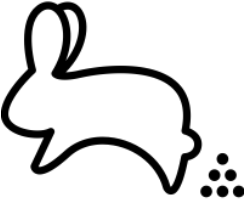
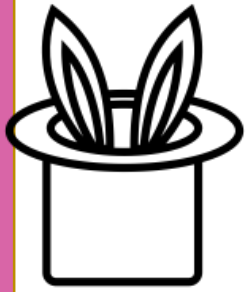


# Algorithmes

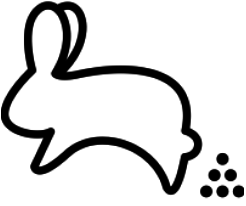
Suivre le lapin blanc pour entrer dans la boîte noire

Olivier Le Deuff



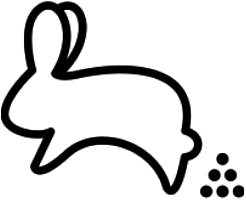


# 1.Introduction



# Introduction : enjeux

- Faire de l'algorithme un concept transversal
  - Hors de l'informatique et des mathématiques
  - Mobilisable en SHS
  - Utilisable notamment en SIC
    - Didactisable en infodoc notamment
    - Déployable sur les études journalistiques
- Interroger jusqu'ou vont les dispositifs désormais :
  - Couplage et evolution entre index et algorithme
  - Passages, du tag au like, du like au want du want au what you need, jusqu'au who you are and what you deserves...te be.. Or to buy
- Former à la gestion aux algorithmes
  - Les citoyens
  - Les professionnels
  - Offrir la possibilité d'agir sur les algorithmes

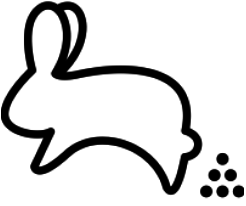


# De l'index à l'algorithme

Comme dans un passé lointain, mais de plus en plus et de façon exponentielle au cours du XXe siècle et jusqu'à aujourd'hui, ces "faits" se produisent grâce à l'infrastructure des techniques et technologies documentaires, non seulement dans les activités scientifiques et professionnelles, mais aussi en tant que dispositifs de médiation dans la vie quotidienne. Avec la récursivité, l'échelle et l'ubiquité croissantes des infrastructures sociotechniques, **les algorithmes et les index sont devenus à la fois plus opaques et plus mobiles, dissimulant les hypothèses logiques et psychologiques qui étaient autrefois très claires dans les classifications et les structures taxonomiques traditionnelles, descendantes et universelles, ainsi que dans d'autres objets et sujets professionnels des technologies de l'information.** (Day, p.16)

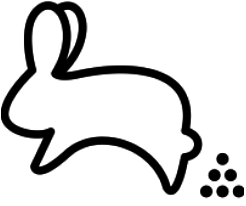
# Evolution des index et de la prise en compte du sujet

Un tel index idéal comprendrait non seulement les termes formels d'indexation et d'utilisation (par exemple, titre, sujet, auteur et résumé), mais aussi des dispositifs permettant d'orienter la recherche vers un utilisateur, car celui-ci serait constitué comme un sujet de besoin dans une situation socioculturelle et à un moment donnés. En d'autres termes, **les algorithmes personnaliseraient la recherche en créant des index techniques qui feraient référence à des citations relatives à l'expérience psychologique et au positionnement social.** Ce que nous avons décrit, bien sûr, ce sont ces algorithmes d'informatique sociale qui tentent aujourd'hui de créer des recherches personnalisées. Les algorithmes de recommandation et les algorithmes sociaux récurrents utilisent des psychologies sociales vivantes pour affiner l'indexation ou le positionnement social documentaire du moi. Ils "aident" le moi dans ce positionnement en aidant à déterminer les types de culture et de social.(Day)



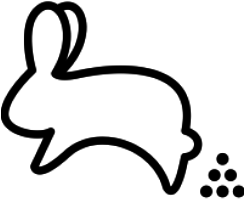
# Besoin d'étudier les algorithmes

Les effets d'interface produits par l'ordonnancement des informations algorithmiques impliquent leurs propres formes de dissociation et la compréhension critique exige donc que nous étudions les mécanismes à l'œuvre sous la surface. (Rieder, p.86°)



## Une autre vision de la compétence...

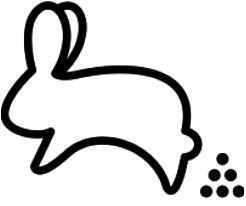
Les "digital natives" souvent loués sont certainement des utilisateurs compétents, mais il y a peu de raisons de croire que cette compétence axée sur la tâche se traduit par une conscience mécanologique (Rieder)



# L'étape suivante de l'information retrieval

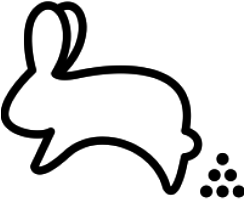
Si le but de la recherche avancée d'informations est de "faire faire aux ordinateurs des choses semblables à celles des humains" (Swanson, 1988, p. 97), l'invention de techniques algorithmiques est son principal vecteur de progrès. (Rieder, p.94)





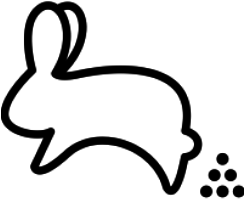
# Cultures algorithmiques

Lorsqu'il s'agit de réaliser des logiciels, la programmation occupe une place centrale et nous pouvons certainement la considérer comme une technique culturelle au sens d'une compétence fondamentale. Être capable de lire et d'écrire du code, de travailler avec des constructions telles que des variables et des structures de données de base comme des listes, d'organiser le flux de contrôle par le biais de boucles et d'instructions conditionnelles, d'utiliser différents moyens pour entrer et sortir des données et de maîtriser des aspects connexes, par exemple, ce que signifie exécuter un programme et comment le faire, ne sont pas des choses trop différentes de la maîtrise du langage ou de l'arithmétique de base. Toutefois, si la programmation reste (pour l'essentiel) une exigence fondamentale, la pratique beaucoup plus large de la fabrication de logiciels s'appuie sur un large éventail de concepts et de techniques qui couvrent les domaines déjà mentionnés comme l'analyse des besoins, la modélisation des processus et des données, l'architecture logicielle, la conception d'interactions, la gestion de projets, les tests, la maintenance, les DevOps et bien d'autres (Rieder)



# Calcul d'une forme nouvelle du social

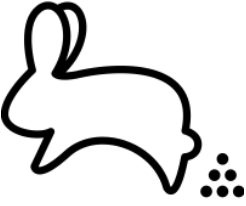
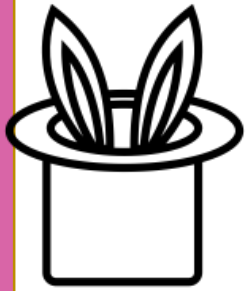
Si les logiques de personnalisation s'installent aujourd'hui dans nos vies, c'est parce qu'elles calculent une forme nouvelle du social, la société des comportements, où se recompose la relation entre le centre de la société et des individus de plus en relation entre le centre de la société et des individus de plus en plus autonomes. » (Cardon)



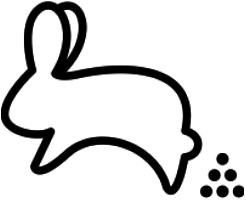
# Ouvrir la boîte noire

Omniprésents, ces calculs restent pour nous mystérieux. Ils orientent des décisions, appareillent des processus automatiques et justifient des choix politiques, mais nous interrogeons rarement la manière dont ils ont été produits. Nous regardons leurs effets sans examiner leur fabrication. Quelles sont les données qui servent au calcul ? Comment l'information a-t-elle été quantifiée ? Quels sont les principes de représentation qui animent le modèle statistique mis en œuvre pour classer tel objet plutôt que tel autre ? Qui pilote le codage des calculs et quels sont ses objectifs ?

Habités par un sentiment d'incompétence, nous préférons ignorer les conditions de fonctionnement de la société des calculs, en laissant les clés aux statisticiens, aux informaticiens et aux économètres. La complexification des modèles algorithmiques mis en œuvre dans les nouvelles infrastructures informationnelles contribue à imposer le silence à ceux qui sont soumis à leurs effets. (Cardon)

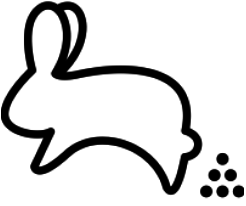


## 2. Définitions



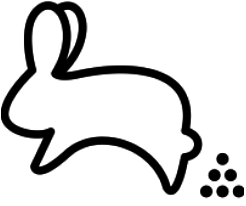
# Répondre à un problème

Écrire un algorithme, c'est dessiner un chemin de résolution pour un problème donné, une manière d'accéder rapidement et avec justesse (ou à une erreur près) à la réponse recherchée. (Jean, p. 19-20).



... par une série de règles

- est "un ensemble fini de règles qui donne une séquence d'opérations pour résoudre un type de problème spécifique" (Knuth, 1997, p. 4)

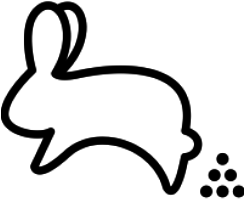


# La recette de cuisine ?

## La métaphore des ingrédients

« Avant toute chose, un algorithme n'est pas une recette de cuisine... Mais, oui, la recette de cuisine est un algorithme. » (Jean, p.42)

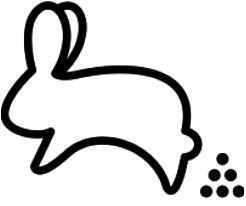
« a donc un sens similaire à celui de "recette, processus, méthode, technique, procédure, routine, rigolade' (Knuth, 1997, p. 4). (cité par Rieder)



# Une application « numérique »

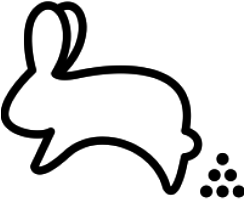
Les algorithmes développés aujourd'hui sont spécifiquement pensés pour être utilisés par un ordinateur, et uniquement par un ordinateur. Ils sont inutilisables à la main, d'où leur nom spécifique : algorithmes numériques. C'est le cas des algorithmes d'apprentissage profond, ou deep learning, qui permettent par exemple à un programme (après un « entraînement » sur des millions de cas) d'identifier un chien ou une voiture. (Jean, P.42)





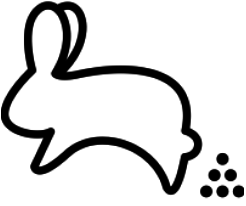
# Scalabilité ou le passage à l'échelle (industrielle)

Quand on écrit un algorithme, on doit toujours définir cette complexité, car elle permet de confirmer si un algorithme est scalable, c'est-à-dire s'il est raisonnable de l'utiliser pour des problèmes comprenant de très nombreuses données d'entrée. À l'heure du big data, on comprend que la question devienne cruciale ! (Jean, p.53)



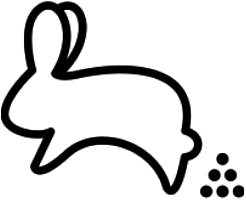
# L'exemple euclidien

- Le mégaclassique The Art of Programming de Donald Knuth, s'abstient également dans un premier temps de donner une définition et commence par montrer un algorithme, à savoir celui-ci :
- l'algorithme E (l'algorithme d'Euclide). Etant donné deux entiers positifs  $m$  et  $n$ , trouver leur plus grand diviseur commun, c'est-à-dire le plus grand nombre entier positif qui divise équitablement  $m$  et  $n$ .
- E1. Divisez  $m$  par  $n$  et laissez  $r$  être le reste. (Nous obtiendrons  $0 \leq r < n$ .)
- E2. Si  $r = 0$ , l'algorithme se termine ;  $n$  est la réponse.
- E3. Réglez  $m \leftarrow n$ ,  $n \leftarrow r$ , et retournez à l'étape E1 (Knuth, 1997, p. 2) (cite par Rieder)



# Cinq fonctionnalités

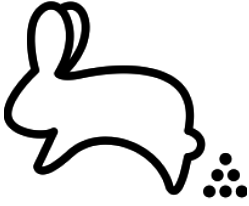
- 1) doit prendre fin après un certain nombre d'étapes,
- 2) doit être bien défini,
- 3) peut avoir une certaine influence,
- 4) doit avoir un ou plusieurs résultats, et
- 5) Doit consister en des opérations "suffisamment élémentaires pour pouvoir en principe être effectuées avec précision et dans un laps de temps limité par une personne utilisant un crayon et du papier".(Knuth, 1997, p. 6... cite par Rieder)





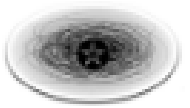

# Différence entre programme et algorithme

« De conception ordinaire, un algorithme n'est pas un programme mais une abstraction d'une solution au problème sous la forme d'une séquence d'étapes terminales, qui lorsqu'il est implémenté dans un langage de programmation, émet une liste d'instructions informatiques qui permettent de réaliser les étapes de la solution. (Colburn, 2000, p. 168) Cette distinction entre un "algorithme", une séquence d'actions qui peut être décrit dans une langue formelle ou non formelle, et un "programme", une mise en œuvre de cette séquence dans un langage qu'un ordinateur peut "comprendre" et exécuter est importante. » (Rieder)

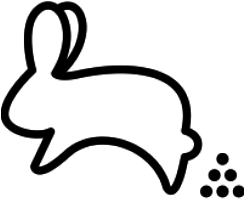
« On peut donc décliner un algorithme dans plusieurs langages de programmation. » Anthony Masure



# Des familles d'outils selon Cardon

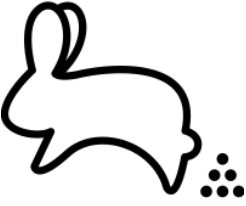
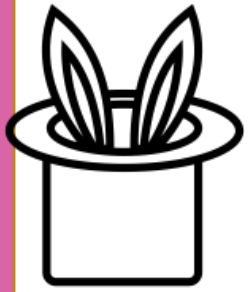
	<b>À côté</b> 	<b>Au-dessus</b> 	<b>Dans</b> 	<b>Au-dessous</b> 
<b>Exemples</b>	Médiamétrie, Google Analytics, affichage publicitaire	PageRank de Google, Digg, Wikipédia	Nombre d'amis Facebook, Retweet de Twitter, notes et avis	Recommandation Amazon, publicité comportementale
<b>Données</b>	Vues	Liens	Likes	Traces
<b>Population</b>	Échantillon représentatif	Vote censitaire, communautés	Réseau social, affinitaire, déclaratif	Comportements individuels implicites
<b>Forme du calcul</b>	Vote	Classements méritocratiques	<i>Benchmark</i>	<i>Machine learning</i>
<b>Principe</b>	<i>Popularité</i>	<i>Autorité</i>	<i>Réputation</i>	<i>Prédiction</i>

# Le risque du Jaberwocky ?

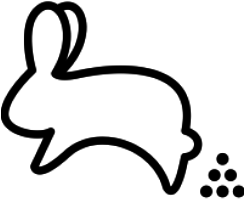


- Finir par produire un « ensemble » de règles et de calculs qui soit une forme de monstruosité difficile à qualifier
- ... et qui soit la source d'interprétation difficile
- Quasiment incontrôlable si on délègue la systémique à des processus automatisés et qui peuvent interpréter des données par apprentissage





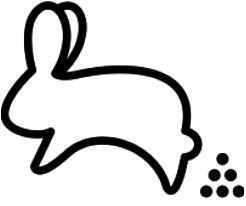
# 3. Situations



# Un poids trop fort des intermédiaires

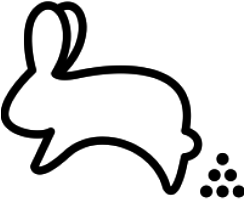
- De nouvelles médiations dont on n'a pas suffisamment conscience
- Contradiction entre les discours de l'accessibilité facilitée et les filtres souvent
  - Oubliés
  - Négligés
  - Dissimulés
  - Pour nous faire croire que tout cela est « naturelle » ou « le meilleur des mondes possibles » et qu'il n'est pas possible de faire autrement.





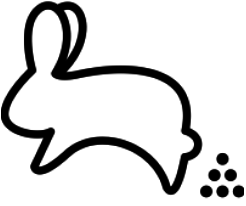
# Nouvelles taxinomies

- Nouvelles logiques de classement
- Une multiplication catégorielle potentielle
  - Quelles vues d'ensemble ?
  - La piste du graphe ?
- Des filtrages par recoupement
- Des filtrages par actions passées.



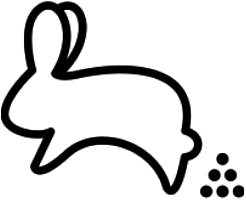
# L'automatisation des décisions

- Les journalistes filtrent l'information sur la base d'un jugement humain avant de la publier
- Les moteurs de recherche (Google News) filtrent a posteriori une information déjà publiée sur la base des jugements émis par l'ensemble des internautes
- Appui sur l'intelligence collective ou sagesse des foules



# Le besoin de données (data driven)

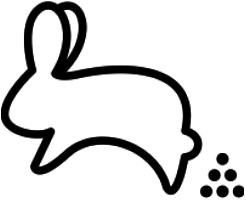
Mais une recommandation sophistiquée nécessite des données. Les nouveaux utilisateurs posent un défi que les chercheurs appellent le "problème du démarrage à froid" : ils n'ont pas encore de données et, sans données, les recommandations fondées sur des données ne fonctionnent pas. (Nick Seaver, 2018)



# La captologie

L'un des moteurs de cette montée en puissance de la pensée behavioriste dans l'industrie du logiciel a été le *Persuasive Technology Lab* de BJ Fogg à Stanford. Fogg a fondé ce laboratoire en 1998 pour développer le domaine qu'il a appelé "captologie", un nom dérivé de l'acronyme "computers as persuasive technologies" (les ordinateurs comme technologies persuasives).“ (Seaver, P.4)

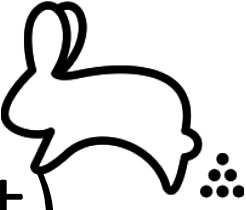




# Construire des pièges

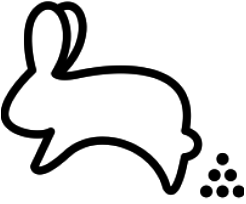
- Influence de l'anthropologie et des travaux de Mason propos des techniques de piège pour capturer les animaux dont l'efficacité repose sur la façon on comprend l'esprit de l'animal pour justement le tenter et donc le piéger.
- "technologies de l'enchantement"
- Référence à l'ouvrage *Tipping point* de Malcom Gladswell qui conçoit les sites web comme des pièges

# Captivation metrics (mesure d'enchantement)



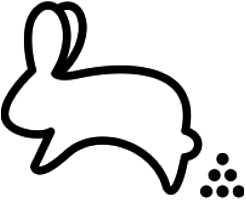
« mesures, qui ne visaient pas à prédire avec précision les évaluations, mais à mesurer la capacité d'un système à attirer l'attention des utilisateurs, ou "engagement. »(Seaver)





# Extension des données

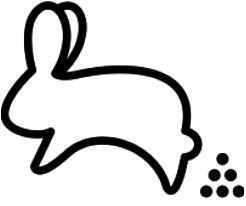
Au fur et à mesure que le centre de gravité de la communauté de la recherche s'est déplacé vers l'industrie et que les entreprises se sont orientées vers la diffusion en continu, elles ont accumulé des données qui pourraient remplacer les notations explicites qui avaient défini le domaine auparavant. Les journaux de données d'interaction pouvaient être lus comme des évaluations "implicites" : les utilisateurs qui arrêtaient une vidéo en cours de route, sautaient des éléments recommandés ou écoutaient des chansons plusieurs fois étaient tous interprétés comme des données d'évaluation". En recherchant des signes de "satisfaction" dans les journaux, les développeurs ont trouvé que les utilisateurs étaient satisfaits : tout comme le fait d'écouter une chanson de manière répétée pouvait indiquer une préférence pour celle-ci, l'utilisation continue d'un service était considérée comme un signe de satisfaction. Dans un billet de blog décrivant leur progression "au-delà des cinq étoiles" des évaluations explicites, les ingénieurs de Netflix ont écrit qu'ils se concentraient désormais sur le "plaisir de nos membres" - mesuré par le temps passé à regarder des vidéos et le temps passé à payer des abonnés (Amatriain et Basilico, 2012). Lorsque Mike m'a parlé de son objectif d'accrocher les utilisateurs, il s'est également vanté de la sophistication de Willow en matière d'analyse des données : "chaque changement qui se produit sur le service a été exactement mesuré pour son impact sur l'écoute et la rétention". (Seaver)



# Extension des métriques

En recherchant des signes de "satisfaction" dans les journaux, les développeurs ont trouvé que les utilisateurs étaient satisfaits : tout comme le fait d'écouter une chanson de manière répétée pouvait indiquer une préférence pour celle-ci, l'utilisation continue d'un service était considérée comme un signe de satisfaction. Dans un billet de blog décrivant leur progression "au-delà des cinq étoiles" des évaluations explicites, les ingénieurs de Netflix ont écrit qu'ils se concentraient désormais sur le "plaisir de nos membres" - mesuré par le temps passé à regarder des vidéos et le temps passé à payer des abonnés (Amatriain et Basilico, 2012). Lorsque Mike m'a parlé de son objectif d'accrocher les utilisateurs, il s'est également vanté de la sophistication de Willow en matière d'analyse des données : "chaque changement qui se produit sur le service a été exactement mesuré pour son impact sur l'écoute et la rétention". (Seaver)





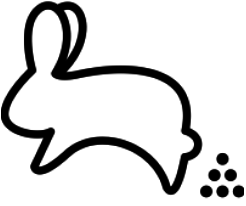
# Pièges potentiels

- Risque de la logique de l'économie de l'attention
- Les « bulles de filtre » (Eli Pariser)
- Les renforcements monomaniacques ou « radicalisés »
- L'absence d'autres mondes et d'autres possibles
- Le raté involontaire d'informations pourtant pertinentes



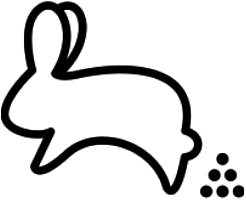
# Des pièges dans des cosmogonies préconçues

Dans cette optique, un piège n'est pas simplement l'application unilatérale d'une force technique, mais plutôt un effort fondamentalement incertain de se mettre en rapport avec d'autres, ce qui produit ainsi un monde. Nous pourrions dire que les infrastructures sont déjà des pièges - des arrangements de technique et de cadre épistémique conçus pour attirer et retenir des types particuliers d'agents envisagés, selon des préconceptions cosmologiques culturellement spécifiques. (Seaver)



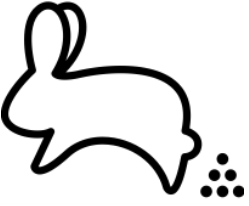
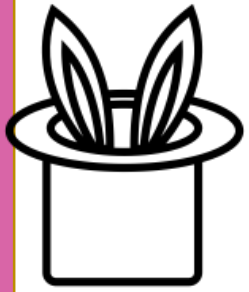
# Transformation économique

- Capitalisme numérique (Shoshana Zuboff)
- Indistinction des belligérants (Patino)

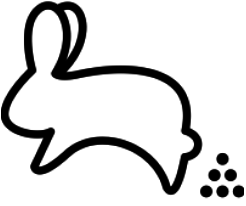


# La presse court-circuitée

- La presse transformée en entrepôts de données web
- ... qui devient des produits à consommer et à proposer
- Le menu n'est pas vraiment choisi
  - Mais sollicité et anticipé par d'autres intermédiaires comme Google

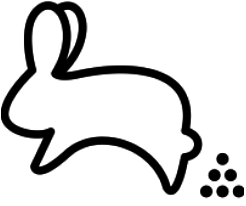


## 4. Biais et limites



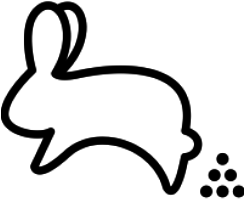
# Les fameux biais

Le biais algorithmique, c'est ce qui vient fausser les algorithmes et qui, sans qu'on s'en aperçoive toujours au premier abord, peut entraîner des discriminations écartant ou défavorisant une catégorie d'individus. Ces biais sont, par définition, non intentionnels. (...) ils sont le fruit d'impensés qui ont échappé aux concepteurs et aux penseurs des technologies. (Jean, p.21)



# Une vision universelle... intrinsèque

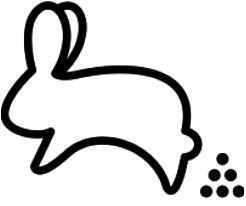
Tes modèles et tes algorithmes peuvent contenir des biais, car tu les écris à partir de ta connaissance du monde – et, ici, de tes connaissances du fonctionnement du corps humain... Lesquelles dépendent de ta compréhension de ce que te disent les médecins.(Jean, p.88-89)



# Source des biais

- Distinction entre les biais explicites, assumés et les implicites qui peuvent prendre de l'ampleur dans le *machine learning*.
- Les biais viennent souvent des données notamment lorsqu'on a pris une source de données provenant d'un pays et qu'on l'applique à tous. (exemple des données de crânes racontées par AJ)
- Les biais viennent parfois des données ou modèles hérités qu'on utilise

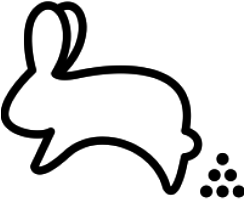




# Le faux procès

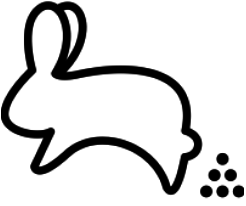
On les accuse de racisme, de sexisme, de favoritisme... Mais c'est un faux procès ! Car, si sophistiqué qu'il soit, un algorithme ne fait jamais que ce pour quoi on l'a programmé – même dans le cas d'une intelligence artificielle apprenante. L'algorithme n'a pas de conscience, il n'a pas d'autonomie, il n'a pas de pouvoirs magiques. Disons-le avec force : fustiger les algorithmes, c'est prendre la question par le mauvais bout. ( Jean, p.114°)

Comme je le disais, éliminer nos biais cognitifs, c'est éliminer nos singularités, les différences qui font que nous nous comprenons mieux les uns les autres. (Jean, p.129)



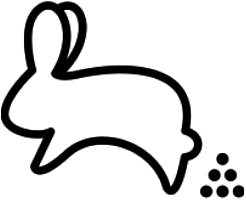
# Biais de genre et biais racistes

- L'exemple de l'algorithme des greffes et ses effets problématiques aux USA... en attendant ailleurs.
- Les effets de pénalisation directes ou indirectes
- Les algorithmes qui traitent mal et désavantage les photos de « non-blancs » (Récent exemple sur twitter)



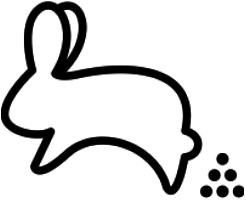
# Fin de la théorie

Les corrélations n'ont pas besoin de causes. Dans un article qui a fait grand bruit, Chris Anderson, un des gourous de la Silicon Valley, a annoncé la « fin de la théorie ». Les calculateurs des *big data*, explique-t-il, peuvent désormais chercher des corrélations sans se préoccuper d'avoir un modèle qui leur donne une explication. Les données massives et les mathématiques permettraient de faire l'économie des sciences de l'homme (Cardon)



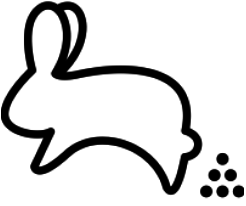
# Corrélations sans cause

Qui sait pourquoi les gens font ce qu'ils font ? Le fait est qu'ils le font et on peut l'enregistrer avec une fidélité sans précédent. Avec assez de données, les chiffres parlent d'eux-mêmes. » Prenant acte de notre méconnaissance des causes qui sont à l'origine de l'action des individus, les calculateurs abandonnent la recherche d'un modèle permettant de l'expliquer *a priori*. Aussi est-ce un nouveau rapport à la causalité qui s'est mis en place dans certains secteurs de la statistique, conférant aux modèles dits « bayesiens » une victoire posthume sur la une troisième secousse ébranle les repères de la statistique standard : les corrélations n'ont pas besoin de causes. (cardon)



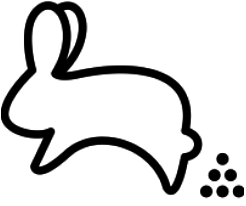
# Technologies de la spéculation

- Anticiper et prévoir...
  - À partir de l'existant
- Peut-on alors anticiper l'imprévisible voire l'inconcevable ?



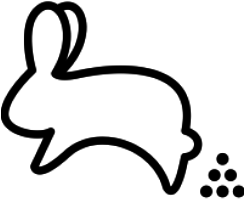
# Modélisation et influence

- Autres considérations précisées par A.J, est-ce que mon modèle développé à partir de ma représentation du monde n'est pas faussée, exemple :
- « il a fallu du temps pour qu'on comprenne que la terre tourne autour du soleil... est-ce que certaines de nos modélisations ne sont pas fausses ? »
- « comment savoir si la mesure que je réalise ne perturbe pas le phénomène que je souhaite analyser et quantifier ? » ( Jean, p.99)



# Délégation et privation de l'entendement

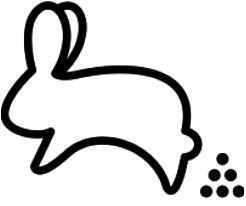
Les infrastructures des big data cherchent à guider sans contraindre, à orienter sans obliger. Elles constituent un exemple typique de ce que Cass Sunstein appelle des nudges, ces outils du « paternalisme libertaire<sup>21</sup> » qui, par défaut, suppléent les choix des individus en les persuadant qu'ils agissent au mieux de leurs intérêts. En fait, les algorithmes rêvent de délester les humains de ce qu'il y a de plus mécanique dans leurs activités, assurant qu'ils les libèrent pour des tâches cognitives plus hautes, plus complexes ou plus ambitieuses. (Cardon)



# Etudier les infrastructures globalement

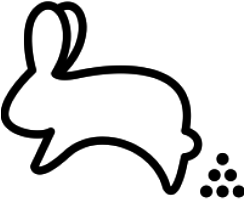
Il n'en reste pas moins que l'enjeu politique que posent les nouvelles boîtes noires du calcul algorithmique est celui de la capacité à les débrayer et à « passer en manuel ». Le risque que présentent les nouvelles infrastructures de calcul est d'architecturer les choix en les fermant sur des processus irréversibles. (Cardon)





# Produire ou connaître des algorithmes

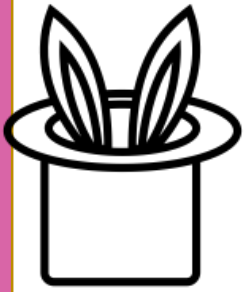
- La culture algorithmique repose sur le besoin d'en connaître et d'en intégrer
- Ce n'est pas nécessairement la capacité à en produire de nouveaux
- Plusieurs méthodes selon les situations
- L'enjeu est de diminuer les temps de traitement et d'améliorer l'efficacité des résultats (exemple. Aurélie Jean et son algo à Bloomberg)



# Révéler le monde

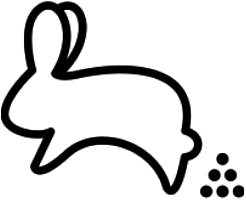
La technologie, dans ce sens, n'est pas seulement quelque chose qui peut être connu, mais un moyen de savoir, une forme fondamentale de savoir. Elle n'est ni une simple application de la science, ni une simple manifestation de forces sociales ou économiques, mais un mode de révélation du monde (Rieder, P.103)

Il faut alors révéler les méthodes des prétendus révélateurs...



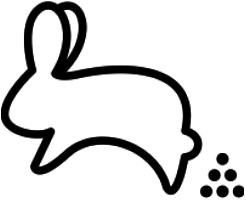
# 5. Algorithm Accountability

Vers une responsabilité algorithmique



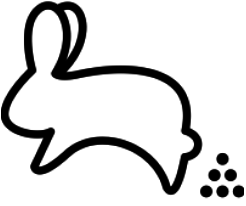
# Vers une éthique

En juin 2018 Grégory Renard (spécialiste des plateformes conversationnelles et fondateur de xBrain en Californie) et moi-même avons publié, sur le modèle du serment d'Hippocrate, l'un des premiers codes éthiques d'IA : le serment Holberton-Turing. (Jean, p.138-139)



# Ethique et quantum literacy

Nous proposons de prendre en compte, de ces deux points de vue, une « littératie" nouvelle et émergente que nous proposons d'appeler une « littératie quantique". Nous voulons ainsi attirer l'attention sur l'inadéquation de principe de la réflexion sur les chiffres et les lettres, les mathématiques et le langage, en tant que deux domaines distincts dont le premier concerne les nécessaires tandis que le second concerne les nécessaires. Nous soutenons que la condition algorithmique s'étend au-delà de ces modes de relation spatiale et les modifie fondamentalement en mettant au premier plan les logiques temporelles sur lesquelles reposent à la fois les interfaces et les pratiques de navigation. La navigation dans un temps non linéaire modifie le "saut d'échelle" critique, quoique plutôt prévisible (Smith, 1992) et introduit des activités critiques et créatives imprévisibles dans et par l'appareil humain de la technologie des données (ethics of coding)



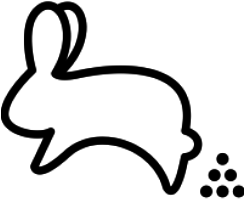
# Des enjeux de transparence

- Obligation de déclaration et de mise en ligne des algorithmes notamment pour les questions « publiques » (exemple de Parcoursup)
- Documenter les algorithmes, c'est-à-dire les processus utilisés et son déploiement
- L'algorithme mathématique ne suffit pas.



# Les pistes de réintrogénierie

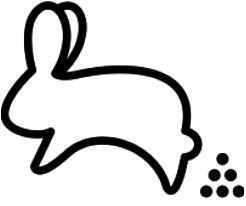
- La piste de « *algorithmic accountability reporting* » (enquête sur la responsabilité des algorithmes) (Diakopoulos, 2015).
  - déconstruire les enjeux de pouvoir qui président à la conception des algorithmes et qui les accompagnent, afin de repérer certaines pratiques discriminatoires, notamment. Ils usent, pour cela d'une méthode proche de la rétro-ingénierie qui leur permet de simuler le fonctionnement de ces lignes de codes afin de les reconstituer.
- Diakopoulos précise néanmoins que cette méthode n'est efficace que si elle est accompagnée d'entretiens approfondis avec les acteurs de ce milieu.
- Car, même si les algorithmes sont, par nature, des artefacts technologiques, ils agissent et réagissent à l'intérieur d'un système foncièrement social qu'il convient d'analyser (Weber et Kosterich, 2018).
- Ouvrir la boîte noire des algorithmes exige, de ce fait, d'étudier, tout à la fois, le fonctionnement de ces lignes de codes, les motivations de leurs créateurs et les interactions avec les usagers qu'ils finissent toujours par engendrer (Pasquale 2015; Diakopoulos 2015).



# Vers une responsabilité algorithmique à « reporter »

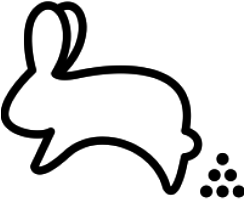
Les journalistes commencent à adapter leurs fonctions traditionnelles de surveillance et de responsabilité à cette nouvelle source de pouvoir dans la société. Ils étudient les algorithmes afin de caractériser leur pouvoir et de délimiter leurs erreurs et leurs préjugés dans un processus que j'appelle "reportage de responsabilité algorithmique"..  
(Diakopoulos)





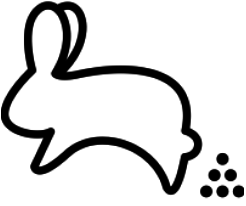
# Répondre aux conditions préalables de la fraude

La transparence journalistique à son niveau le plus élémentaire peut être définie comme l'exposition publique des facteurs cachés qui façonnent la production d'informations au public (Allen 2008). En tant que directive morale plus générale pouvant servir d'indicateur de la vérité, la transparence est ancrée dans le principe de la dignité humaine d'Emmanuel Kant : selon lui, même si la transparence n'est pas une condition suffisante pour un comportement éthique, son absence est une condition préalable à la fraude. (Data driven storytelling)



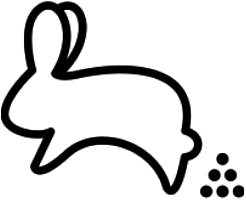
# Exemple : Netflix décrypté

- Plusieurs algorithmes utilisés
  - L'algorithme PVR (Personalized Video Ranker : classement personnalisé en fonction des catégories)
  - Le TOP N (filtrage basé sur l'ensemble du catalogue)
  - Trending now (Lié à l'évènementiel, par exemple Noël mais aussi les requêtes liées à l'actualité)
  - Continue Watching (prends en compte les visualisations précédentes pour supposer l'envie de reprise)
  - Video-video similarity (« parce que vous avez regardé... »)
  - Evidence ou element reassurance (personnalisation des métadonnées et des miniatures affichées pour mieux séduire)
  - Search (proposition semiguidée en fonction du catalogue et des goûts potentiels)
- [Source](#)



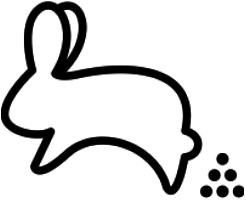
# Algorithm watchers

Mais on peut imaginer des agents algorithmiques (que je nomme algorithm watchers, en anglais) qui analyseraient le comportement en temps réel d'un algorithme d'usage en train d'apprendre. L'idée est que ces agents algorithmiques analysent tout d'abord la représentativité des données d'apprentissage. Pour ce faire, l'agent algorithmique fournit à une copie miroir de l'algorithme d'usage des échantillons (Jean,p. 140)



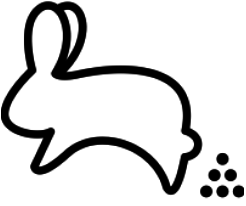
# S'échapper et renverser les processus

La question à poser aux pièges n'est peut-être pas de savoir comment y échapper, mais plutôt comment les reconquérir et les orienter vers de nouvelles fins au service de nouveaux mondes. (SeaverP.13)



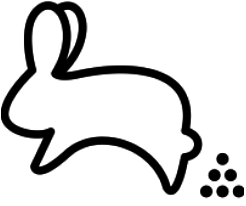
# Une nouvelle mécanologie (Rieder)

- Intégrer les processus algorithmiques et informatiques dans une approche mécanologique à la Simondon
- Etudier non pas les externalités et donc les usages mais étudier le cœur des dispositifs
- Décrypter, documenter, expliquer, didactiser les processus
- Penser l'articulation entre les « organologies » et les mécanologies en n'isolant pas le processus technologique des autres aspects.
- Expérimenter les processus algorithmiques en dehors des plateformes dominantes pour envisager des modèles alternatifs ou... dérivés.



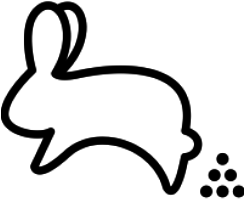
# Data driven storytelling

Nous pensons que ce mouvement vers des reportages basés sur les données, qui est apparent à la fois dans la communauté de recherche sur la visualisation des données et dans la communauté du journalisme professionnel, a le potentiel de former une partie cruciale de l'information du public, un mouvement parfois appelé démocratisation des données - rendre les données compréhensibles pour le grand public. Ce nouveau développement passionnant dans l'utilisation de la visualisation de données dans les médias a révélé l'émergence d'une communauté professionnelle dans le groupe déjà complexe de disciplines impliquées dans la visualisation de données



# La piste du design

- Anthony Masure. Voir sa [thèse](#).
- Permettre l'appréhension par les non-spécialistes
- Défendre l'idée d'une accessibilité à l'algorithme notamment aux logiques de filtrage (que les usagers connaissent déjà sous forme de facettes).



# La piste du design graphique

- Ouvrir les algorithmes dans des possibilités de manipulation graphique
  - Pour les rendre plus visibles
  - Pour mieux expliquer les potentialités
  - Pour offrir l'usager la possibilité de faire des choix
  - Pour faire monter en compétence l'usager dans son algorithm literacy
  - Pour finalement améliorer les algorithmes eux-mêmes
  - Pour retrouver l'esprit web de la personnalisation de l'information qui reste sous le contrôle de l'usager lui-même (cf. Projet de Tariq Krim <https://www.polite.one/>)