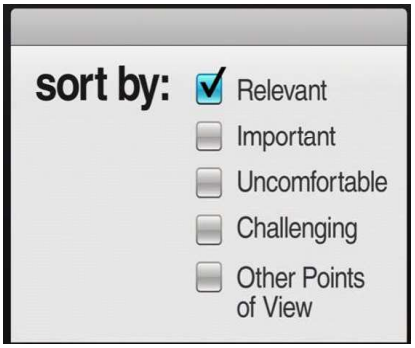
SYSTEMES DE RECOMMANDATIONS



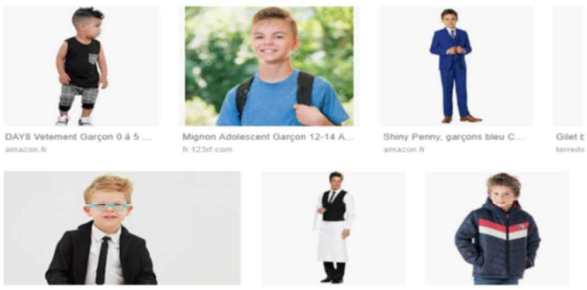
Prise de décisions vs opinions encapsulées dans l'algorithme

ALGORITHMES DE FILTRAGE FONDES SUR LE CRITERE DE PERTINENCE





RESULTATS DE GOOGLE IMAGE



Garçon

Fille

STEREOTYPES DU GENRE DANS LE LANGUAGE

BIAIS DES BASES DE DONNEES TRADITIONNELLES

- **APPELLATIONS:**
  - Père = homme de famille
  - Femme juge
- **ORDRE:**
  - Mr & Mrs
  - Frères et sœurs
  - Maris et femmes
- **METAPHORE:**
  - Femmes sont hystériques, bavardes, autoritaires
  - Hommes sont plus joviaux, grégaire, attentif
- **PRESENCE DES FEMMES DANS LES BASES DE DONNEES:**
  - Un bon indicateur du biais du genre
  - Bristish National Corpus: Il y a plus de Mr que de Mrs + Miss combinés

→ **Les algorithmes qui apprennent sur ces bases de données traditionnelles reflètent ces biais**

DES ASSISTANTS VOCAUX PERSONNELS AVEC VOIX FEMININE…

COMMENT DEVELOPPER UNE IA RESPONSABLE ?

MESURER LES BENEFICES DE L'INTELLIGENCE ARTIFICIELLE…

Advertising : Machine learning can improve ad distribution as it can help identify potential new consumers (Stitelman et al., 2011)

Justice: IA calibrated algorithm in the context of legal court decisions can help reducing criminality of about 24.8% with fairer decisions toward afro-americans and hispanics (Kleinberg et al.,2017)

Health: Use trained algorithm to predict the Restaurant Hygiene inspections reduced inspectors bias (Glaeser et al., 2018)

International trade: Machine translation system has already had a significant effect on international trade on this platform, increasing export quantity (Brynjolfsson et al., 2018)

Housing market: Performance of Different Algorithms in Predicting House Values (Mullainathan and Spiess, 2017)

MAIS MESURER A LA FOIS LES BENEFICES ET LA NON MALFAISANCE

LOI POUR UNE RÉPUBLIQUE NUMÉRIQUE DU 8 OCTOBRE 2016

Article 49 En savoir plus sur cet article...

I. - Le livre Ier du code de la consommation est ainsi modifié :
1° L'article L. 111-7 est ainsi rédigé :

« Art. L. 111-7. - I. - Est qualifiée d'opérateur de plateforme en ligne toute personne physique ou morale proposant, à titre professionnel, de manière rémunérée ou non, un service de communication au public en ligne reposant sur :
« 1° Le classement ou le référencement, au moyen d'algorithmes informatiques, de contenus, de biens ou de services proposés ou mis en ligne par des tiers ;
« 2° Ou la mise en relation de plusieurs parties en vue de la vente d'un bien, de la fourniture d'un service ou de l'échange ou du partage d'un contenu, d'un bien ou d'un service.
« II. - Tout opérateur de plateforme en ligne est tenu de délivrer au consommateur une information loyale, claire et transparente sur :
« 1° Les conditions générales d'utilisation du service d'intermédiation qu'il propose et sur les modalités de référencement, de classement et de déréférencement des contenus, des biens ou des services auxquels ce service permet d'accéder ;
« Ce décret précise, par ailleurs, pour tout opérateur de plateforme en ligne dont l'activité consiste en la fourniture d'informations permettant la comparaison des prix et des caractéristiques de biens et de services proposés par des professionnels, les informations communiquées aux consommateurs portant sur les éléments de cette comparaison et ce qui relève de la publicité au sens de l'article 20 de la loi n° 2004-575 du 21 juin 2004 pour la confiance dans l'économie numérique.
« Ce décret fixe également les modalités selon lesquelles, lorsque des professionnels, vendeurs ou prestataires de services sont mis en relation avec des consommateurs, l'opérateur de plateforme en ligne met à leur disposition un espace leur permettant de communiquer aux consommateurs les informations prévues aux articles L. 221-5 et L. 221-6. » ;

Article 4 En savoir plus sur cet article...

Après l'article L. 311-3 du même code, il est inséré un article L. 311-3-1 ainsi rédigé :

« Art. L. 311-3-1.-Sous réserve de l'application du 2° de l'article L. 311-5, 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.
« Les conditions d'application du présent article sont fixées par décret en Conseil d'Etat. »

THE ALGORITHMIC ACCOUNTABILITY ACT

Opinion | THE PRIVACY PROJECT

The Legislation That Targets the Racist Impacts of Tech

A proposed law would make big companies determine whether their algorithms discriminate, but it's lacking in some big ways.

By Margot E. Kaminski and Andrew D. Selbst
Ms. Kaminski is a law professor and Mr. Selbst is a postdoctoral scholar.

May 7, 2019

126

RECOMMENDATIONS FOR LEARNING SYSTEMS

- ***Learning system data***
  - Quality of training data
  - Data as a mirror of diversity
  - Variables in which the data pose a risk of discrimination
  - Tracking
- ***Autonomy of machine learning systems***
  - Description biais
  - Attention in communication
- ***Explainability of learning methods and their assessment***
  - Explainability
  - Explaination heuristics
  - Development of standards
- ***Decision-making by machine learning systems***
  - Human role in decisions supported by machine learning systems
  - Human role in the explanation of decisions supported by machine learning systems
- ***Consent to machine learning***
  - The possibility for users to choose whether or not to enable a system's learning capacities
  - Consent within the project framework
  - Consent for the use of a machine capable of continuous learning
- ***Responsability in human-learning system interaction***
  - Monitoring system
  - Declaration of intentions for use

Cerna report Research ethics in machine learning (2017)
http://cerna-ethics-allistene.org/

LA DECLARATION DE MONTREAL POUR L'IA RESPONSABLE LANCEMENT LE 4/12/2018

CHAIRE GOOD IN TECH

Responsible Technologies for Humans and Society

AXES DE RECHERCHE CHAIRE GOOD IN TECH

Observatoire Good in Tech : l'observatoire Good in Tech vise à mesurer l'impact des technologies et de l'IA sur la société.

Axe 1: **Innovation numérique responsable: quelles mesures?**
Quels sont les indicateurs de l'innovation numérique responsable (définition, dimensions, mesure)?
Responsabilité sociale et humaine numérique des entreprises, RSE 4.0

Axe 2: **Comment développer des technologies *responsables by design*?**
Explicabilité, interprétabilité, équité, contrôle des biais, des discriminations, des opinions encapsulées dans les algorithmes.

Axe 3: **Réinventer les futurs**
Quelle société de demain dans un monde numérique?
Comment préserver les principes d'égalité dans un monde connecté?

Axe 4: **Gouvernance de l'innovation et des technologies responsables**
Quels sont les niveaux et mécanismes de gouvernance (Europe, Nation, Entreprise) ?
Quel impact de la gouvernance sur l'acceptabilité et l'appropriation des technologies?

MERCI POUR VOTRE ATTENTION

CHRISTINE.BALAGUE@IMT-BS.EU

@BALAGUE