

VOUS N'AURIEZ JAMAIS CRU QU'UN DAUPHIN PUISSE FAIRE ÇA

Mémoire
DNSEP

Marianne
Plano

Hear
2017

SUR LES ALGORITHMES

#freedomIsTheDearestOfOurPossessions

#timeToWalkOnWater

MARIANNE PLANO

VOUS N'AURIEZ JAMAIS
CRU QU'UN DAUPHIN
PUISSE FAIRE ÇA

SUR LES
ALGORITHMES

Mémoire
DNSEP

Tuteur :
Anne Laforet

Hear
2017 / 2018

Section :
Communication
Graphique

Note sur la traduction

Ce mémoire se base principalement sur des textes en anglais, qui pour certains, n'existent pas en français (notamment les articles de journaux). Lorsque je les cite, je me suis efforcée de les traduire au plus juste, mais c'est un exercice vraiment difficile. Si certaines tournures paraissent maladroites au lecteur, je lui demanderais de m'en excuser et quoi qu'il en soit, de toujours se référer à l'original.

Corps

| | | |
|-----|----------------------------|----|
| | Introduction | 3 |
| I | Contexte | 9 |
| II | Neutralité | 23 |
| III | The Frame Problem | 39 |
| IV | Monétiser l'espace | 55 |
| V | Monétiser les utilisateurs | 71 |
| VI | Conclusion | 85 |

Annexes

| | | |
|-----|--|--------|
| I | Machines universelles | iii |
| II | Parlez-vous l'électricité ? | vii |
| III | « Ne crois aucun scan que tu n'ai toi-même trafiqué » | xv |
| IV | Le scandale de la NSA | xx |
| V | Erstazs d'extraordinaire | xxiv |
| VI | Ce que dit la loi européenne | xxviii |

Lexique

Bibliographie

« Notre défi, et il n'est peut-être pas tenable, serait d'échapper, face au processus d'innovation technique, à tout jugement de valeur, toute réaction affective de répulsion ou de séduction. »*

Régis Debray

*. *Cours de médiologie générale*, Paris : Gallimard, 1991

Quand j'étais petite, il n'y avait qu'un ordinateur à la maison et Internet se payait à l'heure, comme le téléphone. Mais je me m'en souviens pas très bien, comme j'ai du mal à me remémorer le visage de mon instituteur de CE1 ou le nom de ma meilleure copine en maternelle. Cette époque tient plus du fantasme nostalgique que du véritable souvenir. Ma génération a grandi avec la révolution numérique. J'ai toujours fait mes exposés avec Wikipédia (surnommé « le démon » par le corps enseignant des années 2000) et considéré l'encyclopédie *Tout l'univers* qu'on m'avait donnée comme une relique.

« La révolution numérique a changé tellement de choses qu'on a du mal à se souvenir comment on faisait avant. »¹

Ma mère m'explique que lorsqu'elle était jeune, les étudiants faisaient taper leur thèses à la machine à écrire par des professionnels, ce qui à mes yeux relève du passé-fiction complet. Elle a vécu la révolution de l'ordinateur personnel, de l'Internet illimité, du moteur de recherche et de l'encyclopédie en ligne, qui pour moi sont des choses aussi familières que les frigidaires ou les panneaux de signalisation. Pourtant, je me rends compte que tous ces outils que je manipule si naturellement au quotidien me sont, dans le fond, étrangers. Je regarde mon ordinateur, et réalise que je n'ai pas la moindre idée de ce qui se passe à l'intérieur. L'Internet, avec lequel je vis presque en continu, m'est pareillement, un mystère total. Sans parler des logiciels, boîtes noires impénétrables. Au fond, c'est avec de parfaits inconnus que je passe le plus clair de mon temps. Puis, je découvre la programmation, grâce

1. Gérard BERRY, *L'Hyperpuissance de l'informatique*, Paris : Odile Jacob, 2017

au *web design*. Tout doucement, à ma petite échelle, mais programmation quand même. Et là, c'est la grande démystification. Ce que je comprends, c'est que déjà, l'informatique est un magnifique terrain de jeu, infini. Et surtout, que par cet apprentissage, ce n'est pas seulement les ordinateurs qui s'expliquent, mais le monde dans lequel je vis. Alors je veux en savoir plus. Je veux ouvrir les machines,² pour comprendre ce qu'elles trament. Car l'asymétrie du savoir entre nous est devenue difficilement supportable. Ou devrais-je dire, l'asymétrie entre ceux qui les fabriquent et moi. Entre ceux qui les fabriquent et nous.

L'ampleur des transformations apportées par le numérique dépasse largement le cadre technique. En tant que citoyens des années 2010, nous devons nous serrer les coudes pour voir plus clair, et les démystifier, à notre niveau. Car derrière ces milliards de lignes de code se cachent, de manière banale, des hommes, et donc, des enjeux politiques. Qui sont l'affaire de tous. J'ai choisi de passer mes derniers temps d'étude avec les algorithmes, que je connaissais très mal il y a peu, mais qui pour certains d'entre eux, me connaissaient très, très bien, comme je l'ai appris. Enfin, croyaient me connaître. **#FreedomIsTheDearestOfOurPossessions**

Concepts incontournables de l'informatique, ils prennent une place croissante dans nos sociétés et méritent toute notre attention. Ils jouent aux échecs, prennent des photos, prédissent le crime, construisent des circuits espionnent les citoyens, distribuent l'électricité dans les villes ou encore influencent la politique pour ne citer que quelques-uns de leurs talents. Cela est dû à l'expansion phénoménale des outils numériques, où ils pullulent littéralement, à différents niveaux. Ils sont les briques fondatrices de que Gérard Berry appelle « l'informatisation du monde ».³

Mon mémoire retrace ma quête de leur compréhension et tente de donner quelques clés pour les aborder. Je me positionne avant tout en tant que citoyenne, qui tente de saisir quelques-uns des enjeux amenés par les nouvelles technologies, et la manière dont elles influent sur nos vies. Ce texte est

2. Quand je dis *machines*, je parle aussi bien d'ordinateurs que de programmes ou d'algorithmes. J'entends par ce terme, tout entité, qu'elle soit physique ou conceptuelle, qui permet d'effectuer une série de tâches de manière automatique. Au cours de ma démonstration, c'est dans ce sens global qu'il faut le comprendre, qui est toujours utilisé indifféremment de la nature matérielle ou non de l'objet qu'il décrit.

3. Gérard BERRY, *L'Hyperpuissance de l'informatique*, Paris : Odile Jacob, 2017

orienté vers la politique, et non la technique. Il me faut néanmoins toujours passer par des explications de ce type, pour cerner chacun de mes objets d'étude et être en mesure de décortiquer véritablement leur impact, et ainsi éviter les suppositions hasardeuses qui ne seraient que le reflet de mes désirs. C'est donc très modestement, que pour aborder les algorithmes, je commencerai par le commencement : à savoir, ce qu'ils sont.

Contexte

I

Informatique

1
I

«L'informatique ne traite pas plus d'ordinateurs
que l'astronomie de télescopes.»¹

L'informatique, ce n'est pas que les ordinateurs.^A C'est une science, à part entière. La « science du traitement de l'information »² qui à ce titre, comprend de nombreux domaines. Nos ordinateurs quotidiens en sont une des applications, mais pas un pré-requis. Certains de ses champs sont des mathématiques abstraites, qui n'ont que peu sinon rien à voir avec les transistors ou les câbles. Ils traitent de logique pure, sur laquelle sont conceptuellement, construites les machines. L'informatique n'est néanmoins, pas une science comme les autres. Ce n'est pas le monde naturel qu'elle décrypte mais un monde parallèle qu'elle construit. Pour cela, on dit qu'elle est « une science artificielle »,² une « science de construction ». ² Et sa matière première, c'est l'information.

1. « *Computer science is no more about computers than astronomy is about telescopes.* » Michael R. FELLOWS, Ian PARBERRY, *SIGACT trying to get children excited about CS*, *Computer Research News*, 1993. Cette citation est tiré du livre de Serge ABITEBOUL et Gilles DOWEK, *Le Temps des algorithmes*, Paris : Le Pommier, 2017

A. Cf. Ordinateur, lexique

2. Gérard BERRY, *L'Hyperpuissance de l'informatique*, Paris : Odile Jacob, 2017

UNE FORME UNIQUE

Dans le monde analogique du XX^e siècle, chaque information était attachée à sa matière propre. Le texte et l'image s'incarnaient dans le couple encre/papier, par l'écriture ou l'imprimerie. Le son existait sous la forme d'un sillon plus ou moins profond gravé sur la surface d'un disque vinyle, lisible par des appareils spécifiques, comme le gramophone ou plus tard la platine. La photographie résultait quant à elle, du couple chimie/papier.

« Tout ces objets étaient profondément différents, conçus par des ingénieurs différents et fabriqués dans des usines différentes. À chaque objet son métier. »³

L'informatique a permis de tous les unifier, en les numérisant, c'est-à-dire en « les traduisant sous la forme de nombres stockés dans des mémoires parfaitement interchangeables, puis travaillées par les mêmes ordinateurs ». ³ Aujourd'hui le texte, le son, l'image existent tous sous une forme identique, et les supports analogiques sont peu à peu remplacés ou complétés comme dans le cas du livre papier, par l'ordinateur seul. Cette forme identique, c'est le langage binaire. Toute donnée présente sur un ordinateur existe physiquement sous la forme de nombres binaires, appelé *bits*. ¹

UN SUPPORT LIBRE

« Les données et les programmes^B de l'informatique ne sont pas des objets matériels car ils ne sont pas liés à un support physique privilégié. » ³

Le terme de « dématérialisation », souvent employé comme synonyme de « numérisation » peut porter à confusion. Il ne signifie pas que le support de l'information n'existe plus. Les ordinateurs personnels ou les *data centers* qui s'étalent sur des kilomètres tout autour de la planète, en sont bel et bien des supports physiques. « Dématisation » signifie que l'information n'est plus attachée à un support spécifique, comme la musique était indissociable du vinyle analogique. Peu importe l'ordinateur sur lequel elle est stockée,

3. Gérard BERRY, *L'Hyperpuissance de l'informatique*, Paris : Odile Jacob, 2017

I. Annexe : **Parlez-vous l'électricité ? De la matérialité du bit**

B. Cf. Programme, lexique

elle reste identique. Cette liberté va même plus loin. Des scientifiques de l'université de Harvard ont récemment réussi à encoder une vidéo non pas sur un ordinateur électrique mais dans l'ADN.⁴ L'information étant *in fine* une simple série de chiffres, qu'ils soient écrit à la main, représentés par un courant électrique, ou par des variations à l'intérieure d'une molécule, est indifférent. Ils restent toujours identiques. "

Algorithme

2 I

«L'algorithme est l'idée fédératrice de tous les domaines que les informaticiens investissent.»⁵

Le terme vient du nom d'un grand mathématicien perse du IX^e siècle, Al-Khwarizmi, considéré comme le père de l'informatique et de l'algèbre ainsi que du Grec ancien *arithmos*, nombre. Un algorithme est «une recette, une série de tâches en vue d'accomplir un calcul ou un résultat spécifique, comme les étapes nécessaires pour obtenir la racine carrée d'un nombre.»⁶ Comme le souligne Dominique Cardon,⁷ ce terme a une signification bien plus large qu'on ne le croit, et s'apparente en pratique à la recette de cuisine, en tant que «description d'une méthode par laquelle une tâche peut être accomplie».⁸ Et le grand intérêt de l'algorithme, c'est qu'il «est un procédé qui permet de résoudre un problème, sans avoir besoin d'inventer une solu-

4. Ian SAMPLE, Harvard Scientists Pioneer Storage of Video inside DNA, *The Guardian*, 13.7.17

II. Annexe : **Machines universelles : toutes des ordinateurs**

5. «*The algorithm is the unifying concept for all the activities which computer scientists engage in.*» Andrew GOFFEY, Algorithm, *Software Studies, a Lexicon*, publié par Matthew FULLER, Cambridge: MIT Press, 2008
6. «*An algorithm is a recipe, an instruction set, a sequence of tasks to achieve a particular calculation or result, like the steps needed to calculate a square root [...]*» Ed FINN, *What Algorithms Want*, Cambridge: MIT Press, 2017
7. Dominique CARDON, *À Quoi rêvent les algorithmes ? Nos vies à l'heure des big data*, Paris: Seuil, 2015
8. «*[...] description of the method by which a task is to be accomplished [...]*» Andrew GOFFEY, Algorithm, *Software Studies, a Lexicon*, publié par Matthew FULLER, MIT Press: Cambridge, 2008

tion à chaque fois.»⁹ C'est le concept, l'outil de base pour traiter l'information.

Dans le domaine des mathématiques, un algorithme se traduit par une suite d'opérations qui permettent d'aboutir à un résultat. Une liste d'instructions, qui appliquée à A donnent B. Par exemple, la méthode qui consiste à poser les additions que nous apprenons petits à l'école est un algorithme. Il décompose un problème compliqué en une série d'étapes simples, tel qu'ajouter des nombres à 1 chiffre, l'un après l'autre, ce qui est beaucoup plus facile pour un être humain. Il remplit ainsi une double fonction : réduire l'effort, et la marge d'erreur quant au résultat final.

Un algorithme est «indépendant des langages de programmation^C et indépendant des ordinateurs qui exécutent ces programmes.»¹⁰ Tout comme l'informatique dans son ensemble, il existe à part de tout support, sous une forme abstraite. Il est une recette de base pour faire un plat, qui s'incarne toujours un peu différemment selon qui la cuisine et avec quels ingrédients, tout en restant fondamentalement la même.

SIMPLICITÉ

«La richesse de l'algorithme ne vient pas de la complexité de leurs composants, mais de la manière dont quelques composants simples sont assemblés. Le lecteur pourra faire un parallèle avec le fait que les milliards de molécules que nous connaissons sont composées de seulement quelques dizaines d'éléments chimiques, eux-mêmes composés de trois particules élémentaires : protons, neutrons, électrons.»¹¹ La conception d'algorithmes est une construction, une épreuve de logique au service de l'optimisation. Logique d'économie, qui se retrouve dans le motto de tout programmeur : faire le plus possible avec le moins possible.

9. Serge ABITEBOUL, Gilles DOWEK, *Le Temps des algorithmes*, Paris : Le Pommier, 2017

C. Cf. Langage de programmation, lexique

10. Andrew GOFFEY, Algorithm, *Software Studies, a Lexicon*, publié par Matthew FULLER, MIT Press : Cambridge, 2008

11. Serge ABITEBOUL, Gilles DOWEK, *Le Temps des algorithmes*, Paris : Le Pommier, 2017

Big Data

3 I

Les algorithmes ont toujours été utilisés pour construire et faire fonctionner les ordinateurs. Ce qui les a récemment propulsés dans le débat public, c'est la naissance d'un phénomène appelé le *Big Data*. Ce terme désigne l'émergence de « masses de données » disponibles, qui découlent de la multiplication des dispositifs capables de quantifier le réel en termes informatiques. Au même titre que la fin des supports analogiques, il est une conséquence de « l'informatisation du monde », et concerne tous les domaines de la vie humaine, de la physique quantique au shopping sur net. Ce phénomène est intrinsèquement lié au progrès des ordinateurs, et à l'expansion ultra-rapide de l'espace de stockage et de calcul depuis leur invention, selon la Loi de Moore.^D Désormais, la « transcription du monde physique et de ses habitants sous forme de données métabolisables par les systèmes informatiques n'est [...] plus limité[e], ni même freiné[e] de manière essentielle par une inaccessibilité technique ou économique. »¹² Les données ainsi récoltées par différents dispositifs, qui vont du thermomètre à l'historique de navigation en passant par le GPS, sont dites « brutes ». Cela signifie qu'elles n'ont pas encore été transformées en informations, c'est-à-dire qu'aucun sens n'en a été tiré.

DONNÉE ≠ INFORMATION

« L'information contenue dans une donnée dépend essentiellement du contexte d'interprétation »¹³

Pour expliquer cette notion, Gérard Berry prend l'exemple de la température. Selon son unité, elle aura du sens ou non dans certaines parties du globe. L'exprimer en Fahrenheit ou en degrés Celsius, c'est passer de la donnée à l'information, la rendre compréhensible à l'aune d'un contexte. Analyser les données, c'est transformer le brut en signifiant. L'émergence du *Big*

D. Cf. Loi de Moore, lexique

12. Antoinette ROUVROY, *Des Données et des hommes. Droits et libertés fondamentaux dans un monde de données massives*, Académia, 2016

13. Gérard BERRY, *L'Hyperpuissance de l'informatique*, Paris : Odile Jacob, 2017

Data a amené avec lui le fantasme de la mort de l'interprétation, qu'illustre bien l'article The Data Deluge Makes the Scientific Method Obsolete de Chris Anderson, ancien rédacteur en chef du magazine *Wired*. Il annonce en 2008 que les données massives rendront obsolète l'élaboration de théories par les hommes, et la méthode scientifique qui tire des modèles^E généraux de corrélations. « Avec assez des données, les chiffres parlent d'eux-mêmes. »¹⁴

S'il est vrai que le *Big Data* a permis l'émergence de nouvelles corrélations, jusqu'alors invisibles, qui infirmèrent des modèles précédemment établis, il n'a pas supprimé le travail d'analyse. Au contraire, il l'a rendu d'autant plus précieux. Comme l'explique Robert Matthews, professeur de statistiques à l'université de Birmingham (UK) dans un article très drôle intitulé Les cigognes amènent les bébés (*Storks Delivers Babies*), la corrélation n'est pas la causalité. Dans ce texte, il met à jour un lien troublant entre le taux de natalité de pays européens et leurs populations de cigognes. Ce qui ne signifie pas pour autant qu'il existe un rapport de cause à effet. « La relation empirique entre le nombre de couples de cigognes et le taux de natalité humain dans 17 pays européens fournit un exemple intéressant de corrélation statistiquement très significative, qui n'est pas immédiatement explicable et qui pourtant, n'a aucun sens en termes de causalité. »¹⁵ Le modèle, comme son nom l'indique, est générique et ne peut représenter la complexité de la réalité. Néanmoins, il reste un outil d'analyse, qui participe à l'explication de phénomènes sans pour autant pouvoir les décrire entièrement. La quantité de données, si elle élargit le terrain des savoirs accessibles, ne supprime pas la nécessité d'interprétation. Attribuer une valeur de vérité immanente aux données est un non-sens. À partir du moment où une donnée devient une information, elle devient automatiquement, par extension, une interprétation.

E. Cf. Modèle, lexicque

14. « *With enough data, the numbers speak for themselves.* » Chris ANDERSON, The End of Theory: The Data Deluge Makes the Scientific Method Obsolete, *Wired US*, 23.6.2008

15. « *The empirical relationship between the number of stork breeding pairs and human birth rates in 17 European countries provides a non-trivial example of a correlation which is highly statistically significant, not immediately explicable and yet causally nonsensical.* » Robert MATHHEWS, Storks deliver Babies, *robertmatthews.org*, 6.2000

« Les datas ne parlent pas d'elles-mêmes, il faut savoir les interroger. »¹⁶

BIG DATA X ALGOS

« L'histoire de l'algorithme, c'est l'histoire d'un vide : un vide entre des systèmes de calcul conceptuels et implémentés,^F ou entre l'information et le sens. »¹⁷

Les humains ont beaucoup de mal avec les données brutes. S'ils devaient les analyser à la main, ils seraient en dessous de tout. Premièrement, car le langage informatique ne leur est pas du tout naturel, et ils risqueraient des erreurs en effectuant le travail de mise en valeur. Deuxièmement, car ils sont désespérément lents, et s'ils devaient interpréter l'océan de *Big Data*, milles vies infernales ne leur suffiraient pas. Alors ce qu'ils font, c'est construire de petites machines mathématiques qui analysent ces données récoltées par d'autres machines : les algorithmes. Il faut garder à l'esprit que ce sont bien les hommes qui les écrivent. De ce fait ils sont, premièrement, toujours au service d'intérêts. Les fonctions qu'ils remplissent, quelles qu'elles soient, sont pré-déterminées par leurs concepteurs ou leurs utilisateurs. Pour comprendre ce que fait un algorithme, il est essentiel de commencer par poser la question suivante : à qui profite-t-il ? Deuxièmement, en tant que produits humains, ils restent toujours sujets à l'erreur et à l'amélioration. « Les *Big Data* signifient donc surtout le franchissement d'un seuil de quantité, de complexité, de rapidité de prolifération des données à partir duquel nous serions contraints d'automatiser et d'accélérer (pour tenir compte de l'accroissement continu, à grande vitesse, des masses de données) les processus de transformation des données numériques en informations opérationnelles. »¹⁸

16. « *Data do not speak by themselves, we need smart questionners.* » Luciano FLORIDI, *The Fourth Revolution: How the Infosphere is Reshaping Human Reality*, Oxford : Oxford University Press, 2016

F. Cf. Implémentation, lexique

17. *The story of algorithm is the story of the gap: the space between ideal and implemented computational systems, or between information and meaning.* Ed FINN, *What Algorithms Want*, Cambridge : MIT Press, 2017

18. Antoinette ROUVROY, Des Données et des hommes. Droits et libertés fondamentaux dans un monde de données massives, Académia [academia.edu], 2016

Partout où il y a des données, il y a des algorithmes. L'informatisation du monde implique leur besoin croissant dans tous les domaines de la vie, pour tirer un sens intelligible par l'homme des données.

L'EXEMPLE DE L'IRM

À première vue, quand on ne connaît rien à la médecine comme c'est mon cas, on imagine qu'une IRM (Imagerie par Résonance Magnétique nucléaire), c'est comme une radiographie : une photographie de l'intérieur du corps. En réalité, ce procédé fonctionne par « émission de champs magnétiques et ce qu'il mesure est la vitesse de relaxation (de retour à la normale) du moment magnétique d'un noyau atomique plongé dans un champ magnétique intense et modifié par une impulsion radio. »¹⁹ Autrement dit, ce que la machine récolte ce sont des nombres et non des images. C'est un algorithme, qui à partir de ces données numériques va simuler l'image et ainsi, rendre le corps du patient intelligible par l'œil du médecin.

QUANTIFIER LES HOMMES

La médecine, et les sciences naturelles en général sont traditionnellement des viviers de données, de quantification, classification des phénomènes, à dessein de comprendre le monde. Les sciences humaines se sont aussi attelées très tôt à la récolte d'informations, afin de comprendre les hommes. Un autre détenteur traditionnel d'informations est l'État, qui par l'administration, définit sa population. Au États-Unis, la constitution impose un recensement de la population tous les dix ans. À la fin du XIX^e siècle, avec les vagues d'immigration massives et l'augmentation du niveau de vie, le nombre d'habitants a augmenté de manière exponentielle. Le recensement, encore fait à la main, devint un calvaire. Pour automatiser le traitement de l'information, Herman Hollerith, un ancien professeur du MIT mit au point en 1889 un système des cartes perforées, qui servaient à catégoriser les individus recensés. Une fois la carte percée, elle était insérée dans une machine qui automatiquement, comptait le nombre de citoyens dans chaque catégorie. Ce fut un immense succès. Fort de sa technologie, Hollerith se lança dans les affaires, et fonda une entreprise qui plus tard devint IBM. Les cartes perforées sont le premier format sous lequel furent écrits les programmes

19. Gérard BERRY, *L'Hyperpuissance de l'informatique*, Paris : Odile Jacob, 2017

informatiques. Elle servaient à programmer les premiers ordinateurs, tel que l'ENIAC (États-Unis, 1946). Il est intéressant de constater que l'invention des ordinateurs découle, entre autres, de la nécessité de faciliter la collecte d'informations sur une population.

DONNÉES PERSONNELLES

Les individus eux aussi, deviennent informatiques. De ce fait, ils sont de manière croissante définis par ce qu'on appelle « les données personnelles ». Pour les cerner, Antoinette Rouvroy les divise en quatre groupes. Premièrement, les *hard data*, qui regroupent les données produites par les institutions et administrations publiques (certificats de naissance, de décès, mariages, diplômes...). Elles faisaient également partie du monde pré-informatique, mais leur numérisation amène de nouveaux enjeux. ^{III} Deuxièmement, les *soft data*, qui désignent les traces des comportements en ligne (historique de navigation, historique d'achats, mails échangés...). Elles découlent de l'arrivée d'Internet, qui engendra la numérisation d'activités encore matérielles au XX^e siècle, désormais très facilement enregistrables. Troisièmement, les *meta datas*. Cette catégorie comprend les données relatives aux données (les heures de connections, de publications, coordonnées GPS, la durée d'une communication, le réseau utilisé...). Et pour terminer, quatrièmement, les *données relatives au objets*, qui communiquent entre eux sans intervention humaine (ce qu'on appelle « Internet des objets »), comme l'utilisation d'énergie, d'espace de stockage, les données de performances...

CROISEMENT

L'historique de navigation d'une personne nous en déjà apprend beaucoup sur elle. Mais croisé avec les données de géolocalisation de ses appareils, l'information devient bien plus précise et précieuse. Si l'on ajoutait le casier judiciaire, le portrait gagnerait encore en acuité. Le recoupement de différents types de données démultiplie leur valeur. Le *Big Data* et les ordinateurs qui le supportent, offrent notamment la possibilité nouvelle de comparer d'immenses set de données, pour un faible coût en temps, grâce aux algorithmes. Pour ce faire, toutes sont sauvegardées, même de celles qui n'ont à première vue, qu'un faible potentiel informatif. La valeur vient du

III. Annexe : « Ne Crois aucun scan que tu n'ai toi-même trafiqué. »

croisement, qui met à jour des corrélations, sans pour autant effacer le rôle d'interprète que jouent les hommes dans l'établissement d'une causalité, ou dans le choix des données à comparer.

DATAS NOUVEAUX-RICHES

La grande différence avec le XX^e siècle, c'est que le quasi monopole de l'État sur les données personnelles est terminé. Avec l'arrivée d'Internet, de nombreux aspects de la vie humaine, tel que la consommation, la vie sociale, la mise à disposition et la recherche de savoir se sont elles aussi numérisées. De nouveaux acteurs, curateurs de ces nouveaux espaces immatériels ont émergé, en proposant des services répondant à ces besoins sous forme informatique. Ces puissances du web, parfois appelées « data nouveaux-riches » exploitent des formes inédites de données, dont la collecte était inimaginable dans un monde analogique. Celles-ci sont intrinsèquement liées à la technologie Internet et s'incarnent sous la forme de « traces numériques des utilisateurs ». Elles prennent différentes formes : historique de navigation, données de géolocalisation, temps passé sur les sites, heures de connexion, historique des requêtes dans un moteur de recherche, historique d'achat, publications, accès aux mails de l'utilisateur et à leur contenu... et constituent une nouvelle source d'informations extrêmement prisée. Google, Facebook, Amazon en savent aujourd'hui plus sur les citoyens que leurs gouvernements. Ces mêmes entreprises se sont construites une présence incontournable sur le web, jusqu'à en définir l'expérience pour de nombreux utilisateurs. Et tout leur fonctionnement, des services qu'elles proposent à l'économie qu'elles pratiquent repose sur des algorithmes.

PROBLÉMATIQUE

« On se fait pas de la science en accumulant des données ;
on ne devrait pas faire des affaires et de la politique comme
ça non plus. »²⁰

Avec le passage au numérique, le fonctionnement des affaires humaines repose de manière croissante sur le couple données / algorithmes. Justice,

20. « *We do not do science by mere accumulation of data; we should not do business and politics in that way either.* » Luciano FLORIDI, *The Fourth Revolution: How the Infosphere is Reshaping Human Reality*, Oxford : Oxford University Press, 2016

savoir, consommation, culture, les traitements automatiques de l'information alliés à la disponibilité exponentielle des données en tout genre bouleversent l'organisation de notre société. Ma réflexion porte sur les algorithmes et leurs limites, ainsi que sur les enjeux politiques qu'ils engendrent.

Les algorithmes sont-ils aptes à s'occuper de toutes les affaires humaines ? Quelles sont les nouvelles puissances du numérique et comment leur pouvoir impacte-t-il la vie des citoyens ? Et pour terminer, comment l'identité humaine se réinvente-t-elle à l'aune de l'automatisation ?

ARCHITECTURE

La démonstration qui suit est guidée par le célèbre algorithme PageRank mis au point par Google. Son analyse, fragmentée dans ce texte, permettra de mettre en lumière les enjeux qui découlent de la quantification, et de l'application d'automatismes aux affaires humaines, selon trois thèmes. Le premier est celui de neutralité, une notion souvent associée aux algorithmes et aux machines en générale dans l'imaginaire collectif. Il s'agira de mettre ce mythe à l'épreuve du réel, pour l'infirmier ou le vérifier. Le second, ce qu'on appelle le *Frame Problem* (problème du cadre), une notion incontournable dans la numérisation du monde. Il s'agira de cerner les limites inhérentes aux algorithmes, et la manière dont elles modèlent de manière croissante l'organisation de la société. Quant au troisième, il est si vaste qu'il s'étend sur deux parties. Il concerne l'économie des données, et la manière dont la révolution numérique s'est insérée dans le système capitaliste. Pour commencer, l'analyse traitera de la monétisation de l'espace numérique, pour finalement aboutir à la monétisation des utilisateurs.

Neutralité

II

Google PageRank

1
II

Admettons que je me pose une question. « Qu'est ce qu'un oiseau ? » Mon premier réflexe sera probablement d'ouvrir mon navigateur. Mon deuxième, de taper le terme *oiseau* dans le moteur de recherche automatiquement présent sur la page d'accueil, qui a toutes les chances d'être Google. Mon troisième, de cliquer sur le premier résultat que me fournit la machine, sûre d'y trouver une réponse pertinente.

INTERNET

Un réseau informatique se définit comme un ensemble d'ordinateurs reliés les uns aux autres. Une connexion à un réseau partagé permet aux machines de communiquer entre elles, d'échanger des informations selon différents protocoles. Internet est l'un d'entre eux, le plus célèbre, mondial et public. Dans la vie courante, nous avons tendance à confondre Internet avec l'une des applications qui l'utilise, le système du *World Wide Web* (littéralement « la toile à mesure du monde »). Mais ce n'est pas la seule. Les messageries électroniques, par exemple, se servent aussi de ce réseau, à travers un protocole différent. Le web est un système qui permet de consulter des fichiers sur les différentes machines connectées à l'Internet, grâce à un logiciel appelé navigateur. Ces fichiers sont aussi appelées *pages web*, et sont écrits dans des *langages web*, c'est-à-dire des langages compréhensibles par le navigateur. Ils sont accessibles par le biais d'adresses, commençant par l'acronyme *www* (pour *World Wide Web*). Quiconque se branche sur le réseau

peut par définition, avoir accès à chacun d'entre eux. Le problème, c'est qu'ils se comptent en milliards.

TRI

Les moteurs de recherches sont nés du besoin d'indexer la toile, de référencer les pages selon ce qu'elles contiennent pour pouvoir trouver une information sans pour autant connaître l'adresse où elles se trouvent. Ils constituent l'immense sommaire du monstrueux livre du web, un sommaire dynamique qui bouge selon le titre que les utilisateurs lui donnent, à travers des mots-clés. Les algorithmes constituent une réponse possible au problème majeur du tri. Tapis au cœur des moteurs de recherche, ils déterminent selon une recette prenant en compte des milliers de paramètres quels sont les sites les plus pertinents au vue de la requête présentée. En quelques secondes, ils compilent des milliards d'informations qu'ils traduisent par un classement sans appel, qui offre aux utilisateurs un web simplifié et hiérarchisé.

PAGERANK

En 1998, Sergei Brin et Larry Page, deux étudiants en sciences de l'informatique à Stanford, décident de lancer leur propre machine de tri. Leur idée : modifier les paramètres qui classent les sites. À cette époque, la jeune toile se parcourait déjà par le biais de recherches textuelles, mais le classement obéissait à un algorithme lexical, qui rangeait les sites selon le nombre d'occurrences de mots-clés qu'ils contenaient. Avec la prolifération des pages, s'est très vite fait sentir le besoin non plus seulement de trier les contenus par thématique, mais également de les hiérarchiser, par qualité. Tout l'enjeu se trouvait dans la capacité à mesurer des notions abstraites, tel que la pertinence, la justesse ou la vraisemblance, qu'il parait à première vue impossible de quantifier. Forts de cette intuition, les fondateurs de Google mirent au point une nouvelle machine.

«Plutôt que de demander à l'algorithme de comprendre ce que dit la page, ils vont proposer de mesurer la force sociale de la page dans la structure du web. [...] L'algorithme ordonne les informations en considérant qu'un site qui reçoit d'autres un lien reçoit en même temps un témoignage de reconnaissance qui lui donne de l'autorité.»¹

En s'inspirant du modèle scientifique, qui détermine la valeur d'un article au nombre de fois qu'il est cité par d'autres, l'algorithme PageRank hiérarchise les pages selon leur « quota de citation » par d'autres pages. À savoir, en langage Internet, selon le nombre de liens hypertextes, qui partout ailleurs sur la toile, y renvoient. Il utilise les liens comme proxy^G de la pertinence. Plus une page est citée par d'autres, et plus ces autres pages sont elle-mêmes reprises, mieux elle sera classée.

«L'innovation de Brain et Pain fut d'utiliser la structure ontologique inhérente à la toile elle-même pour évaluer l'information.»²

Là est la grande force de PageRank: le classement n'est pas effectué en rajoutant une variable destinée à classer les pages mais en utilisant une donnée déjà présente, inhérente au fonctionnement du Web. Cette méthode se positionne en mesure « d'autorité »³ en se plaçant « au-dessus »³ du web. Ainsi, elle prétend mesurer les agissements des internautes « sans les influencer » et déceler la pertinence inconsciente que la communauté confère à un contenu. Une force omnisciente, apte à faire régner la justice de l'information.

1. Dominique CARDON, *À Quoi rêvent les algorithmes ? Nos vies à l'heure des big data*, Paris : Seuil, 2015

G. Cf. Proxy, lexique

2. «Brin and Page's innovation was to use the inherent ontological structure of the web itself to evaluate knowledge.» Ed FINN, *What Algorithms Want*, Cambridge : MIT Press, 2017

3. Dominique CARDON, *À Quoi rêvent les algorithmes ? Nos vies à l'heure des big data*, Paris : Seuil, 2015

« Il classe les sites à partir d'un vote censitaire au fondement méritocratique. Dans son principe initial, il considère que le liens enferment la reconnaissance d'une autorité : si le site A adresse un lien vers le site B, c'est qu'il lui accorde de l'importance. Qu'il dise du bien ou du mal de B n'est pas la question ; ce qui importe c'est le fait que A ait jugé nécessaire de citer B comme une référence, une source, une preuve, un exemple ou un contre exemple. Le seul fait de citer enferme le nectar dont le calculateur fait son miel. »³

Le paramètre de citation est loin d'être le seul qui définit le fonctionnement du moteur de recherche Google mais il demeure, le principal, celui qui a jeté les bases de l'empire. La composition exacte des critères de l'algorithme est jalousement gardée secrète, pour, selon l'entreprise, « éviter que les webmasters ne trichent et le favorise artificiellement. »⁴ Les résultats proposés seraient, simplement, objectivement les meilleurs.

QUI EST « IN » ET QUI EST « OUT »

« À la base, PageRank catalogue l'acte humain de juger, comptant les liens et l'intérêt correspondant que des millions de gens ont porté dans différents coins du web. [...] C'est un outil pour hiérarchiser le monde, classant l'information non d'après ses qualités intrinsèques mais d'après une certaine définition de l'importance ou de la popularité. »⁵

Dans le cas de PageRank, la popularité n'est pas directe (comme le serait le fait de se baser seulement sur le nombre de visiteurs) mais indirecte, au travers du nombre liens : c'est ce que Cardon appelle la « force sociale » d'une page.

4. Dominique CARDON, *À Quoi rêvent les algorithmes ? Nos vies à l'heure des big data*, Paris : Seuil, 2015

5. « At his heart, PageRank catalogs the human act of judgment, counting up the links and relative attention that millions of people have paid to different corners of the web. [...] This a tool for creating hierarchy in the world, sorting information not merely by its inherent qualities but by a certain definition of importance or popularity. » Ed FINN, *What Algorithms Want*, Cambridge : MIT Press, 2017

EFFICACITÉ

Il faut commencer par admettre l'efficacité fulgurante de la machine PageRank. Si le moteur de recherche de Google est utilisé quotidiennement par des milliards d'individus chaque jour, c'est pour la simple raison qu'il fonctionne très bien. Cela étant dit, la critique peut débuter.

NOMBRE DE LIENS = POPULARITÉ = MAJORITÉ = PERTINENCE

Le fait que PageRank et ses enfants se basent sur une mesure de la majorité pose trois problèmes majeurs :

1. *La rigidité du classement*

Il faut être populaire pour être bien classé et être bien classé rend populaire. Dans un tel contexte, il devient de plus en plus difficile pour de nouveaux acteurs d'apparaître : le classement comporte une forme d'auto-réalisation.

2. *L'ignorance du contenu*

Ce que cherche à savoir PageRank est : « Le contenu de cette page est-il pertinent ? ». Mais l'algorithme ne peut avoir accès à cette information, qui renferme un monde de débats en elle-même, et ne se quantifie pas. Il doit donc se baser sur un proxy, la force sociale. Le contenu d'un site en lui-même n'est à aucun moment pris en compte par l'algorithme. Si vous entrez le mot *avortement*, vous trouverez des sites de désinformation côte à côte avec le planning familial. Autrement dit, il répond à la question : « Cette page est-elle pertinente pour la majorité des internautes qui s'expriment sur le web ? »

3. *Le biais de la popularité*

Le classement basé sur la force sociale implique de faire confiance au plus grand nombre pour dégager la pertinence. Le choix des liens hypertextes comme proxy n'est absolument pas neutre. La majorité n'est pas la neutralité.

Traduction automatique

2 II

Pendant longtemps, l'approche de la traduction automatique a été celle d'une « Intelligence artificielle forte »^H : on a tenté d'enseigner la grammaire, la syntaxe aux systèmes, ce qui s'est soldé par un vaste échec. Les progrès fulgurants des dernières années sont dû à un changement radical de méthode : au lieu d'enseigner les règles d'une langue, les *data scientists* ont pris le parti de passer par des milliards d'exemples de traductions passées, à partir desquelles la machine est capable d'en déduire de nouvelles. Pour arriver à ce résultat, ils se sont basés sur une nouvelle technologie, intitulée le *Machine Learning*^H.

WORD EMBEDDINGS

La traduction automatique utilise des algorithmes dits de *Word Embedding* (incrustation de mots) comme GloVe,⁶ qui établissent des liens entre les mots, puis les cartographient. Ils sont ainsi capables de comprendre les relations que les termes entretiennent, et d'offrir une traduction appropriée au contexte. Pour établir ces liens, les *data scientists* ont au préalable « nourri » l'algorithme avec des millions de textes trouvés sur la toile, afin de lui « apprendre » les langues.

IL EST AVOCAT

Trois chercheurs de l'université de Princeton ont étudié cette technologie et ainsi découvert qu'elle faisait écho aux discriminations qui gangrènent la société. En effet, étant basée sur toutes sortes de textes présents sur le web, elle reflète fidèlement les associations de mots le plus souvent écrites. Prenons par exemple la traduction automatique de Google qui utilise un modèle similaire à *GloVe*, *Wolrd2vec*. Le turc est une langue non genrée. Lorsque nous traduisons du turc vers l'anglais « *O bir avukat* », « Il ou elle est avocat[e] », nous obtenons « *HE is a lawyer* », « IL est avocat ». Lorsque

H. Cf. *Machine Learning*, lexique

6. *Global Vectors for Word Representation*, [<https://nlp.stanford.edu/projects/glove/>]

nous traduisons «*O bir doktor*», «Il ou elle est médecin», nous obtenons «*HE's a doctor*», «IL est médecin» alors que «*O bir hemşire*», «Il ou elle est infirmière[è]r[e]» donne «*SHE is a nurse*», «ELLE est infirmière». La position la plus haut placée est automatiquement masculine, et la plus basse, féminine. Même chose avec *professor* (enseignant[e] d'université) et *teacher* (enseignant[e] du primaire au secondaire) qui sont automatiquement associés à un genre. «*HE is a professor*» / «*SHE is a teacher*»

«Nous avons appris comment nous perpétuons des préjugés que nous ne savions même pas que nous avions.»⁷

Les biais dont nous parlons ici ne sont pas présents dans le code lui-même. L'algorithme n'a pas été délibérément construit pour être sexiste. Il reflète simplement les données sur lesquelles il a été entraîné. Il reflète simplement les biais de la société.

«Les découvertes de cette étude sont pas surprenantes, puisque les systèmes de *machine learning* sont seulement capable d'interpréter les données à partir desquelles ils sont entraînés.»⁸

Ce phénomène est néanmoins très intéressant car il met l'accent sur l'importance du choix des données. Le non-choix en est un : ici, celui de une majorité fantômes des gens qui s'expriment sur le web. Fantôme, car les internautes ne participent pas consciemment aux discriminations que vont perpétrer, à grande échelle, les applications qui se basent sur leurs mots. Il y a donc une responsabilité des concepteurs, dans le fait de ne pas trier les données. Cela peut paraître surprenant, mais utiliser sans discrimination tous les contenus aboutit à des algorithmes discriminants. Nos miroirs numériques sont formels : la justice, plus que l'injustice, est un choix.

7. «*We have learned something about how we are passing on prejudices that we didn't even know we were doing*» Joanna BRYSON, chercheuse à l'université de Princetown, propos cité dans un article de Matt BURGESS, *Just like humans, artificial intelligence can be sexist and racist*, *Wired UK*, 13.04.17

8. «*The findings of the study do not come as a massive surprise, as machine learning systems are only able to interpret the data they are trained upon.*», Matt BURGESS, *Just like humans, artificial intelligence can be sexist and racist*, *Wired UK*, 13.04.17

COMBATTRE LES BIAIS

« Pour prendre conscience de ces biais, nous avons besoin de transparence algorithmique. »⁹

L'opacité algorithmique défendue par les géants du web pour des raisons commerciales pose des problèmes éthiques. Comment savoir si les modèles qu'ils utilisent ne sont pas biaisés si le code est fermé ? Outre la recette, la transparence doit également comprendre les ingrédients, à savoir les données linguistiques sur lesquelles la machine fut entraînée. Ne pas trier signifie également implémenter des propos haineux et violents, dont regorge la toile. Est-il acceptable de construire un modèle de traduction qui les utilisent ?

« Pour supprimer les biais, nous devons les quantifier dans les modèles. »¹⁰

Afin de débusquer ces injustices, l'équipe de Princeton a mis au point un système de mesure, basé sur un test d'études psychologique de 1990 baptisé « Test d'Association Implicite » (*IAT: Implicit Association Test*). Le test fonctionne ainsi : on a demandé à un panel de volontaires d'associer des concepts avec des images deux par deux, tout en mesurant le temps qu'ils mettaient. Par exemple, *fleur* est très rapidement associé avec *positif* et *insecte* avec *négatif*. Le « temps de latence » est ainsi utilisé pour mesurer à quel point le concept est connoté positivement ou négativement. Ce test a permis de révéler de forts biais envers certaines populations, selon leur origine ou leur genre par exemple. Les scientifiques ont fait passer ce même test au modèle GloVe, en remplaçant le temps de latence par la distance qui sépare deux termes, qui représente la fréquence à laquelle ils sont associés. Les résultats furent sans appel. « Un échantillon de noms associés avec le fait d'être caucasien-américain se révélèrent être associés avec des termes positifs plutôt que négatifs de manière significative, comparé à un échantillon similaire de noms afro-américains »¹¹ Au menu, figuraient également les associa-

9. « *In order to be aware of those bias, we need algorithmic transparency.* »

10. « *In order to debias, we need to quantify bias in those models.* » CALISKAN-ISLAM Aylin, A *Story of Discrimination and Unfairness*, Chaos Communication Congress : Hambourg, 27.12.16,

11. « *A bundle of names associated with being European American was found to be significantly more easily associated with pleasant than unpleasant terms, compared to a bundle of African*

tions homme / carrière, femme / famille ou le sympathique garçon / science, fille / art.

ENJEUX

Révéler ces biais constitue un premier pas vers des modèles plus justes. Néanmoins d'autres questions se posent comme, comment atténuer ces biais tout en conservant l'utilité des modèles ? Ou encore, combien de temps les biais persistent-ils dans les modèles ? Si nous ne faisons rien, vont-ils provoquer un « effet boule de neige » en banalisant les discriminations à travers leurs applications ?¹²

L'algorithme qui termine automatiquement les mots sur votre smartphone se base sur le *Word Embeddings*. Combien de fois vous guide-t-il par jour ? La question des biais est générale, et nous touche au quotidien. Il faut qu'elle soit publiquement débattue et collectivement tranchée, secret commercial ou pas.

PREDPOL® The Predictive Policing Company®

3
II

« *PredPol* utilise l'intelligence artificielle pour vous aider à combattre le crime, en prédisant quand et où il est le plus susceptible de se produire, vous permettant ainsi d'optimiser vos équipes et de mesurer leur efficacité. »¹³

Comme son nom l'indique, *PredPol* est un logiciel qui a pour ambition de prédire le crime. Mis au point en 2009 par une start-up de Santa Cruz (US)

American names » Aylin CALISKAN-ISLAM, Joanna J. BRYSON, Arvind NARAYANAN, Semantics derived automatically from language corpora necessarily contain human biases, Princeton University, 30.8.2016

12. CALISKAN-ISLAM Aylin, *A Story of Discrimination and Unfairness*, Chaos Communication Congress : Hambourg, 27.12.16

13. « *PredPol® uses artificial intelligence to help you prevent crime by predicting when and where crime is most likely to occur, allowing you to optimize patrol resources and measure effectiveness.* » Page d'accueil de *PredPol*

portée par un ancien professeur d'anthropologie, il pointe des zones géographiques où des actes criminels ont le plus de chances de se produire, dans une ville donnée. Selon le principe des algorithmes prédictifs, celui de *PredPol* fonctionne en analysant les données historiques, en s'inspirant d'un logiciel de prédiction des tremblements de Terre. Trois facteurs entrent en jeu :

1. Le temps : Quand ?
2. Le lieu : Où ?
3. Le type : Nature du crime ?

Sur une carte, le logiciel entoure, heure par heure, les zones les plus sensibles selon son expertise. Le but étant comme précisé dans leur slogan, d'optimiser les équipes de policiers. Il est à ce jour utilisé dans plusieurs villes américaines, comme Los Angeles ou Atlanta. Des émules ont également fleuri en Europe, en Italie, en Allemagne et même en France. Avec la différence assez notable qu'en France, il ne sont pas développés par des entreprises privées mais par les institutions elle-mêmes, pour des raisons de sécurité. L'analyse que nous faisons ici de *PredPol* se place dans son contexte américain. Il serait donc possible de prédire les crimes futurs d'après ceux du passé.

INÉGALITÉ

Comme le souligne la *data scientist* Cathy O'Neil, le principal problème du logiciel se trouve dans la nature des crimes implémentés. Les policiers ont le choix : soit se concentrer sur les crimes violents, soit incluent les délits mineurs, comme le vagabondage, la consommation de drogues, ceux qu'on appelle parfois *Anti Social Behaviour* (ASB). Les délits sans victimes, « qui seraient probablement passés inaperçus si un flic n'avait pas été là pour les voir ». ¹⁴ Qui se produisent surtout dans les quartiers défavorisés, et sont une conséquence notoire de la pauvreté.

Prendre en compte ce type délit crée un cercle vicieux, qui conduit à la fabrication d'encore plus de données pointant vers les quartiers pauvres, là où une patrouille plus diversifiée aurait ajouté des données différentes. « La présence de la police crée de nouvelles données qui justifient plus de présence. » ¹⁵ Parallèlement à ce phénomène, O'Neil remarque que dans nos villes ghettoisées, « la géographie est un proxy très efficace pour l'origine

14. « Many of these "nuisance" crimes would go unrecorded if a cop weren't there to see them. »

15. « The policing itself spawns new data, which justifies more policing. »

ethnique. »¹⁶ Bien sûr, le logiciel n'implémente aucune donnée de ce type, ce qui serait illégal. Mais dans les faits, c'est tout comme.

Notons également que la « délinquance en col blanc » ne fait pas partie des données qu'il est possible de cartographier. Notamment car les policiers ne sont pas formés pour faire face à ce type de criminalité, qui pourtant est un véritable carnage. Alors que les fraudes économiques ont provoqué les crises successives des années 2000, qui jetèrent des gens à la rue et détruisirent des millions d'emplois, « la finance reste sous surveillée, grâce à l'argent de l'industrie et de puissants lobbies. »

¹⁷ La tolérance 0 doit concerner tous les délits. Ou ne pas être.

« Le résultat est que nous criminalisons la pauvreté, en croyant que nos outils sont non seulement scientifiques mais également justes. »¹⁸

On est bien loin de la technique aveugle aux biais humains. Ce qui est assez troublant dans l'utilisations de tels algorithmes, c'est la justification de discriminations *par* la technique (et non à cause d'elle), sous couvert d'une prétendue neutralité du calcul. Cela conduit à une déresponsabilisation des hommes qui agissent, et ne font que « suivre ce dit la machine »; machine vendue, ne pas l'oublier, par une entreprise à but commercial.

« La décision de condamner une personne à une peine de prison est encore prise par des êtres humains. Mais une multitude de petites décisions : condamner un automobiliste à une amende, accorder ou refuser un prêt à la consommation, etc., sont déjà prises par des algorithmes et nous avons tous déjà entendu l'antienne qui prend les algorithmes, ou les ordinateurs, comme boucs émissaires [...] Une telle dictature de la "médiocrité assistée par ordinateur" est à ranger dans la même catégorie que les autres formes de pouvoirs abusifs, face auxquels aucune procédure de contestation n'est prévue. »¹⁹

16. « *In our largely segregated cities, geography is a highly effective proxy for race.* »

17. « *Thanks largely to industry's wealth and powerful lobbies, finance is underpoliced.* »

18. « *The result is that we criminalize poverty, believing all the while that our tools are not only scientific but fair.* »

Cathy O'NEIL, *Weapons of math destruction: how big data increases inequalities and threatens democracy*, Crown: New York, 2016

19. Serge ABITEBOUL, Gilles DOWEK, *Le Temps des algorithmes*, Paris : Le Pommier, 2017

SOLUTIONS ÉCONOMIQUES

«Je ne vais pas avoir plus d'argent. Je ne vais pas avoir plus de flics. Je dois faire mieux avec ce que j'ai, et c'est ce que fait la police prédictive...Si ce vieux flic de la rue peut changer d'avis sur ça, je sais que mes officiers peuvent le faire aussi. »²⁰

PredPol et ses copies, dans le fond, ne répondent pas vraiment au problème de criminalité mais à un autre fléau, qui fait tout autant rage : les coupes budgétaires dans la fonction publique. Si les services de polices se dotent de ces logiciels, c'est d'abord pour palier au manque d'effectifs. Mais, comme le dit Nils Hamer, un policier de Nuremberg, « Le logiciel [de police prédictive] un bon outil de soutien, mais il faut tout de même garder son flair policier et son bon sens humain. Un tel logiciel ne remplacera jamais un policier. »²¹

EFFICACITÉ DOUTEUSE

PredPol annonce des résultats spectaculaires. Dans les villes où il a été mis en place, il y aurait « une baisse de la criminalité de 10 à 30% selon le type de crimes. » D'après d'Ismaël Benslimane, chercheur à l'université de Grenoble, *PredPol* est « d'abord un outil d'enregistrement et de classification des délits. [...] Si *PredPol* est efficace là où il est déployé, que sait-on de l'augmentation de la criminalité où il n'est pas présent ? ». Autrement dit, le logiciel mesure sa propre efficacité, en permettant de récolter des données qui avant n'étaient pas enregistrées. Comme l'explique le chercheur, la prédiction est dans presque tous les cas validée, ce qui conduit à des résultats biaisés.

FONCTIONNEMENT OPAQUE

Outre la perpétration d'injustices flagrantes, ce type de système est anti-constitutionnel tout simplement car l'algorithme utilisé est gardé secret. Il est inadmissible que les pouvoirs publics soient autorisés à suivre les directives

20. « *I'm not going to get more money. I'm not going to get more cops. I have to be better at using what I have, and that's what predictive policing is about... If this old street cop can change the way that he thinks about this stuff, then I know that my [officers] can do the same.* » Charlie BECK, Chef de la police de Los Angeles *Page d'accueil de PredPol*,

21. [Réalisateur inconnu], *Police prédictive*, *Futuremag*, Arte, 19.11.2016

d'un logiciel sans que sa composition exacte ne soit publiée. Les boîtes noires de n'importe quel ordre doivent être bannies des institutions. Il faut que la révolution numérique cesse d'être de le fond de commerce des techniciens, qui s'enrichissent sur l'asymétrie des connaissances et entretiennent l'image d'un fonctionnement quasi-magique de leurs artefacts. Il n'y a absolument aucune magie derrière de tels dispositifs, qui ne sont que des calculatrices à comportements auto-réalisatrices, qui pallient tant bien que mal le manque de moyens humains.

« *PredPol*, pour moi, c'est un algorithme de droite. Il permet de réduire les dépenses publiques, les effectifs dans la police, de faire des économies. »²²

Ce qui leur permet d'exister, c'est le libéralisme des états, qui laissent se creuser les inégalités dans leurs propres pays, à travers des politiques fiscales injustes qui tuent peu à peu les services publics. C'est dans cette brèche là que *PredPol* et ses émules viennent « charogner », il ne faut pas l'oublier.

SÉCURITÉ VS ÉQUITÉ

« Le jugement implicite de la constitution est que libérer quelqu'un qui aurait commis un crime pour manque de preuves, représente un danger moindre pour notre société qu'emprisonner ou exécuter un innocent. Les mauvais algorithmes ont tendance à favoriser l'efficacité. C'est leur nature même que de se nourrir de données qui peuvent être mesurées et comptées. Mais l'équité est [...] difficile à quantifier. C'est une idée. »²³

22. Bilel BENBOUZID, Sociologue, dans un entretien de Camille POLLONI, *Police prédictive : la tentation de « dire quel sera le crime de demain »*, Rue 89, 27.05.15

23. « The Constitution's implicit judgment is that freeing someone who may well have committed a crime, for lack of evidence, poses less of a danger to our society than jailing or executing an innocent person. WMDs, by contrast, tend to favor efficiency. By their very nature, they feed on data that can be measured and counted. But fairness is [...] hard to quantify. It's a concept. », Cathy O'NEIL, *Weapons of math destruction: how big data increases inequalities and threatens democracy*, Crown : New York, 2016

Le film de Spielberg *Minority Report* (2002)²⁴ illustre très bien une société où la sécurité l'emporterait sur l'équité. Lui aussi traite de police prédictive, en mettant en scène une dystopie, où 3 individus aux facultés surnaturelles voient les meurtres à l'avance. Dans ce 2054-ci, les hypothétiques criminels sont arrêtés avant de pouvoir passer à l'acte. Et sont emprisonnés alors « qu'ils n'ont rien fait » puisque les crimes sont empêchés. Impossible de savoir s'ils auraient vraiment commis le crime pour lequel ils sont punis. La sécurité prime. Il nous faut attention à ce que la technique, sous couvert d'être « neutre » ne soit utilisée comme exactement l'inverse : une machine à discriminations, allant à l'encontre de la loi et réduisant à néant les avancées sociales que le XX^e siècle a arrachées. Et, dans le cas de *PredPol*, ne transforme la justice en appareil de sauvegarde automatique de l'ordre social, qui protégerait les puissants et punirait la pauvreté au lieu de faire appliquer équitablement des lois communes à tous.

24. Steven SPIELBERG, *Minority Report*, 2002

The Frame Problem

III

« POURQUOI UTILISEZ-VOUS LE MOTEUR DE RECHERCHE GOOGLE ? »

En 2012, le journaliste et blogger suédois Andreas Ekström a posé cette question à quelques-uns de ses étudiants.¹ Leurs réponses illustrent bien ce que pensent une majorité d'internautes :

1. Parce que ça marche → Critère d'efficacité
2. Parce qu'ils ne connaissent pas d'autres alternatives → Critère d'exclusivité
3. Parce qu'ils sont certains de toujours obtenir le meilleur et le plus neutre des résultats → Critère d'objectivité

Cette idée d'une machine parfaitement neutre assistant les hommes instables, dévorés de sentiments, sous-tend fortement la mythologie des machines. La rationalité numérique, comme remède, comme arbitre des passions. Les nombres en juges impartiaux des affaires humaines, entachées d'intérêts personnels et de pulsions déraisonnables.

« Ces principes ne sont pas nouveaux. Ils partagent de nombreux traits avec les formes de nombreux traits avec les formes de représentation traditionnelles de la société. Simplement, leur manière d'asseoir leur légitimité sur des principes prétendument objectif leur confère une légitimité tout particulière »²

1. L'histoire reprise ici est tirée de la conférence d'Andreas EKSTRÖM, *The moral bias behind your search results*, TEDxOslo : Oslo, 01.2015
2. Dominique CARDON, *À Quoi rêvent les algorithmes ? Nos vies à l'heure des big data*, Paris : Seuil, 2015

CLASSEMENT D'IMAGES

Le fait-divers rapporté ici date de 2011, et se base sur la méthode d'indexation des images utilisée à l'époque. L'arrivée du *machine learning* a bousculé les pratiques, et désormais les moteurs de recherches sont de plus en plus à même de reconnaître directement ce que les clichés représentent. Mais cela n'impacte pas le fond de la démonstration qui suit.

Lorsqu'il s'agissait de sélectionner des images, les moteurs de recherche ne les « identifiaient » pas à proprement parler de la même manière qu'ils identifient les mots. C'est le texte qui les entoure à savoir, premièrement le nom de l'image et deuxièmement sa légende qui indiquaient à l'algorithme, aveugle, ce qui était représenté. Ainsi, si je publie une image de perdrix, j'aurais plus de chance de la voir indexée au mot-clé « perdrix » si je la nome *perdrix.jpg* et si j'indique « La perdrix » dans un texte à côté. La mesure du partage des liens est ici également valable : plus d'autres internautes partagent, publient mon image sous les mêmes termes, mieux elle sera classée dans les résultats de recherche.

Manipulation

1

III

En 2009, l'ex-première dame des États-Unis, Michelle Obama fut victime d'une campagne raciste, qui vit circuler sur la toile un montage photo d'elle, la représentant sous les traits d'un singe. Lorsque l'on entrait son nom dans le champ de recherche Google, on pouvait voir cette image apparaître parmi les premiers résultats. Les auteurs de cette propagande ont tenté de l'insulter par le biais des recherches. Ils se sont assurés de remplir les champs en se calquant sur la méthode de l'algorithme, pour que cette photo apparaissent dans le peloton de tête de la requête « Michelle Obama ». Les modérateurs de Google sont alors intervenus, pour supprimer manuellement l'image de leurs pages. Cet événement montre clairement que les résultats sont manipulables. Pas seulement par Google lui-même, qui peut directement les modifier comme il l'a fait, mais aussi par les internautes, rats de laboratoires récalcitrants, qui détournent l'expérience en ayant conscience du protocole. L'idée du moteur de recherche au-dessus du web, en grand analyste partial de son contenu, est un fantasme.

Le 22 juillet 2011, le terroriste nazi Anders Brejvik perpétra un attentat à la bombe dans la capitale norvégienne d'Oslo (8 victimes) avant d'assassiner de sang-froid 69 personnes sur l'île d'Utoya, alors qu'elles participaient à un camp d'été. Ce n'est pas seulement en vertu des misérables doctrines d'extrême-droite auxquelles il adhérait que cet homme passa à l'acte, mais aussi dans un désir morbide de célébrité. Nikke Lindqvist, ingénieur suédois spécialisé dans l'optimisation des moteurs de recherche, entrepris d'exprimer sa révolte face à la monstrosité de ce carnage, tout en sabotant sciemment l'entreprise de reconnaissance publique de Brejvik.

Utilisant son réseau, il dit à tous ceux qui le suivait sur le net que cet homme voulait contrôler son image et qu'il fallait la saboter. Pour ce faire, il lança une campagne pour l'insulter à travers les recherches Google et ainsi protester contre son acte. Il dit à tous ces lecteurs de trouver des photos de crottes de chiens sur les trottoirs, de bien s'assurer d'écrire le nom du terroriste en légende, de bien nommer les photos d'après lui puis de les publier, de les partager un maximum. Et ça a marché. En Suède, dans la semaine suivant cette campagne de manipulation, en tapant le nom du terroriste dans le moteur de recherche, on trouvait des images de crottes de chien parmi les premiers résultats. Mais cette fois-ci, les modérateurs ne sont pas intervenus. Entre ces deux campagnes de diffamation, il n'y a aucune différence, si ce n'est l'intervention de Google.

POURQUOI LA TRICHE EST-ELLE POSSIBLE ET SERA-T-ELLE TOUJOURS POSSIBLE ?

Comme expliqué en introduction, les algorithmes sont des modèles, c'est à dire des représentations. Pour exister, ces représentations se basent sur des proxys, en l'occurrence le nombre de liens pour la pertinence. Les proxys sont des signes, des facteurs. Prenons un exemple hors du net. Une personne boite. Nous pouvons en déduire que sa jambe la fait souffrir. Mais peut-être que non. Peut-être a-t-elle un caillou dans sa chaussure. Peut-être fait-elle semblant. Peut-être est-elle rechigne-t-elle juste à avancer. Le fait de boiter peut être un signe de douleur. Mais il n'est pas la douleur elle-même. Les algorithmes tel que PageRank raccourcissent automatiquement la causalité à boiter = souffrir. Nombre de liens = pertinence. Le problème avec de tel raccourcis, c'est que :

1. Qui trouve le chiffre à influencer manipule la symbolique entière. Étant des représentations, ils seront toujours falsifiables.
2. Ces raccourcis sont, triche ou pas triche, partiellement faux.

PROXYS, PROXYS, PROXYS...

De retour hors de l'Internet. Vous êtes-t-il déjà arrivé de rencontrer quelqu'un avec un savoir totalement incongru ? Du genre, un passionné de pêche qui vous dresse une zoologie des rivières en moins de 5 minutes, ou un fou de plantes qui vous explique en des termes botaniques obscures comment faire grandir la meilleure, la plus heureuse des courges ? Admettons que notre Monsieur Courge, en plus de la botanique, aime bien bidouiller avec l'ordinateur. Alliant ces deux hobbies, il se crée un petit site, *mon-beau-potager.net*, où il partage ces expériences et donne des conseils. Il y a vraiment très peu de chances qu'avec les mots-clés « faire pousser » et « courge », son site soit le premier dans la liste du moteur de recherche. Et ce sera, non pas parce-qu'il n'est pas pertinent mais parce-qu'il ne remplit pas les critères que l'algorithme associe à la pertinence.

DE COURGE EN COURGE

Si le site de Monsieur Courge reste perdu dans les tréfonds du net alors qu'il renferme des trésors en matière de botanique, c'est sans doute parce que Monsieur Courge n'est pas un webmaster et ne sait pas comment contenter la machine de recherche. Par exemple, il ne sait pas qu'il est possible d'associer des mots-clés spécifiques avec sa page. Il ne sait pas non plus, qu'il existe des dizaines de techniques plus ou moins fourbes pour améliorer son référencement. Monsieur Courge est naïf, et pense lui aussi que les résultats de Google sont les meilleurs.

Pour organiser le web, l'utilisation de machines est simplement indispensable. Devant une telle immensité, il serait impossible d'imaginer une curation de l'espace numérique entièrement à mains d'hommes, où des personnes regardent tous les sites, les évaluent, puis les classent. Mais pour que la curation automatique fonctionne, il faut parler aux algorithmes en termes qu'ils comprennent, en décrivant nous-mêmes l'espace. Les mots-clés sont un exemple, les **#hashtags** en sont un autre.

« Désormais [Google] vous aide à enregistrer le nom de domaine, construire le site, analyser le trafic et trouver des publicités pour votre site, que les algorithmes indexeront ensuite et classeront. »³

Nous voyons donc que l'évaluation de la pertinence par un moteur de recherche englobe aussi la capacité du créateur à communiquer avec la machine, capacité à comprendre les critères de l'algorithme qui n'est en rien liée au contenu du site. Il nous faut rentrer dans le jeu de la machine, et suivre ses règles, pour espérer faire partie du monde numérique. Et il est dans l'intérêt de Google que tout le monde s'y conforme, rentrant ainsi, dans le domaine des données qu'il exploite.

PARLER L'ALGORITHME

Imaginons que je demande l'adresse d'un dentiste à quelqu'un. Il est peu probable que je le regarde droit dans les yeux et lui dise « dentiste Strasbourg pas cher ». Mais c'est ce que je ferais avec un moteur de recherche. L'adaptation se vérifie aussi pour qui cherche l'information. Nous modifions tous notre langage, traduisons nos désirs dans des termes compréhensibles par les algorithmes. Et ce langage n'est *pas* si évident qu'il n'y paraît. Converser avec un moteur demande de l'entraînement et des compétences. Les apprentis codeurs en qui en ont une pratique quasi perpétuelle le réalisent vite. Constamment bloqués par des problèmes techniques, il leur faut partir en quête de « la personne qui a déjà eu ce problème et de celle qui l'a résolu », et cette quête passe 99 fois sur 100 par une requête Google. Avec le temps, ils développent une véritable stratégie des termes, pour utiliser au mieux la machine et être sur de trouver une réponse pertinente. Chercher aussi, s'apprend.

3. « *Now the company will help you register the domain name, build the website, analyse the traffic and serve multiples for the site, which its algorithms will then index and rank.* » Ed FINN, *What Algorithms Want*, Cambridge : MIT Press, 2017

GEEKS ET NON GEEKS

« Et je ferais en sorte que [ma conférence] soit compréhensible par tous, et pas seulement par les geeks et les informaticiens. »⁴

Mais les règles qui régissent la machine sont loin d'être connues de tous. D'une part car les algorithmes sont gardés jalousement secrets, ce qui donne, s'il est utile de le préciser, un immense avantage à leurs créateurs. Et d'autre part, car la transition numérique est à l'heure actuelle loin d'être évidente. Alors que l'espace public se déplace de manière croissante sur la toile, rares sont les citoyens qui en possèdent les véritables clés. Ce qui donne, s'il est aussi utile de le préciser, un immense avantage aux informaticiens.

LA CARTE OU LE TERRITOIRE

« Alors que Google a construit son fonctionnement sur une cartographie précise de la culture sur Internet, son rôle est progressivement passé de celui d'indexeur à celui d'architecte. »⁵

À ses débuts, le PageRank se basait sur la structure existante du web, et tirait son classement, sa cartographie. Mais croissant petit à petit en importance, l'influence s'est inversée, et les hommes se sont mis à dessein à parler à PageRank, car ne pas être cartographié revenait à ne pas exister. Il ne s'agit pas de l'influencer artificiellement mais au minimum de faire en sorte de rentrer dans la machine, de parler, en des termes compréhensibles par les algorithmes. Comme le dit Finn, « la machine culturelle Google assemble une carte qui à terme, menace de remplacer le territoire. »⁶

4. « Und ich werde alles tun so dass es für jeder, und nicht nur für Tekies und Informatiker vestehbar ist. » David KRIESEL, *SpiegelMining – Reverse Engineering von Spiegel-Online*, Chaos Communication Congress : Hambourg, 28.12.2016

5. « While Google built its buisness on accurateley mapping the cultural ontologies of the Internet, its role has graudally shifted from indexer to architect. » Ed FINN, *What Algorithms Want*, Cambridge : MIT Press, 2017

6. « The Google culture machine is assembling a map that at times threatens to upstage territory. »

DÉDRAMATISONS : TAPEZ « DESIGN GRAPHIQUE »
DANS UN MOTEUR DE RECHERCHE OU LA COHABITATION
DE DEUX MODÈLES

Dans son livre *What Algorithms Want*, Ed Finn, théoricien américain du numérique et imagination, avance que le web algorithmique est en passe de supplanter le web tel qu'il existait auparavant, c'est-à-dire sur le modèle d'une curation humaine, sous le forme « site cool du jour » (*cool site of the day*), où le tri n'est pas effectué par un dispositif automatique, et repose sur des « réseaux non cartographiés » (*unmapped networks*). Mais tous les artistes en herbe savent que ce n'est pas près d'arriver. Si vous tapez « design graphique » ou mieux « art numérique » dans la barre de recherche Google, vous avez toutes les chances de tomber sur les pires travaux de l'univers. La curation humaine n'a pas disparue : elle cohabite avec l'algorithmique. En sont la preuve les dizaines de sites qui répertorient des travaux artistiques, type *itsnicethat.com* ou *booksfromthefuture.tumblr.com*. Ces listes ne sont pas faites à partir de machines, mais sont le fruit d'êtres humains qui écument les tréfonds de l'Internet. Ils fonctionnent également beaucoup en collaboration avec leurs visiteurs, qui sont encouragés à soumettre des travaux. On se les échange « à l'ancienne », comme des bonnes adresses où se balader. D'ailleurs, notre Monsieur Courge aurait plus de chance de se retrouver sur *site-de-jardins.org* que parmi les premiers résultats PageRank.

Ces systèmes sont une autre manière de trier l'Internet, sans doute plus efficace que l'algorithmique mais qui sont unimaginables à l'échelle de tout le réseau. De plus, « trouver les bons sites » implique soit de faire partie d'une communauté d'initiés, soit de retourner la toile *et* d'avoir de la chance. Pour la défense du modèle PageRank, nous pouvons dire qu'il est bien plus démocratique pour qui cherche l'information. Pour la défense du modèle « à la main », nous pouvons dire qu'il est plus démocratique pour qui la donne. Disons, en définitive, que les deux modèles ne s'excluent pas l'un l'autre. C'est leur collaboration active qui construit la toile.

LA TRICHE NOUS REND PLUS SAGE

Un algorithme est aveugle. S'il peut se révéler un formidable outil dans la recherche d'informations, il reste néanmoins une machine, seulement capable de « mimer » la compréhension et non de l'exercer. Il est important de se souvenir de la cécité intrinsèque des dispositifs automatiques face aux cas particuliers, qui sortent de leurs dispositifs de compétences, dont la manipulation fait partie. Comme le dit Luciano Flordini, « parfois nous oublions que les systèmes d'intelligence artificielle les plus efficaces sont ceux qui ont la chance d'avoir un environnement pré-modélisé autour d'eux. [...] La véritable difficulté [...] est de faire face [...] à l'imprévisible. C'est ce qu'on appelle le « *Frame Problem* » (problème du cadre). »⁷

TONDEUSES HIGH-TECHS

Pour illustrer cette notion, Flordini prend l'exemple très parlant des tondeuses à gazon autonomes. « Comme leur nom l'indique, ce sont des machines autonomes qui peuvent tondre le gazon. Elles sont aussi stupides que votre vieux réfrigérateur. Pour qu'elles fonctionnent bien, vous devez installer un fil d'enclos, qui définit la zone à tondre. Le robot l'utilise pour localiser la limite du gazon et parfois aussi pour trouver un point de recharge. Vous devez adapter l'environnement au robot pour vous assurer qu'ils puissent communiquer. De la même manière, plongez les AI dans leur soupe digitale, l'Internet, et elles y batifoleront joyeusement. »⁸

7. « Sometime, we may forget that the most successful AI systems are those lucky enough to have their environment shaped around them. [...] The real difficulty is to cope [...] with the unpredictable. [...] This is known as the Frame Problem. »

8. « As their names indicate, they are autonomous machines that can mow the lawn. They are as stupid as your old refrigerator. In order to function properly, you need to set up a border wire that defines the area to be mowed. The robot can then use it to locate the boundary of the lawn and sometimes to locate a recharging dock. You need to adapt the environment to the robot to make sure the latter can interact with it successfully. Likewise, put artificial intelligence in their digital soup, the Internet, and you will find them happily buzzing. »
Luciano FLORIDI, *The Fourth Revolution: How the Infosphere is Reshaping Human Reality*, Oxford : Oxford University Press, 2016

Construire l'enclos

2

III

« De nos jours, l'enveloppement de l'environnement dans des termes compréhensibles par les TIC (Technologies d'Information et Communication) a commencé à supplanter tous les aspects de la réalité et se constate partout, au quotidien. »⁹

PURES INFORMATIONS

« Sur le plan industriel et sociétal, on voit de plus en plus de métiers bien établis se faire concurrencer, voire laminier par de nouveaux entrants, qui comprennent qu'il est plus important de posséder et de canaliser l'information que de posséder les objets physiques comme dans les activités traditionnelles. »¹⁰

L'avènement des « TIC » et de la société de l'information ont également profondément modifié l'organisation du commerce. Ils ont donné naissance à ce qu'on appelle « l'économie du partage » (*sharing economy*), qui est basée non plus sur la possession de biens comme l'est traditionnellement le capitalisme mais sur la possession de la seule information. Uber ou AirBnb sont emblématiques de cette mutation. Ils ne possèdent rien : aucun véhicule, aucun logement. Rien si ce n'est des serveurs, sur lesquels transitent les données. Ils exploitent des « ressources inutilisées » (*unused ressources*), qui physiquement appartiennent toujours à leurs utilisateurs. Ed Finn va plus loin, et parle « d'économie de l'interface » (*interface economy*), arguant que l'interface est finalement la seule et unique chose que ce type d'entreprise vend. Nombre d'entre elles se présentent d'ailleurs comme des « emballeurs » (*wrappers*). À travers leurs applications, elles réussissent à « insérer un niveau d'abstraction informatique entre les clients et le chemin traditionnel qu'ils empruntaient pour accéder à différents services. »¹¹ Ce faisant, elles

9. « *Nowadays, enveloping the environment into an ICT-friendly infosphere has started prevailing all aspects of reality and is visible everywhere, on daily basis.* », Luciano FLORIDI, *The Fourth Revolution: How the Infosphere is Reshaping Human Reality*, Oxford: Oxford University Press, 2016

10. Gérard BERRY, *L'Hyperpuissance de l'informatique*, Paris: Odile Jacob, 2017

11. « [...] inserting a computational level of abstraction between consumers and their traditional pathways to services [...] » Ed FINN, *What Algorithms Want*, Cambridge: MIT Press, 2017

modèlent l'information dans des termes calculable, la faisant rentrer dans le cadre des algorithmes.

De la même manière que Facebook mesure l'amitié en like, Uber note les personnes en étoiles. Les transactions sont dissimulées derrière « la magie du logiciel », ramenant l'acte d'achat plus à un jeu qu'un échange de valeurs. Finn explique que ces processus d'abstraction tentent de faire disparaître la notion de travail aux yeux des utilisateurs, et véhiculent ainsi « le rêve capitaliste d'un commerce instantané accessible en un clique, du choix illimité et de la satisfaction immédiate ». ¹² Cette dissimulation délibérée du travail pose de gros problèmes éthiques, et permettent à des entreprises, dont l'exemple type est celui des coursiers à vélo, de nier leur statut d'employeur pour celui de « relieur » (*binder*). Ainsi, ils échappent à la fiscalité et aux obligations d'assurance qui régissent l'emploi traditionnel, aboutissant à une précarisation des conditions des travailleurs dans leur ensemble, considérés comme des indépendants.

« Le cœur de métier de ces nouvelles entreprises est l'exploitation algorithmique des données de mobilité et de consommation des utilisateurs. » ¹³

Mais le travail est bien loin d'avoir disparu. Il s'est au contraire, étendu aux utilisateurs des applications en plus des prestataires de services. Nous allons maintenant nous pencher sur l'envers du *Frame Problem* et la notion de *Digital Labor*, en étudiant les conséquences qu'engendre l'entrée des hommes dans les processus algorithmiques.

12. « *a capitalist [...] fantasy: the frictionless commerce of one-click purchasing, limitless choice and near-instant gratification* » Ed FINN, *What Algorithms Want*, Cambridge : MIT Press, 2017

13. Antonio CASSILLI, Dominique CARDON, *Qu'est-ce que le digital labour ?*, Paris : Éditions de l'INA, 2014

«L'une des conséquences de l'enveloppement croissant du monde en termes compréhensibles par les TIC est que les hommes peuvent accidentellement devenir un rouage du mécanisme. Le problème est simple : parfois no TICs ont besoin de comprendre et d'interpréter le présent, et ont donc besoin de machines à sens comme nous pour accomplir ce travail. C'est une tendance récente, désignée par le terme de calcul informatique humain (*human based computation*)»¹⁴

DIGITAL LABOR

«Par Digital Labor, nous désignons les activités numériques quotidiennes des usagers de plateformes sociale, d'objets connectés ou d'applications mobiles. Néanmoins chaque post, chaque photo, chaque saisie et même chaque connexion à ces dispositifs remplit les conditions évoquées dans la définition : produire de la valeur [...], encadrer la participation [...], mesurer [...].»¹⁵

L'économie numérique est un commerce d'informations. Toute personne fournissant des données apporte de ce fait une valeur aux services de l'application qu'elle utilise. Antonio Cassilli va plus loin, arguant que « nous pourrions identifier l'acte paradigmatique du Digital Labor dans le fait même d'être en ligne » car toute présence en ligne laisse des traces qui sont ensuite monétisables. Il y a là un peu le même problème éthique que le distributeurs de CD d'autrefois qui touchait 5 euros sur un un disque et l'artiste 10 centimes. Celui qui possède le canal de distribution du service, et sait organiser l'information ramasse tout, même si le mérite de toute la valeur produite est le fait de différents acteurs.

14. «*One of the consequences of envelopping the world to transform it into an ICT-friendly place is that humans may become inadvertently part of the mechanism. The point is simple: sometimes our ICTs need to understand and interpret what's happening, so they need semantic engine like us to do the job. This fairly recent trend is known as human based computation.*» Luciano FLORIDI, *The Fourth Revolution: How the Infosphere is Reshaping Human Reality*, Oxford : Oxford University Press, 2016

15. Antonio CASSILLI, Dominique CARDON, *Qu'est-ce que le digital labour ?*, Paris : Éditions de l'INA, 2014

RE-CAPTCHA : COMPLETELY AUTOMATED PUBLIC TURING TEST TO TELL COMPUTERS AND HUMANS APART

Pour certaines tâches, comme la récupération d'un mot de passe ou l'accès à certains services, Google utilise la technologie du re-CAPTCHA (re-Test public de Turing complètement automatique ayant pour but de différencier les humains des ordinateurs), pour « vérifier que l'utilisateur est humain », qu'il ne s'agit pas d'un programme qui récupérerait automatiquement des informations. Concrètement, elle s'incarne par de la reconnaissance textes déformés, que les intelligences artificielles ne peuvent identifier comme telles, dont la reconnaissance certifie votre humanité. « Sans le savoir, toute personne se prêtant à la tâche contribue de fait à la numérisation de textes du services propriétaires Google Books. [...] D'autres modalités "d'authentification par Digital Labor dissimulé" ont été introduites: reconnaître des numéros de rue pour améliorer Google Street View, apparier des images ou "taguer" des contenus pour calibrer les algorithmes de vision numériques qui permettent à Google de faire de la reconnaissance faciale, de la reconnaissance d'objets [...] ou de mouvements. »¹⁶ Les internautes sont utilisés à leur insu comme « traducteurs » pour les algorithmes de Google, et participe ainsi à agrandir l'étendue des données que l'entreprise exploite. Ils implémentent gratuitement la réalité dans la machine, la rendant exploitable. « Les utilisateurs humains effectue une double tâche: il peuvent prouver qu'ils sont humains et aider la machine à numériser un texte qu'elle ne peut déchiffrer automatiquement. »¹⁷

« INTELLIGENCE ARTIFICIELLE ARTIFICIELLE »

Amazon Mechanical Turk (le Turque mécanique d'Amazon) est une « plateforme de micro travail » créé par le géant de la vente en ligne.¹⁸ Le nom provient d'un automate joueur d'échecs de la fin du XVII^e siècle, présentée comme une machine intelligente, qui en fait était un vaste canular. À l'inté-

16. Antonio CASSILLI, Dominique CARDON, *Qu'est-ce que le digital labour ?*, Paris : Éditions de l'INA, 2014

17. « Human-users can now double-task: they can prove that they are human and help to digitalize some machine-unreadable text at one stroke. » Luciano FLORIDI, *The Fourth Revolution: How the Infosphere is Reshaping Human Reality*, Oxford : Oxford University Press, 2016

18. [<https://www.mturk.com>]

rieur du dispositif était caché un être humain qui manipulait le mannequin. La plateforme est accessible à des entreprises, qui proposent à la communauté de travailleurs appelés les *turkers*, des tâches, les *HITs* ou *Human Intelligence Tasks* (Travaux d'Intelligence Humaine), que les ordinateurs sont incapables d'effectuer, mais qui néanmoins sont extrêmement répétitives. Il s'agit de taguer des images, reconnaître des visages, bref le même type de tâches que proposent les re-CAPTCHAS, cette fois rémunérées quelques centimes. De la même manière, les êtres humains jouent le rôle de « machines à sens ». D'où le sous-titre ironique *Artificial Artificial Intelligence* (intelligence artificielle artificielle). Ces micro-tâches, répétitives et non-qualifiées sont ce qu'on pourrait appeler le travail à la chaîne du XXI^e siècle.

CALCULATEURS MODERNES

En anglais, ordinateur se traduit par *computer*, littéralement « calculateur ». Il fut un temps, il n'y a pas si longtemps, avant l'invention des ordinateurs dans les années 1940, où *computer* ne désignait pas une machine mais bien une personne. Une personne dont le travail consistait à effectuer des calculs, pour répondre à des besoins scientifiques, économiques ou militaires. Ce travail était également dissimulé, effectué par des anonymes, le plus souvent des femmes, utilisées comme des machines. S'il assume des dizaines de fonctions de nos jours, l'ordinateur n'est au départ rien de plus qu'une machine à calculer. L'ENIAC, un des premiers ordinateurs modernes fut inventé aux États-Unis pour effectuer des calculs de tirs balistiques pendant la seconde guerre mondiale, car les personnes en charge, étaient « trop lentes » par rapport aux besoins du front. Les *turkers* d'Amazon sont nos *modern computers*, invisibles entités qui nourrissent les machines à calculer du XXI^e siècle : les algorithmes.

TRAVAIL, ÉTHIQUE

Micro-tâches, optimisation du travail, flexibilité du travail...toutes ces mutations résultent de l'adaptation de notre ancien système à la nouvelle dimension informatique du monde, tout en conservant le même cadre idéologique. Antonio Cassili signale que « les marchés du travail numérique perturbent la notion même de travail aussi parce-que le temps dédié à l'activité travaillée déborde et infiltre le temps de vie » et ajoute que « cette hybridation des rôles de production et de consommation semble s'inscrire dans

un dynamique sociale et économique plus vaste, qui dépasse la seule économie du numérique. »¹⁹ Il devient impossible de considérer le travail dans les mêmes termes qu'avant, et avec sa raréfaction ou sa dilution dans la vie, l'équation travail = droit de subsistance perd complètement son sens. L'organisation de nos sociétés se doit d'épouser ses changements.

PONTS

Dans son livre « *Code, the Hidden Language of Computer Hardware and Software* » (*Code, le langage caché du matériel informatique et des logiciels*), le programmeur Charles Petzold explique que « le système binaire crée un pont entre l'arithmétique et l'électricité ». ²⁰ Le pont entre la culture et l'informatique n'est pas si évident. C'est nous tous qui le construisons chaque jour, en traduisant nos vies en termes compréhensibles par les algorithmes. Et comme le soulignent Luciano Floridi et Antoinette Rouvroy, « le risque que nous courrons est qu'en enveloppant le monde, nos technologies pourraient modéliser nos environnements physique et conceptuel et nous contraindre à nous ajuster à eux, parce-que c'est mieux, plus facile, ou parfois, le seul moyen de les faire fonctionner. » ²¹ En d'autres termes, de sacrifier certains aspects de notre humanité à l'efficacité de nos dispositifs.

19. Antonio CASSILLI, Dominique CARDON, *Qu'est-ce que le digital labour ?*, Paris : Éditions de l'INA, 2014

20. « *The binary number system bridges the gap between arithmetic and electricity.* » Charles PETZOLD, *Code, the hidden language of hardware and software*, Microsoft Press, 2000

21. « *The risk we are running is that, by envelopping the world, our technologies might shape our physical and conceptual environments and constrain us to adjust to them because that is best or easiest, or indeed sometimes, the only way to make things works.* » , Luciano FLORIDI, *The Fourth Revolution: How the Infosphere is Reshaping Human Reality*, Oxford : Oxford University Press, 2016

Monétiser l'espace

IV

Recherches monétisées

1 IV

Il serait inconscient de ne pas profiter de toutes ces brebis égarées sur le net, en quête désespérée d'informations. Elles, nous, représentons ce qu'on pourrait appelé *a dream audience*, un public de rêve, pour qui saurait nous guider. Le coup de maître de Google a été d'instaurer une taxe sur sa carte algorithmique du web. Et pour ce faire, il a créé *AdSense*, un système qui évalue la valeur des mots-clés, pour ensuite les revendre à des annonceurs.

«D'un point de vue commercial, PageRank crée un index basique pour la circulation des idées, une monnaie essentielle dans l'économie de l'attention. Quand Google a commencé à vendre des publicités en échange de ses résultats de recherche en utilisant le système *AdSense*, il a réussi à monétiser cette attention à un niveau jusqu'alors inimaginable.»¹

Lorsque vous effectuez une recherche, une série de liens commerciaux apparaissent avant les résultats de recherche, ce qui leur donne une visibilité maximum. C'est ce que vend *AdSense*. Un passe-droit sur l'algorithme PageRank. Outre la possibilité d'acheter une place dans le référencement, il vous

1. «*From a business perspective, PageRank creates a basic index for the circulation of ideas, an essential currency in the economy of attention. When Google began selling advertisements against its search results with the market binding system AdSense, it succeeded in monetizing that attention at a previously unimaginable scale.*» Ed FINN, *What Algorithms Want*, Cambridge: MIT Press, 2017

permet également de placer sur votre site des publicités sponsorisées par l'entreprise. À chaque fois que quelqu'un clique sur l'une d'elles, Google vous rémunère. Fondamentalement, les rangs à vendre s'opposent à la prétention de classement méritocratique de Google. Bien que la nature publicitaire des liens soit spécifiée, ils obtiennent une place extrêmement privilégiée, qui rend la confusion facile. Combien de fois avez-vous cliqué sur le premier lien avant de réaliser qu'il s'agissait d'un marchand de chaussures ? Google a presque réalisé le vieux rêve des publicitaires : faire passer la pub pour de l'information.

RÉMUNÉRATION

« Google a réussi à étendre le domaine du capitalisme à la langue elle-même, à faire des mots une marchandise, à fonder un modèle commercial incroyablement profitable sur la spéculation linguistique. »²

À chaque clic sur un lien sponsorisé par Google, les entreprises sont rémunérées en conséquences, selon une somme calculée en temps réel par un autre algorithme. Il fonctionne d'après 3 paramètres, qui incluent l'enchère sur un mot-clé (le prix maximum que l'entreprise serait prête à payer à Google pour une redirection) et le calcul du score de qualité de la publicité (laissé libre à Google) ainsi que le calcul de son rang, qui définit l'ordre d'apparition. Une grande formule, qui évalue la valeur économique de la langue, monétisée, quantifiée d'après un système d'enchère permanentes.

« Quand Google prolonge une phrase que vous avez commencé à taper dans la case de recherche, il ne se borne pas à vous faire gagner du temps : il vous ramène dans le domaine de la langue qu'il exploite. »²

Un dernier clou dans le cercueil de la neutralité de PageRank, qui quoique qu'il arrive nous guide implicitement vers les plus offrants. Nous sommes constamment pris par la main, invités à parcourir des sentiers pré-définis de l'algorithme, qui correspondent « au domaine que [Google] exploite ». Littéralement.

2. Frédéric KAPLAN, *Vers un capitalisme linguistique – Quand les mots valent de l'or*, *Le Monde diplomatique*, 1.11.2011

TUNNELS

Un autre aspect préoccupant des algorithmes est leur proportion à différencier ces chemins d'une personne à l'autre, en analysant les traces que nous laissons. De celles-ci, ils tirent d'hypothétiques préférences basées sur notre passé pour nous vendre un futur auquel nous aspirerions. C'est ce qui se passe à petite échelle, lorsque vos précédentes requêtes vous sont proposées en priorité quand vous commencez à taper un terme dans un moteur de recherche. Même chose avec l'algorithme de complétion automatique inséré dans les claviers de smartphones. Lorsque vous débutez un texte, il vous fait des propositions et tente de prédire le mot que vous allez écrire, en utilisant conjointement les *Word Embeddings*, et vos précédentes rédactions. À force d'utilisation, chacun finit par avoir des propositions complètement différentes, calquées sur ses habitudes. De village global, le web mute en un ensemble de tunnels, de plus en plus cloisonnés, entre lesquels un monde commun peine à exister.

Le NewsFeed de Facebook

« Tout ce qui vous plaira. »

2

IV

CE QU'EST FACEBOOK

Facebook, le réseau social lancé par Mark Zuckerberg en 2004, permet aux utilisateurs de se constituer un cercle d'amis en ligne. Chacun se crée une page à son nom, qu'il partage ensuite avec d'autres selon le principe de la « demande d'ami ». Facebook vous permet de publier facilement du contenu tel que des photos, du texte, de créer des événements ou encore de partager des liens. Ces publications sont seulement accessibles à votre cercle, et inversement. La plateforme comprend également un service de messagerie instantanée. Le NewsFeed constitue l'élément central du réseau. En français *fil d'actualités*, il fait office de page d'accueil personnalisée, et rassemble les publications de vos amis ou des pages auxquelles vous vous êtes abonné (des pages que vous aimez, *likez*). À ces « actualités », sont mêlées des publicités, disposées tout au long du fil, qui financent l'existence du réseau.

INFORMATION

Au delà d'un réseau social Facebook est devenu un moyen d'information. À travers les articles de journaux partagés sur leur NewsFeed ou des opinions partagées par leur cercle, il donne aux utilisateurs une idée du monde, de la situation politique notamment. Selon une étude de 2015 du centre de recherche américain Pew Research, 61% des américains nés après l'an 2000 utilisent Facebook comme première source d'informations.³ La réalité, en tant qu'expérience d'un monde commun, se retrouve ainsi segmentée par affinités. Le réseau fonctionnant sur le principe du *like* et de la « demande d'ami », les utilisateurs n'ont pas accès à des avis opposés au leur, et vivent cloisonnés dans un espace de non-débat, d'entre-soi. Cette tendance fut soulignée en 2011 par le militant Eli Pariser, qui la baptisa *Filter Bubble* (bulle de filtres). La philosophe Antoinette Rouvroy apparente ce phénomène à une extension de l'espace privé au détriment du commun. Chacun son monde. Et l'algorithme EdgeRank qui « classe » (*rank*) votre fil d'actualités, semble le renforcer.

CEILLÈRES

Cet algorithme a été conçu pour choisir les publications visibles par chacun, ainsi que l'ordre dans lequel elles apparaissent. La raison d'un tel tri étant le nombre trop important de partages : une personne peut avoir des milliers « d'amis », ou suivre des centaines de pages. Montrer chronologiquement l'ensemble de leurs publications la noierait sous une montagne d'informations, pour la plupart insignifiantes à ses yeux. EdgeRank classe donc les posts d'après l'intérêt hypothétique que chacun leur porterait.

QUI QUE J'AI ÉTÉ, CE N'EST PAS QUI JE SUIS

Cet intérêt est quantifié suivant les activités passées de l'utilisateur sur le réseau. Plus vous avez interagit avec certaines pages ou personnes, plus vous êtes susceptible de voir leurs actualités. Concrètement, l'action d'EdgeRank implique que si je publie un contenu, je ne sais pas s'il apparaîtra

3. MITCHELL Amy, GOTTFRIED Jeffrey, KATERINA Eva Matsa, *Millennials and Political News: Social Media – the Local TV for the Next Generation?*, *Le Monde Diplomatique*, 1.6.2015

sur le fil d'actualité de mes amis : je n'ai aucune idée de comment et par qui il sera reçu. Lors de l'élection de Donald Trump à la présidence des États-Unis le 8 novembre 2016, de nombreux démocrates ont été abasourdis. Enfermés dans leurs bulles Clinton, ils sont restés aveugles à la montée en puissance de Trump, convaincus de leur victoire. De même, les républicains furent exemptés de toute actualité démocrate. Chacun est resté complètement imperméable à l'autre, convaincu que le monde entier était de son avis. Chacun dans sa « caisse de résonance » (*echo chamber*).⁴ Le réseau limite l'espace de débat. Tout comme la possibilité de changer d'avis.

Il est impossible de blâmer entièrement Facebook quant aux phénomènes des filtres. Ces bulles n'ont pas eu besoin du numérique pour exister. Notre cercle social, pour la plupart, se compose de personnes avec lesquelles nous partageons une vision du monde, des idéaux, et qui évoluent dans un milieu similaire au notre. C'est nous qui ajoutons / acceptons nos amis en ligne, nous en sommes donc partiellement responsables. Ce qui est inadmissible, c'est le caractère arbitraire d'EdgeRank. Cette machine sélectionne ce que nous voyons, d'après une recette tenue secrète. Cela signifie que même si je tente de diversifier mon apport idéologique de par les pages que je choisis, ou les personnes que j'admets dans mon cercle, je ne verrais probablement pas leurs publications. Nous sommes responsables de nos filtres, mais l'algorithme en est un de plus sur lequel nous n'avons aucun contrôle.

« C'est ce qui arrive quand un réseau immensément puissant, que nous partageons avec 1.5 milliards d'utilisateurs se vend au plus offrant. »⁵

PUB MON AMOUR

Facebook est, comme beaucoup d'entreprises, financé par la publicité. Plus les utilisateurs passent de temps sur la plateforme, plus sa valeur grimpe aux yeux des annonceurs. C'est le seul but qu'EdgeRank poursuit : les garder

4. Mostafa M. EL-BERMAWY, *Your Echo Chamber is Destroying Democracy*, *Wired US*, 18.11.16

5. « *This is what happens when the immensely powerful network we share with 1.5 billion users is also publicly traded corporation.* » Cathy O'NEIL, *Weapons of math destruction: how big data increases inequalities and threatens democracy*, Crown : New York, 2016

le plus longtemps connectés. En (ne) leur montrant (que) ce qu'ils ont (supposément) envie de voir.

Clickbait

3
IV

Wikipédia définit le *clickbait* comme « un terme péjoratif s'appliquant à du contenu web dont le but principal est de pousser les utilisateurs à cliquer sur un lien pour aller sur une certaine page »⁶ (en français « pute à cliques »). Il désigne les articles qui vous aguichent avec des phrases racoleuses, type « **VOUS N'AURIEZ JAMAIS CRU QU'UN DAUPHIN PUISSE FAIRE ÇA** » ou « **IL DEVIENT MILLIONNAIRE EN DORMANT!!! DÉCOUVREZ SON SECRET** ». Le *clickbait* se pratique sur les réseaux sociaux mais aussi partout ailleurs sur le web. Le seul but de ces entêtes est de vous pousser au clic, pour générer des bénéfices, selon l'équation universelle : clic = visiteurs = publicité = argent. Cette pratique est symptomatique de la culture du proxy. La métrique devient une fin en soi, un moyen de générer des revenus et perd du même coup, toute valeur de mesure.

JOURNAUX SOCIAUX

Le *clickbait* est symbolique de l'influence des algorithmes sur les médias. L'accès à la presse passe désormais de manière croissante par les réseaux sociaux, que l'on nomme maintenant les *gatekeepers*, gardiens. Les rédactions ont donc dû s'adapter à eux, notamment au format que le NewsFeed utilise pour diffuser leurs articles. Et du même coup, se sont mis à racoler les lecteurs. « Les plateformes algorithmiques modèlent désormais effectivement toute la production culturelle, des auteurs forcés au badinage sur Twitter pour promouvoir leurs livres, aux systèmes sophistiqués qui nous recommandent de nouveaux produits. »⁷ La qualité d'un article importe certes toujours, n'exagérons rien, mais encore faut-il qu'il soit lu. Et avant la lecture,

6. « *Clickbait is a pejorative term for web content whose main goal is to get users to click on a link to go to a certain webpage.* » *Clickbait*, [en.wikipedia.org/wiki/Clickbait]

7. « *Algorithmic platforms now shape effectively all cultural production, from authors engaging in obligatory twitter badinage to promote their new books to the sophisticated systems recommending new products to us.* » Ed FINN, *What Algorithms Want*, Cambridge : MIT Press, 2017

il y a le clic. Son succès est donc grandement dépendant de son « emballage algorithmique ». Une nouvelle fois, il faut apprendre à jouer selon les règles de la machine. Une nouvelle fois, se profile l'ombre du *Frame Problem*.

FAKE NEWS

« Les algorithmes de Facebook favorisent le côté viral. Ils sont entraînés pour montrer aux gens le genre d'actualités qui vont leur plaire, et permettent ainsi à la plupart des *Fake News* de passer inaperçues. »⁸

Sont appelées *Fake News* les fausses nouvelles qui circulent sur les réseaux, type « **OBAMA N'EST PAS NÉ SUR LE SOL AMÉRICAIN** » ou « **LES CLINIQUES D'AVORTEMENT REVENDENT DES FŒTUS À LA SCIENCE** ». Ce problème devient endémique et est extrêmement préoccupant, plus encore en période électorale. On parle de « post-vérité », qui est un terme que je trouve personnellement absolument détestable. La « post vérité » désigne l'idée d'un discours sans liens avec le réel matériel, en opposition avec la « vérité », qui est définie comme adéquation du discours au réel. La « post-vérité » considère les événements passés comme relatifs alors qu'ils sont absolus. Elle fait rentrer les faits dans la relativité de l'interprétation, qui n'est pas en elle-même condamnable et reste essentielle dans le travail de la presse. Si les journalistes donnent sens aux faits, chacun de manières différentes selon leurs sensibilités politiques, leurs connaissances, leur discernement (ce qui regroupe le terme « interprétation »), ils sont tous soumis à la même réalité matérielle, indifféremment de leurs opinions. À mes yeux, le combat contre la « post-vérité » ne consiste pas à chercher la neutralité, c'est-à-dire à dépouiller les faits de toute signification mais simplement, à les respecter. L'interprétation est un exercice de pensée, de fondation du futur, au cœur du travail journalistique, qui n'a pas à être remis en question en lui-même, tant qu'il se base sur le réel. Mais peut-on éviter les *Fake News* dans un système global basé la seule popularité ?

8. « Facebook's algorithms prize virality. They're trained to show people the kind of news they're going to like, allowing so much of this *Fake News* to go unchallenged. » Cathy O'NEIL, *Weapons of math destruction: how big data increases inequalities and threatens democracy*, Crown : New York, 2016

COMMENT DES ADOS MACÉDONIENS SE FONT DE L'AGENT
DE POCHE SUR LA PROPENSION DES GENS À CROIRE N'IMPORTE
QUOI ET POURQUOI CLAIREMENT ON NE PEUT PAS
TANT LEUR EN VOULOIR

«Le mensonge est souvent plus plausible, plus tentant pour la raison que la réalité, car le mensonge possède le grand avantage de savoir à l'avance ce que le public souhaite entendre ou s'attend à entendre.»⁹

La petite ville de Veles, située en Macédoine, a fait l'objet de beaucoup d'attention pendant les présidentielles américaines. Il se trouve que plus de 150 sites de *Fake News* pro-Trump,¹⁰ furent montés par des personnes y étant domiciliées. Ces pages diffusaient des articles aux titres évocateurs tel que «**LE PAPE DONNE SA BÉNÉDICTION À TRUMP**» ou «**VOILÀ COMMENT LES LIBÉRAUX ONT DÉTRUITS L'AMÉRIQUE**». Une journaliste irlandaise, Samanth Subramanian, dresse pour *Wired* le portrait de Boris (pseudonyme), un adolescent de la ville qui s'était lancé dans la propagande d'ultra droite le temps de la campagne électorale.¹¹

L'histoire commence en février 2016. Sur son site *Daily Interesting Things*, le jeune homme publie la copie d'un article trouvé sur le net, qui raconte comment Donald Trump aurait giflé un homme en désaccord avec lui lors d'un gala en Caroline du Nord. Ce qui n'est, précision, jamais arrivé. Il partage ensuite le lien de sa page sur des groupes Facebook dédiés à la politique américaine. À son grand étonnement, le lien est re-partagé plus de 800 fois. Ce mois-là, Boris gagne \$150 dollars grâce aux publicités sponsorisées par Google présentes sur son site. Tout bien considéré, il lâche l'école se lance à temps plein dans le *Fake News* business.

«Google a construit un modèle algorithmique qui formate aujourd'hui l'ensemble du monde digital.»¹²

9. Hannah ARENDT, *Du mensonge à la violence*, Calman-Lévy : Paris, 1972

10. Dan TYNAN, *How Facebook powers money machines for obscure political 'news' sites*, *The Guardian*, 24.8.2016

11. SUBRAMANIAN Samanth, *The Macedonians Teens who Mastered Fake News*, *Wired*, 15.2.2017

12. «Google has built an algorithmic model that now now shapes the whole digital world.» Ed FINN, *What Algorithms Want*, Cambridge : MIT Press, 2017

Subramanian explique qu'entre février et novembre 2016, il aurait gagné environ \$16.000 (le salaire mensuel moyen en Macédoine étant de \$374) simplement grâce aux GoogleAds de ses pages. Et Boris n'était pas tout seul : de nombreux copains à lui se sont aussi lancés dans l'aventure lucrative du mensonge politique sponsorisé par la Silicon Valley. « Ce que Veles a produit était [...] une manufacture de pure immoralité, non seulement sans aucune idéologie mais sans aucun intérêt, ni aucune empathie envers les enjeux de cette élection. Ces gars sur Facebook s'en foutaient que Trump gagne ou perde la maison blanche. Il voulait juste de l'argent de poche pour s'acheter des trucs — une voiture, des montres, de meilleurs téléphones, des coups au bar. »¹³ Une politique basée entièrement sur la popularité, où les gagnants sont ceux qui amassent le plus de clics aboutit inexorablement tôt ou tard au mensonge. Et le fait que ce soit des ados au fin fond de la Macédoine post-industrielle qui booste Trump et profiter de la *click economy* (économie du clic) et s'acheter des iPhones ne manque pour sûr, pas d'ironie.

MEA CULPA

« Tous les systèmes [algorithmiques] peuvent être manipulés. »¹⁴

Suite aux différents scandales autour des Fake News, Facebook a décidé de prendre des mesures et de travailler à la modification de son algorithme, pour qu'il soit capable de détecter automatiquement les mensonges. « Il se trouve que construire un algorithme qui trie les articles d'après leur pertinence et leur fond est extrêmement compliqué — peut-être même impossible. Ce qui fait que malgré tous les efforts de l'entreprise, les éditeurs continuent de se jouer de lui, en utilisant des formes de *clickbait* toujours plus subtiles et originales [...]. »¹⁵ L'un des dangers inhérent aux algorithmes comme curateurs de l'espace culturel est très bien illustré par les phénomènes de *clickbait*

13. « *What Veles produced, though, was something more extreme still: an enterprise of cool, pure amorality, free not only of ideology but of any concern or feeling about the substance of the election. These Macedonians on Facebook didn't care if Trump won or lost the White House. They only wanted pocket money to pay for things—a car, watches, better cell phones, more drinks at the bar.* » SUBRAMANIAN Samantha, The Macedonians Teens who Mastered Fake News, *Wired*, 15.2.2017

14. « *All [algorithmic] systems can be gamed.* » Cathy O'NEIL, *Weapons of math destruction: how big data increases inequalities and threatens democracy*, Crown : New York, 2016

et de *Fake News*. Il consiste en l'élévation du processus au rang de valeur et de ce fait, de son possible détournement. En l'occurrence, pour faire de l'argent. L'exemple du hack des résultats d'image Google évoqué précédemment est également symptomatique de ce problème. Excepté que dans ce cas là, le détournement poursuivait non pas un but économique mais politique. Quels que soient les intérêts à l'œuvre, il reste dans la nature de l'algorithme analyste des hommes d'être manipulable.

Pour endiguer ces torrents de mensonges, le réseau à également mis en place un système de validation des organes de presse, en accordant une « pastille bleue » à certaines sources. Grâce à elle, les articles publiés par ces journaux voyagent plus facilement dans les fils d'actualité que sans, sont favorisés. « Les rôles du curateur, de l'éditeur, et du critique sont plus importants que jamais comme nous traçons les limites de l'informatisation et luttons pour apprendre et nous souvenir des choses que nous systèmes informatiques ne font ou ne peuvent pas faire. »¹⁶ Cette curation « à la main » par Facebook soulève naturellement de nouveaux problèmes, comme son pouvoir de censure. Avons-nous vraiment envie de Facebook comme éditeur de la presse mondiale ? (rires) Mais c'est un autre débat. C'est d'ailleurs un très vieux débat, qui concernait aussi les éditeurs papier, et qui met en jeu les mêmes conflits d'intérêts. Ce qui plus est intéressant, c'est que cette sélection par un comité représente un pas en arrière de l'informatisation. Nous sommes revenus à une validation d'un contenu culturel par des hommes de chair et d'os. Avec lesquels nous pouvons être en désaccord, qui possèdent un pouvoir complètement disproportionné, certes. Mais hommes quand même. La campagne présidentielle américaine aura au moins eu un point positif : montrer à la face du monde que vérité et *click economy* sont incompatibles.

15. « *It just happens that building an algorithm to rank stories according to things like accuracy and substance is extremely hard—perhaps impossible. And so, despite the company's best efforts, publishers are still gaming it with ever more subtle and inventive forms of click-bait [...].* » Will OREMUS, *Facebook Is Telling News Outlets How to Do Their Job*, *Slate*, 24.10.2017

16. « The role of the curator, the editor, and the critic is more important than ever, as we draw the lines of effective computability and struggle to learn and remember the things that our computational systems do not or cannot know. » Ed FINN, *What Algorithms Want*, Cambridge : MIT Press, 2017

NEW NEWSFEED

Au début du mois de Novembre 2017, Facebook a testé une modification de son fil d'actualité dans 6 pays dont le Sri-Lanka, la Slovaquie et la Serbie. Ce changement incarnerait une réponse partielle aux *Fake News* et également à une plainte récurrente des utilisateurs qui trouvaient la vie sociale de leurs amis enterrée sous la presse. Au lieu d'avoir comme d'ordinaire, un seul fil, certaines pages ont été reléguées dans un onglet supplémentaire, un clic plus loin, appelé *Explore Feed*. Stefan Dojcinovic est un journaliste indépendant Serbe, qui publie un journal d'investigation en ligne, *KRIK*. Comme il l'explique dans un article au *New York Times*, la Serbie n'est pas une « démocratie complètement fonctionnelle »¹⁷ (rires), ce qui implique une censure de la presse par l'État, sous une forme « douce » (*soft*). Concrètement, l'argent public finance seulement les organes d'État, et les indépendants sont victimes de toutes sortes de déstabilisations, qui vont de la calomnie dans les tabloïds aux contrôles incessants du fisc. Son opposition lui interdit les moyens de diffusion classiques comme l'imprimé ou la télévision.

Facebook est le réseau social n° 1 en Serbie. Le seul point positif que l'on lui peut reconnaître est son indépendance vis à vis des pouvoirs d'États, qui permet à la presse de se faire entendre sous certains régimes autoritaires. Pour Dojcinovic, banni des circuits classiques, le réseau était un bon moyen de visibilité et accessoirement la principale source de redirection vers son journal. De son point de vue, la modification du NewsFeed a été terrible, faisant dramatiquement chuté le taux de fréquentation de son site. Dans cette version du NewsFeed, seules les « pages sponsorisées » (les pages qui payent) ont le droit d'apparaître dans le fil d'actualité général, et les autres pages étant reléguées sous l'onglet. En web, un clic est un monde. En appliquant de telles méthodes, Facebook reconduit les inégalités de la matérialité, en ostracisant la presse à faibles moyens. Parfois, les *Fake News* viennent de grands journaux. La vérité n'est *oh surprise*, pas non plus proportionnelle aux moyens du journal. Comme nous l'avons dit plus haut les faits ne se laissent pas cerner par un chiffre, un proxy quelconque. Leur définition sort du cadre de quantification stricte.

17. *Serbia escaped the dictatorship of Slobodan Milosevic in 2000, but it hasn't developed into a fully functioning democracy.* Stevan DOJCINOVIC, [Hey, Mark Zuckerberg: My Democracy Isn't Your Laboratory](#), *New York Times*, 15.11.2017

Nous pouvons nous douter que le but d'une telle opération est plutôt de pousser les journaux à systématiquement payer pour de la visibilité. Quoiqu'il en soit, Facebook a avorté le test qui fut un échec retentissant. Il faut souligner que ce test n'a pas été fait aux États-Unis mais dans de petits pays éloignés, au faible pouvoir international, dont la voix porte peu jusqu'à l'Amérique. Ce choix démontre un cynisme effrayant de la part de la plateforme, qui n'a aucune honte à prendre les utilisateurs de pays défavorisés pour des rats de laboratoires. Comme le dit Dojcinovic: «Hey Mark Zuckerberg, ma démocratie n'est pas ton laboratoire.»¹⁸ Cet événement met également en lumière l'importance du design d'interface dans l'âge de l'information. La relégation d'un contenu derrière un simple clic peut avoir des conséquences dramatiques. Les questions de hiérarchie sont centrales dans le débat autour des algorithmes qui ne sont rien de plus que ça: des machines à tri. Le rôle éthique du designer, comme curateur de l'espace informatif, devient incontournable.

PRESSE RE-INDÉPENDANTE

Cette dépendance des médias à Facebook est assez préoccupante, car il peuvent du jour au lendemain faire voler des journaux en éclats. Certains choisissent d'exister seulement par eux, sans même avoir de site web externe. Mais cette dépendance n'est que la conséquence d'une première dépendance: celle des utilisateurs. Il faut également interroger nos usages. Devant de tels agissements, il devient difficile de continuer à utiliser les services de Facebook, d'un point de vue éthique, ou simplement si en tant que citoyen, nous désirons un accès à une information fiable.

ON NE CROIT PAS AUX ÉVÉNEMENTS

Ils existent. Ou n'existent pas. Quelles que soient nos convictions. Je pense que les lecteurs de *Fake News* ont une part de responsabilité non négligeable dans leur prolifération, tout en étant victimes. L'attitude qui consiste à tenir pour vrai seulement ce qui nous plaît est extrêmement dangereuse, et de plus profondément égocentrique. La réalité n'obéit pas à notre bon plaisir. Tout autant que les rédacteurs de fausses nouvelles, cette attitude menace

18. «Hey, Mark Zuckerberg: My Democracy Isn't Your Laboratory.» Stevan DOJCINOVIC, Hey, Mark Zuckerberg: My Democracy Isn't Your Laboratory, *New York Times*, 15.11.2017

l'existence d'un monde commun, fondé sur une réalité palpable par tous, qui finit quoiqu'il arrive toujours par s'imposer. Toutes les négations à dessein du réel se soldent par des échecs, qu'elles soient collectives, à un niveau étatique ou individuelles, au niveau de nos bulles personnelles. Il faut rester constamment vigilant face à soi-même, et ne jamais prendre « ses désirs pour des réalités ».

La solution ne peut être entièrement technique même si une part du problème l'est. S'informer s'apprend, et les FakeNews sont aussi le symbole d'une fracture sociale et économique, qui se joue dès l'enfance avec la qualité de l'éducation à laquelle chacun peut avoir accès. S'informer prend un temps, qu'il est impossible de compresser par des processus automatiques. Tout comme les algorithmes ne peuvent réduire à l'infini la place occupée par les données, il est un seuil de temps en-dessous duquel la distinction entre mensonge et vérité devient problématique. Ce temps est également le reflet d'inégalités : tout le monde n'y a pas accès et il est clair que combiner deux jobs en élevant ses enfants ne laisse pas énormément de place à autre chose que des problèmes matériels. La crédulité est donc également, un symptôme d'injustices sociales. Elle représente un terrain exploitable, et fragilise les populations quant à la manipulation des puissants. Ce qui me gêne particulièrement dans l'algorithme EdgeRank, c'est qu'il se nourrit de cette propension des utilisateurs à croire ce qui les arrange, tout simplement car elle est plus lucrative. C'est un peu « le client est roi, si monsieur veut une réalité alternative, il n'a qu'à demander ». L'éthique de Facebook est basée sur le profit et seulement lui, ce qui pose à terme, pas mal de problèmes à nos démocraties.

Monétiser

V

les utilisateurs

« Si c'est gratuit, c'est vous le produit. »

Adage de la Silicon Valley

« — Aoru, Jenkins, Pendragon... Combien de noms as-tu en fin de compte ?!

— Autant qu'il en faut pour vivre libre. »

Le Château Ambulant, Hayao MIYAZAKI, 2006

Données personnelles

1
V

TRACER

« Nous avons tous fait l'expérience d'être suivis sur le web par une pub tirée de quelque chose que nous avons cherché ou lu. Vous savez, quand regardez une paire de bottes et qu'après, pendant une semaine, ces bottes vous suivent partout. »¹ Que des bottes vous tracent sur Internet n'est pas vraiment le problème. Le problème vient des données collectées pour que ces bottes puissent vous tracer. Comme le dit la sociologue Zeynep Tufekci, « Nous construisons une dystopie juste pour pousser les gens à cliquer sur des pubs. »² Et il faut que ça s'arrête.

1. « We all had the experience on the web of being followed by an multiple based on something we searched or read. You know, you look at a pair of boots and, for a week, those boots are following you around everywhere you go. »

2. « We're building a dystopia just to make people click on ads. » Zeynep TUFEKCI, *We're building a dystopia just to make people click on ads*, TEDGlobal: New York US, 9.2017

DATA BROKERS

Les *data brokers* sont des entreprises qui collectent les données personnelles que nous disséminons sur le net pour les revendre. Historiques de recherches, profils de réseaux sociaux, listes de sites consultés, listes d'achats, géolocalisation etc... Les moindres traces sont agrégées, puis recoupées entre elles pour profiler au mieux les individus. Cette valorisation, qui transforme les données en informations, qui les « fait parler » en les recoupant s'appelle le *data mining*. Plus le profil est ciblé, plus la donnée se vend chère. Le marché est estimé à 54 milliards d'euros en Europe.³ Cette monétisation des traces entre dans le cadre du *Digital Labor* que nous avons évoqué. Les utilisateurs sont eux-même les producteurs des données qui sont ensuite revendues par les plateformes. Comme le dit Antonio Cassilli, l'adage « si c'est gratuit, c'est vous le produit » signifie en fait « si c'est gratuit, c'est que tu y travailles ».⁴ L'utilisateur est à la fois consommateur d'un service et producteur de richesse.

Lors de l'élection de Donald Trump en novembre 2016, une entreprise britannique, *Cambridge Analytica* s'est retrouvée en première ligne dans la presse. Il s'agit d'une société d'influence numérique, qui prétend pouvoir changer le vote des citoyens à la demande. À partir des données personnelles rachetées aux *data brokers*, elle est capable de dresser un profil psychologique des individus, pour cibler les indécis et les amener à changer de camp. Un *data miner* professionnel.⁵

« Nous avons une base de donnée très vaste concernant tous les adultes aux États-Unis. Nous détenons des milliers d'informations sur chaque personne : consommation, style de vie, démographie, géographie, transactions bancaires, réseau sociaux... »⁶

3. John Paul LEPERS, *Data brokers : la grande manipulation ?*, *Vox Pop*, Arte, 12.1.2017

4. Antonio CASSILLI, Dominique CARDON, *Qu'est-ce que le digital labour ?*, Paris : Éditions de l'INA, 2014

5. Pour comprendre comment marche le *data mining* et quel pouvoir il confère, je conseille vivement la conférence *SpiegelMining - ReverseEngineering von Spiegel Online* (*SpiegelMining - Inverser la machine du Spiegel en ligne*) qui eu lieu lors du 33^e Chaos Computer Congress. C'est une excellente démonstration, en direct, de la mise en valeur de données par l'informaticien allemand David Kriesel.

6. Alexander NIX, PDG de *Cambridge Analytica* propos recueillis dans l'émission de John

Ce ciblage leur permet de mettre en place des campagnes de publicités sur mesure. *Cambridge Analytica* classe ainsi les électeurs selon l'intérêt qu'ils portent à des thèmes spécifiques, comme les salaires, la sécurité ou l'immigration pour ensuite « délivrer le message qui saura capter [leur] attention »⁷. Cette pratique se nomme le *micro targeting* (micro-ciblage).

Le rôle de Facebook

2
V

DONNÉES ENTRANTES

Parallèlement à sa mission de « rendre le monde meilleur en connectant les gens »,⁸ Facebook revend les données personnelles de ses utilisateurs à des acteurs extérieurs, comme les *data brokers*. La journaliste Carole Cadwalldr, dans son article pour *The Guardian*, [The great British Brexit robbery: how our democracy was hijacked](#), (Le grand braquage britannique : comment notre démocratie a été piratée) explique que ce sont eux qui fournirent la majeure part des données utilisées par *Cambridge Analytica* pour profiler les électeurs. Mais leur contribution est allée au-delà. Comme elle le raconte, des psychologues de l'université de Cambridge avaient précédemment récolté légalement des données sur le réseau, dans un but académique, qu'ils analysèrent. Ils en tirèrent un papier, expliquant comment les utilisateurs trahissent leur personnalité, leur orientation sexuelle, leur opinions politiques par de simples *like*. L'un des scientifiques de l'équipe fut ensuite recruté par le *data miner*, pour trouver de nouvelles stratégies. Il mit au point un test de personnalité, que les utilisateurs de Facebook pouvaient faire en ligne, qui cette fois était rémunéré. Celui-ci autorisait non seulement la récolte de leur profil mais aussi de ceux de leurs amis. Pratique naturellement autorisée par le réseau. Vendre vos dents ou vos cheveux, c'est ringard. En 2017, on préférera vous acheter vos faiblesses intimes, vos vulnérabilités psychologiques pour ensuite vous vendre un mur entre le Mexique et les USA, en tapant pile là où ça fait mal.

Paul LEPERS, *Vox Pop*, [Data brokers : la grande manipulation ?](#), Arte, 12.1.2017

7. Giovanni DONI, Employé de *Cambridge Analytica* propos recueillis dans l'émission de John Paul LEPERS, *Vox Pop*, [Data brokers : la grande manipulation ?](#), Arte, 12.1.2017

8. « *Make the world a better place by connecting people* » Slogan de Facebook

PSYCHOLOGIE

Pendant les élections américaines de 2010 et 2012, les chercheurs de Facebook ont conduit une expérience intitulée *The Voters Megaphone* (Le porte-voix des électeurs), pour encourager les utilisateurs à aller voter. Ceux-ci pouvaient cliquer sur un bouton « A voté », qui publiait automatiquement cette information sur leur page. Des études précédentes montraient que « la satisfaction personnelle d'avoir voté était moins susceptible de vous pousser à vous exprimer que le possible jugement de votre famille et amis ». ⁹ Facebook a voulu tester cette affirmation, en direct. Après comparaison des archives de précédents scrutins, les chercheurs ont estimés que l'ajout du bouton « A voté » avait poussé environ 350 000 électeurs supplémentaires aux urnes. Deux ans plus tard, Facebook est allé plus loin. La question cette fois, portait sur une influence psychologique directe. En utilisant un logiciel de linguistique, ses chercheurs ont classé des actualités en deux groupes : positives ou négatives. Pendant une période donnée, ils ont exposé un panel test à des nouvelles déprimantes, et un autre à des nouvelles joyeuses. Résultat, le « panel dépression » publiait plus de post négatifs et le « panel joyeux », plus de posts positifs. Précisons que ce test a provoqué une grosse polémique étant donné que les utilisateurs n'étaient pas informés de leur nouveau statut de cobayes. Tout comme dans l'exemple du *NewsFeed*, le réseau s'est servi d'eux pour tester son pouvoir. « Qu'est ce qui se passerait s'ils jouaient avec les émotions des gens le jour des élections ? » ¹⁰ Que Facebook ai un tel pouvoir d'influence est assez préoccupant. Mais pas de panique : comme il le vend, il n'est pas le seul à en profiter.

DONNÉES SORTANTES

Encore un peu de *Minority Report*. Dans le film de Spielberg, nous assistons à une scène où le héros passe dans un centre commercial, où des publicités s'adressent directement à lui par son prénom, et sont contex-

9. « *Studies have shown that the quiet satisfaction of carrying out a civic duty is less likely to move people to vote than the possible judgment of friends and neighbors* » Cathy O'NEIL, *Weapons of math destruction: how big data increases inequalities and threatens democracy*, Crown : New York, 2016

10. « *What would occur if they played with people's emotion on the election day?* » Cathy O'NEIL, *Weapons of math destruction: how big data increases inequalities and threatens democracy*, Crown : New York, 2016

tualisées par rapport à sa vie personnelle. C'est un peu ce que permet le réseau social. Tout comme Google vend les premiers rangs de ses résultats de recherche, Facebook se rémunère en commercialisant des espaces pub dans le fil d'actualité de ses utilisateurs. Les pubs sont directement mélangées aux publications de pages suivies, photos d'amis et autres événements à venir, placées par EdgeRank proportionnellement à l'investissement de l'annonceur. L'idéal étant, de faire passer la pub pour de l'information, comme chez Google. En tant que page Facebook « sponsorisée », vous pouvez outrepasser l'arbitraire de l'algorithme et vous assurer un morceau du fil de vos *followers* (abonnés), en payant. Cette garantie ne s'accompagne naturellement d'aucune preuve, le code étant toujours fermé. Ces pratiques ne sont pas sans rappeler, cyniquement, les passe-droits que les concepteurs de jeu vidéo en ligne, octroient contre de l'argent à certains joueurs. Au lieu d'effectuer une tâche répétitive ou d'attendre pour passer au niveau suivant, vous pouvez tout simplement payer.

En tant qu'annonceur, il vous est également possible de cibler des catégories de personnes, pour que vos publicités apparaissent spécifiquement dans leurs actualités. Ces catégories sont directement tirées des informations que les utilisateurs ajoutent à leur profil, dans les rubriques « Travail », « Loisir », « École » et autres, par un algorithme qui les propose automatiquement à la vente. Une illustration pour le moins littérale de la monétisation de l'information. En résumé, Facebook fournit donc à la fois les données pour profiler les consommateurs et les outils pour les influencer. **#businessplan #génie**

TANT QUE VOUS PAYEZ...

Dans un article publié le 14 septembre 2017, le site d'information américain *ProPublica* révèle que pour \$40, le réseau permettait aux annonceurs de cibler des publics antisémites, qui s'étaient déclarés sur leur profils comme « *Jew Haters* » (Anti Juifs), « *Nazi* » ou d'autres variantes du même goût. Après le signalement de *ProPublica*, l'entreprise a retiré ces catégories et certifier empêcher les dérives futures de l'algorithme, en limitant le nombre de catégories disponibles ou en les vérifiant avant d'autoriser le ciblage.¹¹

11. Julia ANGIN, Madeleine VARNER, Ariana TOBIN, Facebook Enabled Advertisers to Reach 'Jew Haters', *ProPublica*, 14.9.2017

« Si tu suis l'argent, tu ne sais jamais sur quoi tu va tomber. »¹²

L'article signale également que le réseau a annoncé début septembre 2017 avoir découvert que des publicités avait été placées pour un montant avoisinant \$100.000, par de « faux comptes » affiliés à la Russie. Plus qu'un rôle de censeur que devrait ou ne devrait pas adopter la plateforme, cela questionne l'économie qui sous-tend son fonctionnement. Ce que cette controverse met en cause n'est pas l'algorithme mais la vente d'influence au plus offrant.

FILTER TARGETING

Micro targeting et *filter bubbles* ne sont au fond, que les deux faces d'un même pièce. Comme le souligne Eli Parser dans une interview à *Wired*: « Je n'avais pas réalisé, quand j'ai eu cette image d'un tas de sources médiatiques puis d'une membrane ou d'un film qui entoure une personne que ces sources les traverseraient, que le système dans son ensemble deviendrait conscient de ce phénomène, dans le sens où les médias se construirait autour de lui. C'est ce qui s'est passé. Vous pouvez cibler des niches ou des communautés très spécifiques, et atteindre beaucoup de ces gens, simplement en comprenant comment l'algorithme marche. »¹³

Implications politiques

3
V

« La convergence du *Big Data* et du marketing procure désormais aux politiciens des outils bien plus puissants. Ils peuvent cibler des micro-groupes de citoyens, à la fois pour un vote ou de l'argent, en les séduisant chacun avec un message méticuleusement affûté, que personne d'autre n'a de chances

12. « *If you follow the money, you'll never know what you'll get.* » Inspecteur Lester FREAMON, personnage de *The Wire*, David SIMON, 2002

13. « *It hadn't really fully occurred to me, when I first had this image of a bunch of media sources and then a membrane or filter that surrounds a person that those sources get through, that the whole system would become self aware in a certain sense—that the media organizations would grow autotrophically toward those bubbles. I think certainly that has happened. You can target very particular niches or communities and reach a lot of those people, and do it by understanding how that algorithm works and what it lets in.* » Jessie HEMPLE, Eli Parser Predicted the Future. Now He Can't Escape It, *Wired*, 24.5.17

de voir. Ce peut être une bannière sur Facebook ou un mail de collecte de fonds. Mais chacun permet aux candidats de vendre discrètement de multiples versions d'eux-mêmes — à nous de deviner ensuite laquelle d'entre elle se pointera à l'investiture. »¹⁴ Les candidats ne sont donc plus forcés de tenir un discours commun, derrière lequel les citoyens choisissent ou non de se rallier. Étant donné qu'en France comme au États-Unis, ils ne sont pas légalement tenus de réaliser leur promesses de campagne, ils peuvent les adapter « à la tête du client » et une fois élu, gouverner de la manière qui leur sied. D'après Zeynep Tufekci, le danger principal des bulles se trouve dans « l'ignorance mutuelle du voisin », et permet de servir des discours extrêmes à certains segments, tout en se gardant bien de les montrer à d'autres, qui pourrait être « repoussés voir même dégoûtés » par ce genre de messages. Il devient impossible de rassembler la totalité des publicités, d'avoir accès à l'ensemble de la campagne, protégée par les bulles.¹⁵

14. « *The convergence of Big Data and consumer marketing now provides politicians with far more powerfull tools. They can target microgroups of citizens for both votes and money and appeal to each of them with a meticulously honed message, one that no one else is likely to see. It might be a banner on Facebook or a fund-raising e-mail. But each one allows candidates to quietly sell multiple versions of themselves — and it's anyone's guess which version will show up for work after inauguration.* »
15. « *According to Zeynep Tufekci, [...] these groups pinpoint vulnerable voters and then target them with fear-mongering campaigns, scaring them about their children's safety or the rise of illegal immigration. At the same time, they can keep those ads from the eyes of voters likely to be turned of (or even disgusted) by such messages.* » Cathy O'NEIL, *Weapons of math destruction: how big data increases inequalities and threatens democracy*, Crown: New York, 2016

MANIPULATION

Un ancien employé de *Cambridge Analytica*, parti à la suite de la campagne en faveur de Trump, explique que l'entreprise s'auto-définissait comme « une arme psychologique » (*psychological warfare*). Il raconte que les méthodes utilisées sont « les mêmes que celles qu'utilisent les militaires pour provoquer un changement massif d'opinion. [...] On faisait ça pour gagner des élections dans le genre de pays en développement qui n'ont pas trop de règles. [...] C'était comme travailler pour le MI6. »¹⁶

De telles pratiques dépassent clairement le cadre de la publicité. Ce que vend ce type d'entreprise, c'est un outil d'étude et de manipulation psychologique des populations basé sur l'espionnage de leurs activités en ligne. Et qui a suffisamment d'argent, peut se l'offrir pour la bagatelle de quelque millions de dollars, sans que cela pose de problème légal aux USA. Bienvenue dans la propagande du 21e siècle.

ASYMÉTRIE

« Il peut y avoir une faible transparence au sein des communautés dans lesquelles nous vivons (on connaît peu nos voisins, et nos collègues de travail ont leur vies privées rigoureusement protégée) pourtant les *data miners*, les hackers et les institutions peuvent accéder à de nombreuses informations nous concernant. Il n'y a pas de symétrie. »¹⁷

Nous constatons que le monde désormais informatique permet très facilement à un petit nombre d'acteurs de tout savoir sur une très large population, sans que l'inverse soit vrai, et sans nécessiter d'autres dispositif que le réseau Internet. Cela est du à ce que Gérard Berry appelle la nature « imma-

16. « *the same methods the military use to effect mass sentiment change. [...] We were just doing it to win elections in the kind of developing countries that don't have many rules. [...] It was like working for MI6.* » Source anonyme, propos recueillis par Carole CADWALLADR, *Cambridge Analytica affair raises questions vital to our democracy*, *The Guardian*, 4.3.2017

17. « *There can be little transparency within the community we live in (we hardly know our neighbours, and our fellow workers have their privacy rigourously protected) yet data-miners, hackers, and institutions can be well informed about us. There is no symmetry.* » Luciano FLORIDI, *The Fourth Revolution: How the Infosphere is Reshaping Human Reality*, Oxford: Oxford University Press, 2016

térielle de l'information », ¹⁸ qui efface virtuellement toutes les contraintes de circulation liées au support qu'elle subissait dans le monde analogique, qui compliquait sa concentration (mais ne l'empêchait pas pour autant).

CULPABILISATION

Il est complètement légitime de mettre en garde les personnes des risques qu'elles encourent en se dévoilant sur Internet. Mais se protéger seul contre de telles machines est tout simplement impossible. Le discours déresponsabilisant des plateformes, qui consiste à dire « vous avez vous-mêmes fourni ces informations et en utilisant nos services, avez donné votre accord pour l'utilisation que nous faisons de vos données » est écœurant. Ce système se nourrit du déséquilibre des connaissances en informatique, et des vides législatifs qui résultent de la mutation trop rapide de nos sociétés, d'analogiques vers informatiques. Les gouvernements démocratiques se doivent de protéger leurs peuples contre de telles atteintes à leur liberté de choisir *en leur âme et conscience*. D'un côté, en encadrant le commerce des données par des lois, et en les faisant respecter. De l'autre, en offrant à tous l'accès à une éducation informatique minimum. Ils doivent retrouver leur rôle de contre-pouvoir face au secteur privé. Pour cela, encore faudrait-il qu'ils ne soient pas complices de cette surveillance... **IV**

SAVOIR = POUVOIR

À travers une étude de l'évolution des manières de punir en France du XVI^e siècle à nos jours, le philosophe Michel Foucault explique que notre société a vu une mutation des moyens de contrôle et de la manière dont le pouvoir s'exerce sur les individus. D'une société qui juge ses membres d'après leurs actes, nous sommes passés à une société normalisatrice qui tend à proscrire des comportements. La discipline quelque soit la forme qu'elle revêt, incarne à ses yeux le moyen par lequel s'exerce une nouvelle contrainte : la norme. Dans *Surveiller et punir*, il explique que cette mutation est notamment passée par la science et la collecte de donnée qu'elle induisit.

18. Gérard BERRY, *L'Hyperpuissance de l'informatique*, Paris : Odile Jacob, 2017

IV. Annexe : « Le Scandale de la NSA, ce que peut faire la lutte anti-terroriste aux libertés »

« La constitution de "tableaux" a été un des grands problèmes de la technologie politique et économique du XVIII^e [...] c'est à la fois une technique de pouvoir et une procédure de savoir. Il s'agit d'organiser le multiple, de se donner un instrument pour le parcourir et le maîtriser ; il s'agit de lui imposer un "ordre". »¹⁹

Ces effets de ce savoir / pouvoir s'incarnent aujourd'hui pleinement dans le *Big Data*, qui offre une masse de données sans précédent quant aux comportements humains, qui « laissent tous des traces ». Les plus pessimistes mettent en lien la situation actuelle avec le Panopticon de Bentham, comme illustration de la toute-puissance des « data nouveaux riches » (miners, brokers, milliardaires et autres GAFAM). Il s'agit d'une architecture pénitentiaire imaginée par un théoricien anglais du XIX^e siècle, qui ne fut jamais construite tel quelle mais inspira des prisons à travers le monde. Elle fut pensée de manière à ce que le gardien puisse voir l'ensemble des prisonniers sans être vu. Elle incarne physiquement l'idée de surveillance permanente. Foucault explique que cette architecture est à la fois un moyen d'amasser des informations et un moyen de tester des techniques de discipline. Ce qui n'est pas sans rappeler Facebook qui teste l'influence du *NewsFeed* sur ses utilisateurs, et recueille tranquillement leurs données psychologiques. Foucault explique que « [l'] effet majeur [du Panopticon de Bentham est] d'induire chez le détenu un état conscient et permanent de visibilité qui assure le fonctionnement automatique du pouvoir. »¹⁹ Que la surveillance soit effective n'importe pas : sa seule probabilité suffit à soumettre ceux qui y sont sujets.

19. Michel FOUCAULT *Surveiller et punir* Paris : Gallimard 1975

« S'il y a quelque chose que vous ne voulez pas que les gens sachent, peut-être que vous devriez commencer par ne pas le faire. »²⁰

Eric SCHMIDT, Président du conseil d'administration d'*Alphabet* (Maison mère de *Google*)

« La surveillance est devenue une évidence. Beaucoup de gens à qui j'ai parlé expliquent qu'ils font attention à ce qu'ils tapent dans les moteurs de recherches parce-qu'ils savent que c'est enregistré. Cela limite le terrain de leur expérimentation intellectuelle. »²¹

Edward SNOWDEN, Lanceur d'alertes, dans un entretien tiré du film *CitizenFour*,
Laura POITRAS, 2013

C'est le principe de l'auto-censure, qui assure par le biais même de la surveillance l'obéissance de ses membres à la règle du groupe et étouffe la pensée contestataire en l'empêchant de se construire, sauvegardant ainsi l'ordre établi.

« S'il y a quelque chose que vous en voulez pas que les gens sachent, peut-être que vous devriez commencer par ne pas utiliser Google. »

Sagesse populaire

Mais je pense qu'au fond, le contrôle est une conséquence du véritable but que révèle cette phrase, qui est directement en lien avec les algorithmes. Ce que veut Schmidt, c'est vous garder dans le cadre de la machine : vous ramener vous et votre comportement dans le domaine des données qu'il exploite. Comme lorsqu'il dit « En fait, je ne pense pas que les gens veulent que Google réponde à leurs questions. Ils veulent que Google leur dise ce qu'ils devraient faire. »²² Il poursuit l'idéal du calcul effectif de l'humanité par sa machine culturelle, ce qui implique, par définition, une humanité automatique et répétitive.

20. « *If you have something you don't want anyone to know about, maybe you shouldn't be doing it in the first place.* »

21. « *It's become an expectation that we're being watched. Many people I've talked to mention that they're careful with what they type into search engines because they know that it's being recored. That limits the boundaries of their intellectual exploration.* »

22. *I actually think most people don't want Google to answer their questions. They want Google to tell them what they should be doing next.* Propos cités par Ed FINN, *What Algorithms Want*, Cambridge : MIT Press, 2017

Conclusion

VI

Vie privée

ENVIRONNEMENT

1 VI

« Nous vivons dans une seule et même infosphère, qui n'a pas d'extérieur. »¹

Les premiers ordinateurs fonctionnels, tel que l'ENIAC mis au point durant la seconde guerre mondiale par une équipe de chercheurs américains, étaient d'immenses machines qui s'étalaient sur plusieurs pièces. Le transistor, qui marqua le début de la miniaturisation compulsive des circuits électroniques, n'étaient pas encore inventé, et la technologie qui le remplaçait, le « tube électronique » (*vacuum tube*), était bien plus massive. Dans une conférence lors du 33^e Chaos Computer Congress, James Briddle, artiste britannique, cite l'un des ingénieurs ayant participé au projet, Harry Reed, qui explique son rapport à cette machine particulière : « J'ai toujours considéré l'ENIAC comme un ordinateur très personnel. Maintenant, nous pensons aux ordinateurs personnels comme des objets que nous transportons avec nous. Mais l'ENIAC était un ordinateur à l'intérieur duquel vous viviez. »² Il amène

1. « *We live in a single infosphere wich has no outside* » Luciano FLORIDI, *The Fourth Revolution: How the Infosphere is Reshaping Human Reality*, Oxford : Oxford University Press, 2016
2. « *I always thought of the ENIAC as a very personal computer. Now we think of personal computer as something that we kind of carry around with us. But the ENIAC was a computer that you lived inside.* » Harry REED cité par James BRIDDLE, *A New Dark Age - Turbulence, Big Data, AI, Fake News, and Peak Knowledge*, Chaos Computer Congress : Hambourg, 29.12.2016

ensuite une idée très pertinente, qui se vérifie un peu plus chaque jour à mesure que nous numérisons notre monde.

« Pour moi, le terme d'*ordinateur personnel* s'avère de manière croissante [...] et évidente, être faux. Parce-que nous vivons désormais tous dans une sorte de coquille informatique. Nous savons qu'ils sont justes des bornes, pour accéder à d'autres connections [...] L'ENIAC ne s'est pas miniaturisé dans ces appareils, il s'est en fait étendu tout autour de nous. »³

C'est ce que Luciano Floridi appelle *l'infosphère*. Il explique qu'en tant que membres de cet espace, nous existons maintenant en tant qu'*inforgs*, organismes informatifs. Cette interprétation nous permet aussi de lire les phénomènes tel que le *Digital Labor*, comme l'ajout de pièces humaines au fonctionnement d'un ordinateur global, à l'échelle de la planète. Cette redéfinition de notre être en tant qu'*inforgs* implique une valorisation des informations qui nous caractérisent, qui deviennent constitutives de notre identité au mêmes titre que l'est notre corps. « Si l'information personnelle est finalement reconnue comme étant constitutive de l'identité d'une personne et son individualité, peut-être deviendra-t-il un jour strictement illégal d'échanger certains types d'informations, de la même manière dont le trafic d'organe ou d'esclave est interdit. »⁴

HOMMES PLURIELS

Cette définition informationnelle fait néanmoins planer un autre danger, que nous avons évoqué précédemment. La constitution de notre identité par les seules données et la négation de « l'hors du cadre ». Toute information est un proxy de l'objet qu'elle décrit et non l'objet lui-même, et à ce titre, elle

3. « For me that division of personal computer is obviously and [...] increasingly not true. Because we all live inside this kind of shells of computation now. We know thses are terminals to other connections [...] The ENIAC hasn't contract into these things, it has actually expended out. » James BRIDLE, *A New Dark Age - Turbulence, Big Data, AI, Fake News, and Peak Knowledge*, Chaos Computer Congress : Hambourg, 29.12.2016

4. « If personnal information is finally acknowledged to be a constitutive part of someone's personnal identity and individuality, then one day it may become strictly illegal to trade in some kinds of information, exactly as it is illegal to trade human organs [...] or slaves. » Luciano FLORIDI, *The Fourth Revolution: How the Infosphere is Reshaping Human Reality*, Oxford : Oxford University Press, 2016

comprend intrinsèquement une réduction de la réalité à son format, qui ne peut, malgré toute sa puissance, embrasser la complexité d'un être humain, ou la complexité du monde. La définition informationnelle de l'individu doit coexister avec d'autres définitions, comme la mise en récit personnelle de sa vie. Par exemple, dans un procès équitable, le droit à la parole de l'accusé fait intervenir cette mise en récit, et l'inclut parallèlement aux faits, qui tout en étant indiscutables, sont explicables. « L'ambition totalisante de l'idéologie des *Big Data* présuppose en effet de congédier, de désactiver ou de neutraliser tout ce qui est non numérisable, imprévisible et indécidable qui ne pourrait se percevoir et se raconter qu'à condition de disposer de scènes interruptrices des flux numériques, hétérochroniques plutôt que synchrones au temps "réel". »⁵ Là où les *Big Data* se base sur des « phéromones numériques », qui ne résultent la plupart du temps d'aucune transcription intentionnelle des individus »,⁵ se raconter à soi et aux autres est une manière de donner un sens à son existence, qui comprend en germe la capacité « d'auto-détermination » et induit la possibilité d'agir quand le *Big Data* l'exclut.

« L'inadéquation, le désajointement de nos représentations au monde, au lieu d'être perçue comme un défaut, un accroc dans la trame fluide du réel, devrait être préservée en tant que scène de représentation, ou plutôt comme scène d'exposition des limites de la représentation, obligeant la fabulation, l'invention, l'interprétation, l'institution des individus - et des collectifs - comme "sujets" consistants, c'est-à-dire aussi, dotés d'agency ou de capacité d'agir. »⁵

CONFORMISME

Les algorithmes que nous avons étudiés favorisent tous une société uniforme. PageRank et EdgeRank, en valorisant la popularité. Les WordEmbeddings, en normalisant l'usage de la langue par rapport au plus grand nombre. *PredPol*, en partant du principe que la crimes de demain seront les mêmes que ceux d'hier. Et pour finir, Netflix, qui nous recommande des contenus selon ce que nos émules culturels apprécient. **V**

Nous pouvons discerner deux types de conformismes. Premièrement, le conformisme de notre moi d'aujourd'hui donc de demain, à notre moi d'hier.

5. Antoinette ROUVROY, *La Vie n'est pas donnée*, Académia, 2016

V. Annexe : « *Ersatzs d'extraordinaire* »

Ces algorithmes sont tous basés sur l'analyse du passé et ne peuvent intrinsèquement implémenter la nouveauté. Par là même, ils questionnent notre droit à une « identité renouvelable » (*renewable identity*) comme l'appelle Flordini et la « momifient » (*momify*). C'est ce qu'il nomme « la valeur auto-constitutive de la vie privée » (*self-constituting value of privacy*), comme terrain d'expérimentation de l'individualité. « *We never stop becoming who we are* », nous n'arrêtons jamais de devenir nous-mêmes. Enfermer les individus dans leur passé revient à leur retirer le droit fondamental d'être qui ils veulent, ou plus simplement, d'être qui ils sont. Ces algorithmes nient ce qu'Antoinette Rouvroy appelle le « principe d'autodétermination » des personnes, à savoir leur liberté fondamentale d'être seuls maîtres de leur vie.

Deuxièmement, ces algorithmes induisent un conformisme des comportements des individus entre eux. Ils fonctionnent tous selon des principes probabilistes, qui cherche notre avenir dans le passé de nos semblables, et considèrent qu'ils ont toutes les chances d'être similaires. C'est ce phénomène qui provoque les bulles et l'enfermement culturel, exaltant la division tout en niant l'individualité. Ces systèmes véhiculent un idéal très préoccupant : celui d'un monde où tout événement relèverait d'un lien cause / conséquence quantifiable, et serait le résultat d'une logique transparente. Mais comme le rappelle la philosophe Hannah Arendt, « l'épreuve de causalité — la prévisibilité de l'effet, si toutes les causes sont connues — ne peut être appliqué au domaine des affaires humaines ; mais cette imprévisibilité pratique n'est pas une preuve de liberté, elle signifie simplement que nous ne sommes jamais en mesure de connaître toutes les causes qui entrent en jeu, cela, d'une part, à cause du grand nombre de facteurs, mais aussi parce que les motifs humains, à la différence des forces naturelles, demeurent cachés à tous les regards, à l'observation des autres hommes comme à l'introspection. »⁶

De ce fait, une prédiction basée sur un ensemble de données ne peut être juste, dans le sens où elle prendrait en compte toute la complexité de la réalité, comme l'est la prédiction de certains phénomènes naturels. Quiconque prétend le contraire est un menteur. Comme le dit Arendt, les motifs humains ne se révèlent que dans la réalisation de l'action, à celui qui la provoque ou à un regard autre. C'est la base de l'analyse rétrospective de l'histoire, qui jamais ne pourra se renverser en écriture du futur. La prédiction comporte

6. Hannah ARENDT, *La Crise de la culture*, Gallimard : Paris, 1963

donc une forme de violence arbitraire, qui impacte le sujet au présent sur la base d'un acte hypothétique donc inexistant. Imputer par exemple, un risque de récidive à une personne qui commet un délit pour la première fois, est dans n'importe quel cas fondamentalement injuste, étant donné que la seule donnée pertinente, à savoir une récidive antérieure, est absente.⁷

AGIR

Comme le dit si bien le personnage d'Albus Dumbledore dans le roman de J.K. Rowling *Harry Potter et la chambre des secrets*, «Ce ne sont pas nos capacités qui définissent qui nous sommes. Ce sont nos choix.» Être catalogué d'après un comportement inconscient (qui déjà en lui-même est assez fumeux) est une aberration. À l'heure des *Big Data*, lors de la répartition dans les différentes maisons (groupes d'élèves) de l'école des sorciers Poudlard, Harry Potter serait allé à Serpentard (la maison des mages noirs) et non à Gryffondor, car son caractère l'y prédestinait. Il n'aurait pas eu le choix. Cela reflète une vision de l'homme comme un être préprogrammé qui n'a aucun pouvoir de décision sur qui il est, ou autrement dit sur ce qu'il fait, réhabilitant le fantasme du naturel. Cette vision déterministe nie l'existence de la liberté, en tant que possibilité de commencement, d'introduction de nouveauté au monde.

L'idée d'un soi immuable, que les *Big Data* seraient en mesure de révéler me paraît extrêmement dangereuse, car elle induit une déresponsabilisation du sujet. «Je suis comme ça, j'y peux rien». À moins d'une preuve du contraire, les êtres humains sont tous responsables de leurs actions, et leur personnalité n'est en rien une excuse pour faire le mal. «Nous n'arrêtons jamais de devenir nous-mêmes» souligne également que les choix impactent la constitution de l'être. Je nuancerais tout de même mon propos, en ajoutant que les hommes sont maîtres du sens de la vie qu'ils peuvent avoir. Le contexte, les événements extérieurs, s'ils ne déterminent pas les possibles, les impactent tout aussi fortement que les choix.

7. ANGWIN Julia, LARSON Jeff, MATTU Surya, KIRCHNER Lauren, *Machine Bias*, *ProPublica*, 23.05.2016 — Sur le sujet de la récidive et des algorithmes.

LIBERTÉ

Le droit à la vie privée comprend donc le droit fondamental à l'action. Action en tant que rupture envers notre passé. Action en tant que rupture envers celui du groupe. Action en tant que nouveauté. Nier notre pouvoir de décision, c'est nier notre liberté et notre humanité en nous ramenant au rang de phénomène calculable.

« La substance de l'histoire est composé de ses interruptions, autrement dit, de l'extraordinaire. »⁸

Hannah Arendt

#timeToWalkOnWater

VIE PRIVÉE D'AUJOURD'HUI ≠ VIE PRIVÉE D'HIER

« Les mêmes techniques, qui aujourd'hui distribuent nos données personnelles aux quatre vents, permettent de construire des systèmes très différents, qui respectent la confidentialité des données personnelles. La technique n'impose nullement que de telles données deviennent publiques. »⁹

L'anonymat du XIX^e siècle est bel et bien mort et enterré. De nos jours, l'absence d'informations n'est plus la conséquence d'une impossibilité à les rassembler mais un choix délibéré. Pour rendre ce choix accessible aux citoyens, nous avons besoin d'une protection, qui garantisse nos libertés face à une ingérence trop grande à la fois du secteur privé et des États.

8. *La crise de la culture*, Gallimard : Paris, 1963

9. Serge ABITEBOUL, Gilles DOWEK, *Le Temps des algorithmes*, Paris : Le Pommier, 2017

Changer les choses, trois approches 2

VI

Dans un article pour le magazine *Usbek et Ricca*, le journaliste Vincent Lucchese raconte son expérience de « dégooglisation intense ». ¹⁰ Afin de tester l'efficacité des alternatives pour sortir de l'économie des GAFAs, il a donc supprimé tous les services affiliés à la marque qui lui était pourtant quotidiens. Du mail au moteur de recherche en passant par les sites de streaming vidéo, il dresse un petit panorama de l'éco-système du libre, qui se développe en réaction en outre à la commercialisation outrancière de données à laquelle se livre les géants du web. Lucchese cite par exemple la plateforme *Framasoft*, qui référence les alternatives libres aux services propriétaires et intrusifs.

« Pour préserver la logique du morcellement des données, qui participe à leur sécurité, *Framasoft* oblige à créer un compte différent pour chacun des services libres qu'il propose. Se profile ainsi le dilemme majeur pour les internautes entre confort et vie privée. » ¹¹

Face à la perte de leur vie privée, de nombreux utilisateurs font le choix d'une désintoxication similaire. Et il est indéniable que celle-ci engendre une perte de confort, principalement dans la communication avec autrui. Si demain, je décide de supprimer *Whats App* (filiale de Facebook) et d'utiliser un service de messagerie équivalent qui garantit la vie privée tel que *Signal*, il faut que toutes les personnes avec lesquelles je communique fassent de même. Néanmoins, ces services restent minoritaires, peu connus, et sont de ce fait assez élitistes.

1

L'APPROCHE INDIVIDUELLE

Dans la matinale de *France Inter* du 30 novembre 2017, dans un débat autour de l'interdiction du glyphosate, ce pesticide cancérigène commercialisé par Monsanto, Arnaud Daguin, chef cuisinier et conseiller en stratégie

10. LUCCHESI Vincent, *Ma vie sans Google*, *Usbek & Ricca*, n°20 — Nov/Déc 2017

11. Vincent LUCCHESI, *Ma vie sans Google*, *Usbek & Ricca*, n°20 — Nov/Déc 2017

alimentaire souligne qu'en définitive, ce sont les comportements des consommateurs qui doivent changer en amont pour modifier le modèle agricole dans son ensemble. Comme il le dit à regrets, « le bulletin de vote aujourd'hui, c'est la carte bleue ». Au niveau de web, « la carte bleue » se traduit par votre présence sur un site, ou votre utilisation d'un service. L'approche, d'une attitude citoyenne consciente au niveau individuel, est également celle défendue par Pierre-Yves Gosset le fondateur de Framasoft. « Si les gens ne sont pas prêts à payer un minimum pour avoir une adresse e-mail qui ne soit pas espionnée, il ne faudra pas qu'ils se plaignent de basculer dans une société dystopique dans le style de la série Black Mirror ». Comme le dit Vincent Lucchese, pour lui « la responsabilité repose autant sur les GAFA que sur les utilisateurs, en demande de toujours plus de confort ».¹²

2

L'APPROCHE LÉGISLATIVE

Dans cette même matinale, le 9 novembre 2017, le député François Ruffin défend un positionnement différent. Une auditrice soulève cette question, en lui demandant si à son avis, l'attitude des citoyens au niveau individuel, en appliquant une éthique dans les choix de consommation, serait un levier vers un monde plus juste et un moyen d'expression.

« Personnellement je suis assez sceptique sur le pouvoir du caddie. Si on regarde dans notre histoire, il n'y a aucun droit [...] qui a été construit par le boycott. Ça ne veut pas dire que chacun, individuellement, on ne peut pas réorienter notre consommation [...] Mais je crois toujours que ça passe par la loi. Je crois que ça passe par des règles communes que l'on se fixe. »¹³

J'ajouterais que si l'on prend l'exemple de l'agriculture, manger bio coûte encore plus cher que manger du poison. Il y a là une inégalité aberrante, que seule la loi permettrait de supprimer, car contrairement au pouvoir d'achat, elle est la même pour tous. Si l'on suit la logique du « pouvoir du caddie », une nouvelle fois, le droit à la parole exclut les plus pauvres. Je partage l'avis de François Ruffin, qui soutient que la seule action individuelle est impuissante, mais pas inutile pour autant. Dans le cas de l'agriculture, la construction d'alternatives aux cultures chimiques, que financent directe-

12. Vincent LUCCHESI, *Ma vie sans Google*, *Usbek & Ricca*, n°20 — Nov/Déc 2017

13. François RUFFIN, *Inter'activ*, 7/9, *France Inter*, 9.11.2017

ment les consommateurs de bio, a permis de démontrer leur viabilité, et leur capacité de rendement. Dans le cas du web, l'émergence de plateformes libres montre qu'il est possible de générer un revenu sans publicité et sans tracer les utilisateurs. Néanmoins, cela reste insuffisant.

L'accès à un engagement individuel véhicule lui aussi des inégalités, soit d'un point de vue économique comme avec le bio soit d'un point de vue technique comme avec le mouvement du libre. Nous avons tous droit à une alimentation sans poison. Nous avons tous droit à une vie privée. De plus, faire partager la faute à l'utilisateur contribue d'une certaine manière à légitimer l'espionnage. Alors qu'il est tout simplement scandaleux. En mai 2018, une nouvelle loi européenne votée en 2016, la GDPR entre en vigueur. Celle-ci renforcera le contrôle des citoyens sur leurs données personnelles et représente un premier pas vers une protection décente. ^{VI}

MUTATIONS DÉMOCRATIQUES

Notre époque est marquée par une défiance croissante à l'égard des appareils d'États et des institutions. Détournement de fonds public (affaire Fillon), promesses non tenues (« Mon ennemi c'est la finance. »), gouvernement par ordonnances (lois travail), poids des lobbies (glyphosate), fiscalité inégalitaire (suppression de l'impôt sur la fortune), surveillance (Loi anti terroriste) les citoyens perdent espoir dans une démocratie qui les ignorent, et délaissent les urnes. Pourtant, la protection égale des individus ne peut venir que de la loi, qui doit s'accompagner de moyens mis en œuvre pour la faire respecter. L'asymétrie de l'application surlignée par *PredPol* concerne de nombreux délits, comme l'évasion fiscale qui, pour cause de manque de moyens des brigades d'enquêtes et non formation délibérée des policiers, continuent d'être sous-punis.

Dans leur livre, *Le Temps de algorithmes*, les chercheurs Serge Abitboul et Gilles Dowek expliquent que les gouvernements sont des « processus purement informationnels. [...] Le gouvernement des États et des collectivités locales est l'un des domaines dans lesquels cette vision est la plus exacte : dans les boulangeries nous transformons de la farine en pain, dans les usines du métal brut en voitures, mais dans le gouvernement des États et des collectivités locales, nous ne transformons que de l'information ». Ils utilisent une

VI. Annexe : « **Ce que dit la Loi européenne** »

métaphore très pertinente, tirée de l'informatique pour décrire une situation de déséquilibre entre les citoyens et le pouvoir d'état, qui ne les représentent plus. Aujourd'hui comme ils l'expliquent, s'exprimer seulement tous les 5 ans n'a plus aucune justification si ce n'est l'auto-sauvegarde d'un ordre social :

« Dans l'analyse de ces systèmes, une question essentielle est celle de la quantité d'information échangée par les agents. Si tous les cinq ans, par exemple, nous votons pour élire le chef de l'État, parmi trente-deux candidats, chaque citoyen transmet à la collectivité cinq bits d'information à chaque élection car $32 = 2^5$. Le débit d'information du citoyen vers la collectivité est donc de cinq bits tous les cinq ans, soit un bit par an. En ajoutant les élections législatives, locales, etc., les élections, dans leur ensemble, constituent un canal de communication dont le débit est de cinq ou six bits par an, soit, pour employer une unité plus courante, de l'ordre de 0.0000001 bit par seconde. Par comparaison, le débit montant de notre box domestique est de quelque dix mille milliards de fois supérieur. »

Les techniques modernes d'information et de communication peuvent être utilisées pour surveiller la population comme pour lui permettre de s'exprimer. L'enjeu du vote par voie numérique est aussi celui de la sécurité : il faut être capable de mettre en place un système d'expression absolument sûr et invulnérable au piratage. Si la programmation est capable de faire voler des avions, dont les programmes sont soumis à des tests extrêmement poussés, il est possible de mettre en place des moyens d'expression sécurisés, en les astreignant aux mêmes niveaux d'exigences. La démocratie participative prend racines dans l'impossibilité matérielle du XX^e siècle à entendre la parole des citoyens. D'où la nécessité de déléguer, de députer une personne pour représenter la représentation auprès de l'État. La faible fréquence des élections découle également en partie d'une difficulté logistique à les organiser qui aujourd'hui n'a plus lieu d'être. « Nos institutions cependant continuent à demander aux citoyens de compresser, en simplifiant toujours autant, leurs souhaits, alors qu'il est devenu possible de faire autrement. De manière plus générale, nous utilisons peu, dans le gouvernement des États et des collectivités locales, les possibilités qui, au temps des algorithmes, nous sont offertes. Cette inadaptation des institutions à l'avancée des techniques, dans un monde devenu par ailleurs beaucoup plus complexe, participe à leur obsolescence, et au-delà, à la défiance grandissante des citoyens à leur égard. »¹⁴

14. Serge ABITEBOUL, Gilles DOWEK, *Le Temps des algorithmes*, Paris : Le Pommier, 2017

L'APPROCHE INVENTIVE

« Facebook, c'est un réseau de vieux. »

Lise, 15 ans

Si Facebook était le *must have* des années 2000, sa cote de popularité est en chute libre chez les moins de 18 ans. La raison principale de ce désintérêt est totalement étrangère à l'éthique : aujourd'hui, leurs parents sont sur Facebook, et peuvent les espionner. Ce qui ramène donc, de manière détournée, au problème de la vie privée sur Internet. Ce désintérêt s'est fait au profit d'applications comme *Snapchat* (encore inconnue des parents), qui permet principalement de publier des photos et des vidéos, visibles par un cercle de contacts choisis. Celles-ci s'autodétruisent après un laps de temps défini par l'utilisateur. Il reste néanmoins possible aux membres de votre cercle de faire des captures d'écrans de ces photos pour les conserver (d'où l'arrivée du verbe *screenshooter* dans le vocabulaire français), ce qui n'empêche pas leur suppression automatique des serveurs de la plateforme une fois le temps écoulé. L'application se place donc en opposition avec l'idée de traces numériques, fondant la qualité de son expérience utilisateur sur le caractère éphémère des contenus, qu'elle ne conserve pas. L'idée est d'ouvrir une fenêtre publique à un instant T, puis de la refermer, de diffuser une image « orale », qui de la même manière que les mots dits, s'évanouit rapidement des mémoires et des serveurs.

D'autres modèles de services, en marge de l'idée de personnalisation de l'expérience utilisateur par la récolte de leurs traces, et du confort qu'elle garantirait se développent et sont massivement adoptés. Ils répondent à des besoins différents, et l'exemple de *Snapchat* démontre que l'éphémère, dans un web ultra conservateur, peut-être un atout vendeur. Les utilisations d'Internet sont en constante mutation. Si les géants du Web, qui battirent leurs empires sur la publicité alliée au traitement algorithmique des données personnelles, nous paraissent indétrônables, l'avenir du modèle qu'ils défendent reste inconnu. L'alternative peut également naître d'une interprétation différente de l'espace numérique et de ses possibilités. Contrairement à l'idéal répétitif que prônent leurs algorithmes, le monde change.

Leur puissance économique pose néanmoins d'autres problèmes, comme la constitution d'un monopole par rachat automatique de la concurrence. Mais ces conflits ne sont pas caractéristiques du numérique, et concernent toute entreprise hégémonique, quelque soit son domaine d'activité.

QU'EST-CE QU'UN BON ALGORITHME ?

Un bon algorithme est, avant tout, une machine transparente. Sa recette doit faire partie de l'espace public, afin que tout ceux qu'elle affecte soient en mesure de valider ou d'infirmer son emploi. Si l'État utilise des algorithmes, dans la police par exemple, ceux-ci doivent être validés par l'ensemble des citoyens, comme le sont les lois ou les mesures qu'il prend. Les boîtes noires, comme tout exercice d'un pouvoir arbitraire, n'ont pas leur place dans une société juste.

En résumé, un bon algorithme est une machine qui respecte les mêmes règles de société que les êtres humains. Éthiquement et légalement, la discrimination par exemple, est condamnable. Le fait qu'elle soit l'œuvre d'un dispositif automatique ne la rend pas tolérable et encore moins justifiable. S'il s'avère que le respect de l'équité ou d'une autre règle commune, dans un cas précis, n'est pas automatisable, et donc ne peut être exercé par une machine, alors, comme dans l'exemple du comité de presse de Facebook, l'automatisation doit cesser. Tant que, pour une situation donnée, les êtres humains restent, malgré tous leur biais, plus justes, plus efficaces que les machines, elles ne peuvent pas les remplacer. Et ce même si leur rentabilité économique en fait saliver plus d'un.

Ce qui nous nous ramène au problème de base qui caractérise absolument tous les algorithmes : celui de l'implémentation. Toutes les affaires humaines ne se prêtent pas au calcul, qui ne peut à lui seul, définir les hommes où le monde. Les cas que nous avons étudiés démontrent la violence que l'automatisation peut provoquer, de par son caractère aveugle, ou en tant qu'outil de contrôle (machine au service des puissants) ou encore, en forçant les hommes à s'adapter à elle (*Digital Labor*). Nous devons apprendre de ces expériences, et comprendre l'ontologie de la machine, et du même coup la notre, qu'elle a complètement chamboulé et renouvelé. Pour qu'elles soient à même, de son fondre au plus juste.

QUI EST LA MACHINE

La première machine à calculer fut inventée en 1644 par le mathématicien et philosophe français Blaise Pascal. L'impulsion de travailler sur cette invention lui vint de son père, qui, nommé surintendant de Haute-Normandie, devait s'occuper de comptabiliser les impôts de la province qu'il administrait. Désirant le soulager des affres de calculs compliqués, Pascal mit au point une machine à calculer mécanique, la Pascaline, qui demeura pendant 2 siècles, la seule machine de calcul opérationnelle. Cette notion de soulagement a perduré, et demeure à mes yeux très importante. Elle peut nous guider dans la construction de l'altérité de nos machines, qui nous soutiennent là où nous souffrons, nous remplacent là où nous échouons, et nous enlèvent le fardeau de la survie. L'asservissement des hommes par leur technique est pour moi un non-sens total et un mensonge, qui en fait, recouvre l'asservissement bien plus trivial et violent d'un groupe d'hommes par un autre.

De manière globale, la technique a confisqué notre temps à la mort, que nous ne devons plus perpétuellement combattre, et nous l'a rendu. À nous d'en faire le meilleur usage, en sachant perpétuellement réapprendre, à être humains. Si nous ne sommes pas assez stupides pour nous entretenir ou détruire nos ressources naturelles. Et c'est pas gagné.

Annexes

Machines universelles

I
A

« Un point absolument essentiel est que la nature physique des objets matériels qui servent à stocker les données et faire les calculs est indifférente. »¹

L'ORDINATEUR MÉCANIQUE

Avant l'invention de l'ordinateur au milieu du XX^e siècle, compter nécessitait beaucoup de temps et comportait une grande marge d'erreur. Certaines disciplines, tel que la navigation, utilisaient des technologies qui exigeaient de nombreux calculs. Afin de se repérer en mer, on éditait d'immenses tables, dans lesquelles étaient répertoriées les positions des astres à chaque heure (ou moins) sur une année, selon la position du regardeur. Ainsi, les marins étaient-ils capables de déterminer où ils étaient sur le globe et de se diriger. Les coordonnées imprimées étaient calculées patiemment à l'avance par des êtres humains, et donc, parfois (fatigués, inattentifs) fausses. Ce qui pouvait avoir de lourdes conséquences.

Lassé de découvrir des erreurs dans les tables scientifiques de son époque, le mathématicien britannique Charles Babbage entreprit de mettre au point une machine capable d'effectuer ces calculs et de les imprimer automatiquement : La *machine à différences*. La construction de ce gigan-

1. Gérard BERRY, *L'hyperpuissance de l'informatique*, Paris : Odile Jacob, 2017

tesque calculateur mécanique débute en 1823 mais n'est jamais achevé. En effet, les technologies de l'époque ne permettent pas d'atteindre un niveau de précision suffisant. En 1834, Babbage entreprend de dessiner les plans d'une seconde machine : la *machine analytique*, qui est aujourd'hui considérée comme le premier ordinateur.

« Si vous regardiez à l'intérieur d'un processeur en silicium, vous verriez quelque chose de très similaire [aux plans de l'unité centrale de la machine analytique]. »²

Aussi incroyable que cela paraisse, cette machine de rouages et de vapeur fonctionne comme un ordinateur moderne. Simplement, elle utilise les technologies de son époque. Au lieu de circuits imprimés, c'est avec des roues crantées qu'elle compte, qu'elle compte. Au lieu de l'électricité, c'est à la vapeur qu'elle fonctionne. Elle est un ordinateur, un ordinateur mécanique.

« Mais [la machine analytique] pouvait également faire quelque chose que fait un ordinateur et pas un calculateur : cette machine pouvait regarder sa propre mémoire et prendre une décision. »³

La machine analytique ne fut jamais construite. Il fallut attendre les années 1940 pour que le premier ordinateur fonctionnel soit inventé. Ce sont les progrès en physique, avec l'invention de composants capables d'utiliser les propriétés de l'électricité qui le rendirent possible. Mais pas seulement.

L'ORDINATEUR THÉORIQUE

En 1936, le mathématicien britannique Alan Turing publie l'article « *On Computable Numbers, with an Application to the Entscheidungsproblem* », qui expose sa *Théorie de nombres calculables* dans la revue de la London Mathematical Society. Cet article s'interroge sur un problème mathématique : la

2. « If you did look inside a silicon chip, you would see something very similar to [the analytical engine's CPU plans]. » John GRAHAM-CUMMING, *The greatest machine that never was*, TEDxImperialCollege : Londres, 03.2012
3. « But it could also make something that a computer does and a mechanical calculator doesn't: this machine could look at its own memory and make a decision. » John GRAHAM-CUMMING, *The greatest machine that never was*, TEDxImperialCollege : Londres, 03.2012

décidabilité, c'est-à-dire la capacité ou non de répondre à une théorie mathématique, de prouver qu'elle est vraie ou fausse, selon une série d'opérations effectuées selon les axiomes de cette théorie. Capacité à savoir, si elle est décidable. Afin de prouver ou d'infirmer cette décidabilité, Turing met au point une machine conceptuelle (pas physique) à partir de laquelle il est théoriquement possible d'effectuer n'importe quel calcul, qui explique la notion de « procédure mécanique ». Autrement dit, la notion d'algorithme. On parle « des » machines de Turing au pluriel, car il en donne une description minimale, que chaque machine spécifique utilise différemment. De nombreuses machines sont des machines de Turing.

Rapidement, elles se composent de deux rubans imaginaires, découpés en cases horizontales, illimités de chaque côté. Dans chacune de ces cases, peuvent être écrits / lus un nombre finis de symboles pré-définis, disons 0, 1 ou rien. Une tête de lecture / écriture les surplombe, et peut agir sur eux, selon un protocole minimal : lire le symbole / le remplacer / l'effacer / aller à gauche ou aller à droite. Le premier ruban sert au calcul (c'est l'équivalent du *processeur*). Le deuxième décrit une série d'instructions, que la machine doit suivre pour mener à bien le calcul (c'est l'équivalent du *programme*). Les symboles et opérations utilisés dans les deux rubans sont les mêmes. Cette machine décrit donc un ordinateur : une machine programmable où données et programmes sont stockés sous la même forme. Elle en est une abstraction conceptuelle.

« Le premier grand résultat de Turing est l'existence d'une machine universelle : il existe une machine U telle que tout calcul d'une machine spécifique M sur des données spécifiques D peut être réalisé par la machine universelle U en écrivant d'abord sur sa bande de calcul un programme décrivant complètement la machine M puis les données D . »⁴

Si Babbage fut le premier à incarner physiquement l'informatique avec sa machine programmable, Turing, ainsi qu'un second mathématicien, Alonzo Church en jetèrent les bases théoriques. Conjointes, leurs théories proposent une définition essentielle de la machine informatique. Une manière d'exprimer, en logique pure, un algorithme.

4. Gérard BERRY, *L'hyperpuissance de l'informatique*, Paris : Odile Jacob, 2017

« Turing nous a donné une leçon très importante. Il a posé les bases mathématiques de la science informatique et dit, que l'important, ce n'est pas comment on fait un ordinateur. L'important, ce n'est pas si l'ordinateur est mécanique comme l'était celui de Babbage ou électronique comme le sont les ordinateurs d'aujourd'hui, ou peut-être, dans le futur, des cellules ou de nouveau mécanique, avec les progrès des nanotechnologies [...] Toutes ces choses sont des ordinateurs. Il y a, dans un sens, une essence informatique. Et c'est la théorie Turing-Church. »⁵

Tous les ordinateurs sont des machines universelles.

5. « [...] Turing told us something very important. He laid down the mathematical foundations for computer science, and said, it doesn't matter how you make a computer. It's doesn't matter if the computer's mechanical like Babbage's was or electronical, like computers are today, or perhaps, in the future, cells, or mechanical again, once we get into nanotechnology. [...] All those things are computers. There is in a sens, a computing essence. This is called the Turing-Church thesis. » John GRAHAM-CUMMING, *The greatest machine that never was*, TEDxImperialCollege : Londres, 03.2012

Parlez-vous l'électrcité ? De la matérialité du bit

II
A

« Parce-que nous vivons dans un monde matériel.
Et que je suis une fille matérialiste. »⁶

Madonna, *Material Girl*, 1985

SYSTÈME BINAIRE

À la base de l'ordinateur, il y a le célèbre système binaire. Toute information présente sur un ordinateur existe physiquement sous la forme de nombres binaires, qui peuvent ensuite être manipulés / interprétés différemment selon les instructions que la machine reçoit. Quelque soit la forme sous laquelle est restituée cette information (images, textes, programmes, son etc.), elle est toujours stockée sous cette même forme, essentielle. Logique magique.

6. « 'Cause we are living in a material world. And I am a material girl. »

ZÉRO, UN

Contrairement à notre quotidien système décimal qui utilise 10 chiffres (de 0 à 9) pour compter, le système binaire n'en comporte que deux : 0 et 1. En décimal, lorsque nous arrivons à 9, nous repartons à 0, en ajoutant un chiffre à gauche pour désigner le nombre de dizaines ou de 10 puissance 1 que le nombre représente. Une fois les dizaines épuisées, nous recommençons l'opération en ajoutant à nouveau un chiffre à gauche qui cette fois désigne le nombre de centaines ou de 10 puissance 2.

COMPTER EN BASE 10

| | | | | | | | |
|-----------|------|------|------|------|------|-----|-----|
| 10^1 | 0 | 1, | 2, | 3, | 4, | ... | 9, |
| Unités | | | | | | | |
| 10^2 | 10, | 11, | 12, | 13, | 14, | ... | 99, |
| Dizaines | | | | | | | |
| 10^3 | 100, | 101, | 102, | 103, | 104, | ... | 999 |
| Centaines | | | | | | | |

Ce système est dit positionnel, ce qui veut dire que la place du chiffre dans le nombre modifie sa signification. En effet : $120 \neq 201$. Et sa position indique par quelle puissance de 10 il doit être multiplié. Compter en base 10 est une convention, pas du tout un système absolu, qui provient sans doute du fait que nous ayons dix doigts.

Trêve de maths. Parlons plutôt dauphins. Dans son livre «*Code, the Hidden Language of Computer Hardware and Software*» (Code, le langage caché du matériel informatique et des logiciels), le programmeur Charles Petzold, pour expliquer le système binaire, se projette dans un monde parallèle où nous autres, êtres humains, serions des cétacés, et n'aurions de ce fait, que deux nageoires pour compter au lieu de 10 doigts. Là, nous disposons que d'un chiffre avant de passer à la puissance du dessus ! C'est arrivé à 2 que l'on ajoute un chiffre à gauche, qui désigne ici les puissances non pas de 10 mais de 2. Un tout petit paquet, de juste deux unités.

COMPTER EN BASE 2

| | | | |
|-------|------|------|-----|
| 2^1 | 0, | 1, | |
| 2^2 | 10, | 11, | |
| 2^3 | 100, | 101, | 111 |

Convertit, ça donne ça :

| Base 10 | Base 2 | Soleils |
|---------|--------|---------|
| 0 | 0 | |
| 1 | 1 | ☼ |
| 2 | 10 | ☼☼ |
| 3 | 11 | ☼☼☼ |
| 4 | 100 | ☼☼☼☼ |

(Un autre système de tous les jours qui n'est pas en base 10 est celui des heures, qui fonctionne en base 60 : 60 secondes constituent 1 minute et 60 minutes 1 heure). Qu'est ce que ça peut bien changer de compter ainsi ? LE MONDE.

BINAIRE X ÉLECTRICITÉ

«Le système binaire crée un pont entre l'arithmétique et l'électricité. [...] Un fil peut être un chiffre binaire. Si le courant passe dans le fil, le chiffre binaire est 1. Sinon, le chiffre binaire est 0.»⁷

La clé, c'est l'électricité. Les ordinateurs sont des machines qui stockent et manipulent de l'électricité. Ils sont des champs, à perte de vue de 0 et de 1. De courant non courant.

7. «The binary number system bridges the gap between arithmetic and electricity. [...] A wire can be a binary digit. If current is flowing through the wire, the binary digit is 1. If not, the binary digit is 0.» , Charles PETZOLD, *Code, the hidden language of hardware and software*, Microsoft Press, 2000

| Base 2 | Courant |
|--------|-------------|
| 0 | NON |
| 1 | OUI |
| 10 | OUI NON |
| 11 | OUI OUI |
| 100 | OUI NON NON |

On a aussi fait ça avec des cartes perforées...

| Base 2 | Cartes |
|--------|--------|
| 0 | ○ |
| 1 | ● |
| 10 | ● ○ |
| 11 | ● ● |
| 100 | ● ○ ○ |

Voilà ! La pierre des cathédrales informatiques qui peuplent notre monde contemporain. La toute petite pierre : langage binaire : le 0 ou le 1. « Autour de 1948, le mathématicien américain John Wilder Turkey (1915) sentit que les mots « chiffre binaire » (*binary digit*) auraient un grand rôle à jouer dans les années à venir, à mesure que les ordinateurs prenaient de l'importance. Il décida d'inventer un nouveau mot, plus court, pour remplacer les 5 syllabes de *binary digit*. Il pensa à bigit et binit mais se décida finalement pour [...] le terme *bit*. »⁸ Ainsi naquit-il.

8. « Sometime around 1948, the american mathematician John Wilder Turkey (born 1915) realized that the words binary digit were likely to assume a much greater importance in the years ahead as computers become more prevalent. He decided to coin a new, shorter word to replace the wiedly five syllables of binary digit. He considered bigit and binit but settled instead on the [...] word bit. » Charles PETZOLD, *Code, the hidden language of hardware and software*, Microsoft Press, 2000

PORTES LOGIQUES

« Un interrupteur peut être un chiffre binaire. S'il est ouvert (ne laisse pas passer le courant), le chiffre binaire est 0. S'il est fermé (laisse passer le courant), le chiffre binaire est 1. »⁹

| Base 2 | Interrupteur | Lampe |
|--------|---------------------|-------|
| 0 | Ouvert | ○ |
| 1 | Fermé | ● |
| 10 | Fermé Ouvert | ●○ |
| 11 | Fermé Fermé | ●● |
| 100 | Fermé Ouvert Ouvert | ●○○ |

Ces champs de données codées en bit sont reliés par des portes logiques : des interrupteurs qui exécutent les fonctions logiques de base, type OUI, NON, SI. Ils incarnent les chemins que la machine peut suivre pour traiter les données. Ils sont une traduction physique, électrique de la logique et sont capables d'exécuter les opérations mathématiques de base (en fait juste l'addition, à partir de laquelle vient tout le reste ! Un monde construit sur l'addition, c'est quand même fou). Ils constituent ce qu'on appelle l'architecture d'un circuit. Et c'est le programme qui va déterminer comment les utiliser. Quels algorithmes ils vont exécuter. Pour faire court, les ordinateurs modernes traduisent des maths en électricité. Je trouve ça fascinant.

DES NOMBRES...ET POURQUOI PAS DES LETTRES ?

Certes nous pouvons coder des nombres en système binaire. Mais qu'en est-il du texte ? Pour répondre à cette question, les ingénieurs des années 1960 mirent au point un système de correspondance, codé sur 8 bits, avec toutes les lettres de l'alphabet et quelques caractères spéciaux (type ponctuation). Un simple tableau qui dit : si c'est ce chiffre, alors c'est un a minuscule, si c'est celui-là, un A majuscule, celui un d minuscule etc... Le système ASCII.

9. « A switch can be a binary digit. If the switch is on, or closed, the binary digit is 1. If the switch is off, or open, the binary digit is 0. A lightbulb can be a binary digit. If the lightbulb is lit, the binary digit is 1. », Charles PETZOLD, *Code, the hidden language of hardware and software*, Microsoft Press, 2000

Pour nous autres français des années 2017, l'encodage ASCII est le « truc qui fait des accents bizarres » en opposition à UTF8 qui est le « truc qui ne fait pas d'accent bizarres ». Nous savons tous qu'il fait plus chaud en Å©tÅ©. En effet, nos chers ingénieurs étant de langue anglaise, ils n'avaient pas prévu de code pour les étranges caractères européens (sans parler des kanjis...). Un peu plus tard, fut mise au point un second système d'encodage, l'UTF8, cette fois codé sur 16 bits (2 fois plus lourd) mais qui tend vers l'universalité. Néanmoins, il ne permet pas d'encoder entièrement des langues sans alphabet comme le chinois, qui restent un casse-tête infini pour les ingénieurs. Aux français en revanche, il permet enfin d'apprécier tranquillement l'été (et non l'Å©tÅ©).

NIVEAUX D'ABSTRACTION

« Programmer une machine, c'est comme manger avec un cure-dents. Les portions (*bites*) sont si petites et le déroulement si laborieux que le repas dure une éternité. »¹⁰

Les ordinateurs d'aujourd'hui ne comprennent toujours que les 0 et les 1 : c'est ce qu'on appelle le « langage machine » ou *machine code*. On dit que c'est un langage bas : niveau d'abstraction zéro, les instructions que l'on donne sont directement celles exécutées. Mais pour les hommes, écrire des programmes en binaire est un ENFER (tout le monde n'est pas Neo). Donc, pour alléger la tâche de communication, les ingénieurs ont mis au point des interfaces, à mi-chemin entre les deux mondes : les langages de programmation. Ce sont des langages compréhensibles par les hommes mais qui reprennent les fonctions de base exécutables par la machine. Des ponts.

Ils sont dit de « haut niveau » car il ne parle pas directement le binaire machine. Mais quoi, nous venons de dire que la machine ne parlait que le binaire ! (stupide machine). Pour passer du langage de programmation à la machine, on a aussi inventé un traducteur : le *compiler*. C'est un programme qui contient les instructions pour passer du langage de programmation au binaire. D'un haut niveau d'abstraction à un bas niveau d'abstraction. Ces programmes sont extrêmement compliqués et difficiles à écrire, mais ils sont

10. « *Programming a machine is like eating with a toothpick. The bites are so small and the process so laborious that dinner takes for ever.* » Charles PETZOLD, *Code, the hidden language of hardware and software*, Microsoft Press, 2000

essentiels dans le fonctionnement des ordinateurs. Une fois écrits, ils permettent à tous les utilisateurs du dit langage de s'en servir, sans avoir à les réécrire à chaque fois. C'est l'un des grands principes de la programmation : la construction d'un monde à plusieurs mains.

« Ne crois aucun scan que tu n'ai toi-même trafiqué » Algorithmes de compressions

III
A

LA COMPLEXITÉ ALGORITHMIQUE

Une notion très importante lorsque l'on parle d'algorithme est celle de complexité. La complexité algorithmique désigne la longueur du chemin (nombres d'opérations), le temps et l'énergie qu'il faut pour arriver à un résultat voulu. La taille de la recette en somme. On appelle ça le « coût » d'un algorithme.

« Pour une fonction donnée, il peut y avoir plusieurs algorithmes plus ou moins astucieux, et le temps peut varier énormément ; qu'à cela ne tienne, il n'y a qu'à chercher la complexité optimale. »¹¹

« Complexité optimale » désigne le chemin de plus court, donc le moins coûteux, en espace et en temps. En informatique, que ce soit au niveau de la conception de matériels informatiques ou de logiciels, ce principe d'économie, d'optimisation règne en maître absolu. Il constitue LA règle de base de la

11. Jean-Yves GIRARD, *La Machine de Turing*, Points : Paris, 1991

programmation : le plus court, le plus rapide. Le moins cher. L'informatique est une grande course à l'efficacité.

DES ALGORITHMES POUR FAIRE DES ÉCONOMIES

« Transmettre des données coûte du temps, de l'argent et de la mémoire. »¹²

Il existe différents formats d'images. *Jpeg, tiff, gif...* Ce qui les différencie, ce sont leurs algorithmes de compression, c'est-à-dire les algorithmes utilisés pour encoder les données qui décrivent/constituent l'image. Si vous avez régulièrement à faire au traitement d'images, peut-être avez-vous entendu dire que le format *jpeg* n'est pas un bon format car il détériore les fichiers. Cette détérioration est due à son algorithme, qui sacrifie de l'information pour rendre l'image plus légère, pour qu'elle tienne moins de place (moins de bits).

Tous les formats ont des modes de compression différents. Certains, comme le *jpeg* sont dit « destructeurs » car ils enlèvent de l'information. D'autre, comme le *tiff*, compressent différemment, et gardent la totalité des données, qui est restituée lors de la décompression. Le but étant de créer des algorithmes qui font les images les plus légères possibles tout en conservant le maximum d'informations utiles.

NUMÉRISATION

L'histoire racontée ici se base sur la conférence de David Kriesel, *Ne crois aucun scan que tu n'ai toi-même trafiqué* au 31^e Chaos Computer Congress, à Hambourg (DE).

En 2013, un informaticien de Bonn (DE), David Kriesel, reçoit un appel d'un centre de photocopie Xerox, pour lequel il avait travaillé durant ses études. Cet appel lui fait part d'une découverte pour le moins surprenante. Chargés de scanner des plans pour un bureau d'architectes, les employés se sont rendus compte que les chiffres apparaissant sur le scan étaient différents de ceux sur les plans originaux. Un peu paniqués, ils lui demandent s'il aurait

12. « Datenübertragung kostet Zeit, Geld und Speichern. » KRIESEL David, *Traue keinem Scan, den du nicht selbst gefälscht hast*, Chaos Communication Congress : Hamburg, 2.1.2015

une idée, « d'où ça peut venir ». Devant l'insistance de ses anciens collègues, bien qu'il soit dubitatif, il décide d'aller jeter un œil sur place. Après plusieurs essais, sur plusieurs machines, c'est avéré. Les scanners Xerox *changent* les chiffres.

Horrifié, Kriesel décide de pousser les tests. Il scanne d'autres documents, puis en crée d'artificiels avec d'autres tailles de corps et d'autres polices pour voir si les erreurs se généralisent. Après plusieurs scans, dans différentes résolutions, sur différents documents, il devient clair que le bug est endémique.

Il en informe un sous-traitant de Xerox, qui arrive à reproduire le bug. Choqué, il décide d'alerter l'entreprise, qui ne se sent pas vraiment concernée. C'en est trop. Kriesel décide de publier un article sur son blog, qui raconte ces « scanners qui changent les chiffres ». S'en suit graduellement une escalade de partages, jusqu'à ce que son histoire se retrouve dans de grands médias internationaux type CNN ou BBC. Car il faut préciser que Xerox n'est pas n'importe quelle entreprise. En anglais le verbe « photocopier » n'existe pas. On dit *to xerox something*. Aux États-Unis, elle détient un quasi monopole du secteur.

PATTERN MATCHING

Pour aller plus loin dans le gain de place, il est possible d'utiliser un deuxième algorithme pour compresser, personnifié par le format *Jbig2*, un format qui utilise la technique du *pattern matching* ou « reconnaissance de motifs ». Contrairement aux autres qui traitent indifféremment toute l'image, le *pattern matching* fonctionne par « zones » :

1. Zones de données : les blancs ne sont pas codés pixel par pixel, ce qui fait gagner pas mal de place.
2. Zones de « motifs » selon un critère de ressemblance. Par exemple, sur l'image de base, l'algorithme va reconnaître tous les *e* et ensuite ne sauvegarder qu'un seul dessin de *e* en disant simplement à l'ordinateur où ils sont placés. Au lieu de centaines de *e* un tout petit peu différents, sauvegardés chacun comme des images, nous aurons un seul modèle de *e* qui leur sera substitué. Ce qui signifie un gain de place encore plus grand.

Le problème, c'est que cet algorithme fonctionne mal. Et donc, change certains caractères, avec le *pattern matching*, qui remplace les mauvais signes. Ainsi, un 6 sur le document papier peu devenir un 2 sur le document numérique. Comme le dit Kriesel, ces erreurs sont bien plus dangereuses que les scans basse qualité illisibles que produisent les formats de compression traditionnels. En effet, un caractère *illisible* est différent d'un caractère *faux*. Un caractère illisible est une absence d'information alors qu'un caractère faux est une information erronée. *Grosse différence*. Pour remarquer ce type d'erreur, il faut donc être extrêmement attentif, et à moins que le scan ne montre un contre-sens flagrant (comme sur le plan d'architecture), on peut tout à fait passer à côté. L'erreur n'est pas visuellement caractérisée.

NUMÉRISATION DU MONDE

En introduction, Kriesel prend l'exemple de la controverse sur le lieu de naissance de l'ancien président américain Obama. Certains activistes d'extrême droite avançait qu'Obama ne serait pas né sur le sol américain mais au Kenya, ce qui aurait rendu sa présidence illégitime. Pour faire taire ces polémiques, il décide de publier un scan de son certificat de naissance. Mais un scan peut-il vraiment être un gage d'authenticité, si, sans même parler de modification à dessein, la machine elle-même ne transcrit pas fidèlement la réalité ?

Après enquête de la part de Xerox et de multiples péripéties, il est apparu que ce programme de compression bugué était utilisé depuis *8 ans* sur de très nombreux appareils, à travers le monde entier. L'entreprise a trouvé le bug et mise en circulation la mise à jour du logiciel, qui corrigeait le problème. Mais qu'en est-il de tous les documents scannés pendant ces 8 ans ? Si les exemplaires papiers ont été détruits, il est impossible de savoir si la version numérique est bien identique.

LIMITE

Certes, les algorithmes sont formidables pour optimiser la place occupée par les données. Il en existe des dizaines, dont certains opèrent sans perte. Mais ce ne sont *pas* des baguettes magiques. Ils possèdent eux aussi leurs propres limites. Il arrive un moment où il faut, quoi qu'il se arrive, choisir entre poids et qualité. Où la réduction des données influe sur l'information qu'elle transmettent. Où de « conforme » l'image devient « fausse ».

L'avarice a ici des conséquences tel qu'elle a rendu la machine inutile. Ce que l'on demande à un scan, à la base, ce n'est pas d'être léger. C'est de numériser un contenu analogique, c'est-à-dire de rendre lisible par un ordinateur les données présentes à l'état matériel. Le faible poids en un plus et non une prérogative. Un fichier léger avec des données fausses est purement, simplement, totalement inutile. Voir pire, s'il est pris pour vrai. Le problème n'est *pas* lié au média numérique. Le problème est lié à la gourmandise des concepteurs, qui, dans une course à l'économie en ont oublié leur mission : faire en sorte que le texte sur l'écran soit le même que celui du document. L'avarice conduirait-elle, passé une certaine limite, au mensonge ?

Le Scandale de la NSA, ce que peut faire la lutte anti-terroriste aux libertés

IV A

«Je me souviens du net avant qu'il soit surveillé, et il n'y a rien dans l'histoire de l'humanité qui s'en rapproche. Des enfants pouvaient parler d'égal à égal [...] avec des experts à l'autre bout du monde, de n'importe quel sujet, n'importe quand, n'importe où. C'était libre et sans entraves. Ce modèle s'est érodé, s'est refroidi sous nos yeux, pour se changer en un modèle où les gens auto-censurent leurs opinions.»¹³

Edward SNOWDEN, Lanceur d'alerte, dans un entretien tiré du film *CitizenFour*,
Laura POITRAS, 2013

En 2013, Laura Poitras, une journaliste d'investigation américaine, est contactée par un anonyme se présentant comme un employé de CIA, sous le nom de *Citizen Four*. Il lui annonce détenir les preuves d'une surveillance de masse de la population américaine par la National Security Agency (NSA), l'agence de renseignement américaine. Après plusieurs mois d'échanges cryptés, ils conviennent d'une rencontre en juin, à Hong-Kong, qui inclura deux autres journalistes, Glenn Greenwald et Ewen MacAskill (*The Guar-*

13. «*I remember what the Internet was like before it was being watched and there's never anything in the history of men that has been like it. You could again, have children from one part of the world having an equal discussion [...] with experts in a field from another part of the world, on any topic, anywhere, anytime, all of the time. And it was free and unrestrained. And we've seen the chilling of that, the cooling of that, the changing of that model towards something in which people self-police their own views.*»

dian). Au cours de la semaine qui suit, les trois journalistes restant cloîtrés dans un hôtel, à écouter leur source, à savoir Edward Snowden, qui les guide dans les documents qu'il a dérobés, prouvant l'établissement de cette surveillance massive.

« Lancé en 2007, le logiciel PRISM permet à la NSA d'accéder aux données de Google, Facebook, Microsoft, Yahoo [...] pour approximativement 20 millions de dollars par an, une somme dérisoire pour un service de renseignement. »¹⁴

Les trois journalistes publient successivement des articles reprenant ses révélations et étalent sur la place publique les pratiques de l'État américain. Snowden explique que l'agence enregistre toutes les communications qu'il leur est possible d'intercepter, peu importe de qui elles viennent, peu importe par quel canal elles passent, pour constituer d'immenses bases de données qu'il sera ensuite possible de parcourir rétroactivement, à l'aide (notamment) d'algorithmes. Il ajoute qu'il est possible de regarder en direct les faits et gestes des citoyens, ou de créer des alertes sur certaines personnes lorsqu'elles passent une communication. Un genre de « notification *Big Brother* », proprement effrayante.

« Les gouvernements et les juges n'ont pas attendu le temps des algorithmes pour surveiller leurs sujets et citoyens. Au XX^e siècle, par exemple, un juge pouvait ordonner la mise sur écoute d'une personne coupable ou suspecte d'un délit ou d'un crime et les gouvernements se passaient même souvent des juges. Mais, comme il fallait que quelqu'un consacre du temps à écouter ces conversations, cette surveillance était limitée. »¹⁵ Si la concentration du savoir/pouvoir est accessible aux milliardaires, par le biais d'entreprises tel que *Cambridge Analytica* qui vendent l'influence au plus offrant, elle l'est également aux États, qui entendent, tout comme eux, assurer par ce biais, leur auto-sauvegarde. Poitras explique que ce dispositif d'espionnage a été mis en place à la suite des attentats du 11 septembre, comme mesure anti-terroriste. Entre liberté et sécurité, les gouvernements démocratiques

14. « *Launched in 2007, the PRISM program allows the NSA to access data from Google, Facebook, Microsoft, Yahoo [...] for a mere \$20 million per year, a trifling sum for an intelligence program.* » Edward SNOWDEN, lanceur d'alerte, dans un entretien tiré du film *CitizenFour*, Laura POITRAS, 2013

15. Serge ABITEBOUL, Gilles DOWEK, *Le Temps des algorithmes*, Paris : Le Pommier, 2017

occidentaux prennent de manière croissante le chemin la sécurité. Rappelons que toutes ces collectes de données sur les citoyens américains (différent pour les étrangers) sont effectuées sans aval de la justice et donc, illégales.

« [La collecte passive sur des réseaux mondiaux] est à la fois interne et externe aux États-Unis. Il y a beaucoup de manières différentes de faire ça, mais les partenariats avec les entreprises sont l'une des premières sources domestiques [de la NSA]. Ils le font aussi avec des multinationales, qui peuvent être basées aux États-Unis [...] qu'ils peuvent payer pour un accès [aux données]. »¹⁶

Le problème de la collectes de données n'est donc pas le seulement lié aux entreprises qui y ont accès de par les services qu'elles proposent. L'inaction des États et leurs réticences à prendre des mesures radicales contre elles pour la protection des citoyens prend aussi racines dans l'utilisation qu'eux-mêmes en font. Couper l'accès aux données des entreprises signifieraient également coupé le leur. « "Grâce" aux ordinateurs et aux algorithmes d'apprentissage, les États peuvent désormais surveiller massivement toute la population. Les ordinateurs des services de renseignement réalisent des analyses statistiques sur les montagnes de données qu'ils collectent, afin d'y détecter des comportements suspects : le nom, le prénom, l'adresse, les déplacements, les contacts d'un individu, les films qu'il regarde, la musique qu'il écoute, etc., sont autant d'éléments qui qualifient cet individu et qui feront dire aux algorithmes qu'il est suspect ou non, sur des bases purement statistiques. »¹⁷

Ce type d'algorithme illustre très bien l'idée de boîte noire, qui impute une culpabilité potentielle à des sujets sur la base de critères confidentiels, et semble bien loin de l'adage « innocent jusqu'à preuve du contraire ». Abiteboul et Doweck ajoute que « Plus la surveillance est massive, plus le risque de faux positifs augmente », ce qui signifie que quand bien même ces méthodes seraient efficaces, elles attraperaient au passage des innocents. Nous en revenons à l'idée développée par Cathy O'Neil, qui soutient que ces machines

16. « [World wide passive collection on networks] is both domestic to the US and international. There is a lot of different ways they do it [] but corporate partnerships or one of the primary things they do domestically. They also do this with multinationals that might be headquartered in the US and [...] that they can pay to get the access. » Edward SNOWDEN, Lanceur d'alerte, dans un entretien tiré du film *CitizenFour*, Laura POITRAS, 2013

17. Serge ABITEBOUL, Gilles DOWECK, *Le Temps des algorithmes*, Paris : Le Pommier, 2017

postulent, au contraire de la loi, qu'un coupable dehors est plus dangereux pour la société qu'un innocent en prison.

La sécurité de l'État n'étant pas forcément celle des citoyens, une telle surveillance, si elle prend racine dans la lutte contre le terrorisme, peut vite dégénérer en censure de l'opposition politique. Les lois tel que le Patriot Act, adoptée en 2001 aux États-Unis à la suite du 11 septembre, ou la loi relative au renseignement de 2015 en France mettent en danger la liberté des individus. « Il est de notre responsabilité collective de poser la question du bien-fondé d'une telle surveillance, de lui imposer des limites, de la placer sous le contrôle d'une justice indépendante. »¹⁸

18. Serge ABITEBOUL, Gilles DOWEK, *Le Temps des algorithmes*, Paris : Le Pommier, 2017

Erstazs d'extraordinaire : algorithmes de recommandation

V
A

La mission d'un algorithme de recommandation est de vous conseiller un produit, qu'il soit physique comme sur les sites d'achat en ligne type Amazon ou immatériel, comme sur les sites de streaming type Netflix ou Youtube. Pour ce faire, il se base sur deux paramètres. Premièrement, votre historique, c'est-à-dire ce que vous avez précédemment aimé. Deuxièmement, l'historique des autres utilisateurs, qui ont aimé des contenus similaires à vous. L'algorithme vous conseille donc selon les choix de vos émules culturels. Tout comme PageRank, il se base sur la popularité. La différence réside dans le choix des utilisateurs : alors qu'un moteur de recherche prend en compte tout le monde, un algorithme de recommandation s'intéresse seulement aux personnes qui, dans leur comportement d'achat ou de streaming, vous ressemblent. Cette méthode se base sur l'idée que les similarités trahissent des goûts communs. Le but de cet algorithme étant, que vous consommiez le plus possible, des biens matériels ou pas.

Comme l'explique Dominique Cardon, ce type d'algorithme calqué sur le comportement perpétue les inégalités culturelles.¹⁹ Dans le cas de YouTube, si j'écoute de la musique variée, l'algorithme me proposera de nouveaux morceaux variés. Si je suis plus du genre TOP 50, je serais conseillé en conséquence. Il perpétue l'enfermement, la bulle. Étant incapable d'implémenter autre chose que le passé, il nous pousse à la répétition et fait fit de l'évidence qu'on peut aimer une œuvre qui était pourtant, une minute auparavant, complètement étrangère à nos goûts.

BOULVERSEMENTS DE SUPERMARCHÉS

Pour nourrir son algorithme de recommandation, Netflix rémunèrent des *taggers* qui ont pour mission de qualifier en des termes compréhensibles par la machine les contenus culturels que la plateforme de streaming commercialise. Un autre exemple de *Digital Labor*, assimilables aux turkers d'Amazon, dissimulés sous l'aspect prétendument informatique / magique du « Vous pourriez aussi aimer ». Il est étrange de constater à quel point le format numérique a conduit à des classements, des avis, des notes sur tout contenu culturel quel qu'il soit, dans la logique du cadre précédemment évoqué. Les sites comme Netflix et ses milliards de tags ou Sens Critique et ses milliers de listes illustrent bien cette tendance à l'extrême rangement. Catégories, époque, sensibilités, tout y est répertorié, de manière à faciliter vos choix et vous guider vers un contenu « taillé sur mesure ». « Comme le dit un représentant de l'entreprise, "il y a 33 millions de versions différentes de Netflix", ou un système adapté à chaque client. »²⁰

Chacun a déjà fait l'expérience d'œuvres bouleversantes, qui résonnent longtemps à l'intérieur de lui, parfois toute une vie. Sont-elles pour autant des déclinaisons d'un même thème, d'un même pays, d'une même époque comme le sont les catégories de contenus ? Être guidé par des catégories a toujours un petit côté absurde, car elles sont dans le fond, assez inefficaces dans leurs tentatives d'automatiser cet extraordinaire, qui par définition, ne

19. Dominique CARDON, *À Quoi rêvent les algorithmes ? Nos vies à l'heure des big data*, Paris : Seuil, 2015

20. « As one company representative put it "there are 33 millions different versions of Netflix" or a uniquely tailored system for each individual customer. » Ed FINN, *What Algorithms Want*, Cambridge : MIT Press, 2017

se laisse que rarement approcher. Elles ne disent souvent rien de l'objet qu'elles tentent de décrire. Ce qui est assez drôle, c'est qu'alors qu'elles répondent à une mission de rangement de l'espace culturel, elles créent autant d'incongru que de clarté. Mettre *Le Château*²¹ et *Le Pavillon d'Or*²² dans **#roman #architecture** par exemple, n'a absolument aucun sens. Elles créent leur propre hasard, qui se déguise en ordre.

COMPLEXES

L'algorithme *AutoPlay* de Youtube, qui crée automatiquement une playlist d'après les visionnages précédents des utilisateurs, a favorisé, pour une raison inconnue, certains morceaux de manière spectaculaire. Au nombre des exemples, on trouve l'album *Through The Looking Glass* de l'artiste japonaise Midori Takada, sorti en 1983, puis tombé presque immédiatement dans l'oubli à l'époque. Sans motif aucun, l'algorithme le proposait automatiquement aux utilisateurs intéressés par la musique électronique d'ambiance, ce qui a contribué à le propulser sur le devant de la scène, lui conférant des millions de vues.²³

Interrogé sur une bizarrerie similaire observée avec l'algorithme de Netflix, qui favorisait la série *Perry Mason* (1957), leur vice-président produit, Todd Yellin explique qu'en « ajoutant de la complexité au monde de la machine, vous rajoutez une serendipité imprévisible. *Perry Mason* va en sortir. Ces fantômes dans la machine seront toujours un sous-produit de la complexité. Et parfois nous appelons ça un bug, parfois nous appelons ça une caractéristique. »²⁴ Traduction : nous ne savons pas non plus pourquoi l'algorithme propose cette série. Les architectes des machines sont eux-mêmes noyés dans la complexité qu'ils ont créée. Lorsque les algorithmes proposent

21. Franz KAFKA, *Le Château* traduit de l'allemand par Bernard LORTHOLARY, Flammarion : Paris, 1926

22. Yukio MISHIMA, *Le Pavillon d'Or* traduit du japonais par Marc MÉCRÉANT, Gallimard : Paris, 1956

23. Martyn PEPPERELL, *How YouTube autoplay gave a lost Japanese classic new life, Dazed and Confused*, 3.4.2017

24. « The more complexity you add to a machine world, you're adding serendipity that you couldn't imagine. *Perry Mason* is going to happen. These ghosts in the machine are always going to be a by-product of the complexity. And sometimes we call it a bug, and sometimes we call it a feature. » Ed FINN, *What Algorithms Want*, Cambridge : MIT Press, 2017

des séries, l'effet de ces coquilles, sorties de nulle part, est moindre, voir même, peut s'avérer intéressant. Dans un algorithme comme *PredPol* en revanche, leur présence, si avérée, serait bien plus problématique. Quoi qu'il en soit, il reste très préoccupant que les concepteurs d'algorithmes créent et utilisent des machines dont ils admettent eux-mêmes ne pas comprendre le fonctionnement.

Ce que dit la loi européenne

VI A

Du côté de l'Europe, nous sommes un peu mieux protégés que nos confrères des États-Unis vis-à-vis des entreprises. Vis-à-vis des États, chaque pays applique sa loi, et la liberté des citoyens en France a récemment été questionnée par la loi sur le renseignement, à la suite des attentats.

Une loi, la GDPR (General Data Protection Regulation), votée le 27 avril 2016 par les institutions européennes entrera en vigueur le 25 mai 2018. Elle renforcera les droits des citoyens vis-à-vis d'organisations privées ou publiques concernant l'utilisation de leurs données personnelles. Elle comprend plusieurs mesures :

- Le droit facilité pour les personnes d'exiger l'obtention des données les concernant récoltées par un groupe (qui existait déjà mais est renforcé), gratuitement et dans un délai de un mois.
- Le droit de faire supprimer des données par l'utilisateur dans certaines circonstances, si elles ne sont plus nécessaires au groupe, si l'intérêt n'est pas légitimé, s'il retire son consentement ou si elles furent illégalement collectées.
- La nécessité pour les organisations d'obtenir le consentement clair des utilisateurs quant à la collecte de leurs données
- Une obligation de transparence quant à la manière dont ces données sont utilisées. Les entreprises de plus de 250 employés doivent justifier la conservation des données et faire état d'un dispositif de sécurité suffisant, qui garantissent leur usage exclusif.

- La nécessité de signaler aux citoyens si des données les concernant ont été dérobées
- La mise en place d'amendes si la loi n'est pas respectée, pouvant aller jusqu'à 20 millions d'euros

Il ne reste plus qu'à la faire appliquer... **#sameLawForAll**

Lexique

Implémentation

« Un algorithme est une abstraction, qui possède une existence indépendante de ce que les spécialistes de l'informatique aime à nommer les " détails d'implémentation ", c'est à dire, son incarnation dans un langage de programmation spécifique pour une architecture de machine spécifique [...] »¹

L'implémentation est l'étape qui consiste à adapter un algorithme au contexte dans lequel il va être utilisé, c'est-à-dire au langage de programmation dans lequel il sera écrit, ainsi qu'à la machine qui l'exécutera. C'est le passage de l'abstrait au concret, qui lui permet de fonctionner au mieux dans une situation précise. Chaque langage, chaque machine est régit par des contraintes différentes, ce qui nécessite de repenser l'algorithme pour les intégrer. C'est l'adaptation de la recette aux ingrédients disponibles.

Langage de programmation

Un ordinateur ne comprend que les 0 et les 1. Pour lui dicter des actions, il faut donc les décrire ainsi. C'est ce qu'on appelle le « *machine code* ». Mais pour les êtres humains, parler ce langage machine est un enfer. Pour se faciliter la tâche, ils ont donc inventé une nouvelle manière de communiquer avec elle, à mi-chemin entre eux deux : les langages de programmation. Ils

1. « *An algorithm is an abstraction, having an autonomous existence independent of what computer scientists like to refer to as "implementation details," that is, its embodiment in a particular programming language for a particular machine architecture [...]* » Andrew GOFFEY, Algorithm, *Software Studies, a Lexicon*, publié par Matthew FULLER, Cambridge : MIT Press, 2008

se composent d'un ensemble de règles syntaxiques, qui décrivent des actions basiques réalisables par l'ordinateur. Ils permettent d'écrire des programmes en des termes compréhensibles par les humains. Il en existe des centaines, qui répondent tous à des besoins différents. Si la machine comprend leurs instructions, c'est grâce à un autre programme qui « traduit » les instructions en 0 et en 1 : le *compilier*.

Loi de Moore

En 1975, Gordon Moore, créateur d'Intel (une célèbre entreprise américaine fabriquant des micro-processeurs) prophétisa que le nombre de transistors par unité de surface dans le matériel informatique allait doubler tous les deux ans. C'est ce qu'on appelle la « Loi de Moore ». Celle-ci fut parfaitement respectée jusque récemment, décrivant fidèlement les progrès de l'industrie du *hardware*, qui permirent l'augmentation exponentielle de la puissance de calcul des machines. Aujourd'hui, elle tend à devenir fausse, limitée par les propriétés physiques de la matière.

Machine Learning

Le terme *machine learning* ou « apprentissage automatique » désigne une branche de la recherche en intelligence artificielle, qui fonctionne grâce à un type d'algorithmes particulier, capables « d'apprendre ». L'université de Stanford le définit comme « la science de faire agir les ordinateurs sans qu'il aient été explicitement programmés pour le faire ».² En pratique, cela se traduit par l'utilisation de montagnes de données, avec lesquelles les data-scientists « nourrissent » les algorithmes, afin qu'ils soient capables, par exemple, de reconnaître ce que représente une image. La méthode utilisée couramment et qui porte ses fruits est dite de « l'apprentissage supervisé ». Pour la reconnaissance d'objets, on rentre une série d'images ainsi que ce qu'elles représentent dans la base de donnée. À partir de ce corpus, la machine va être capable d'en tirer des invariants, des motifs, pour reconnaître ensuite d'autres images représentant les mêmes éléments, pour ensuite les ajouter aussi à sa base, et ainsi affiner sa reconnaissance. On parle de phase « d'entraînement » de l'algorithme. L'apprentissage est dit supervisé car il faut donné des informations de base à la machine pour qu'elle soit ensuite capable d'en tirer des conclusions.

2. Lee BELL, *Machine learning versus AI: what's the difference?*, Wired UK, 1.12.16

Apprentissage non-supervisé

Les êtres humains et les animaux n'ont pas besoin de base pour tirer des conclusions quant au monde qui les entoure. Ils apprennent d'une manière dite « non supervisée ». Pour l'instant, le processus derrière cet apprentissage empirique reste assez mystérieux et les tentatives de *machine learning* sans guide préalable se sont soldées par des échecs.

Le conte des deux intelligences artificielles

Pour bien comprendre ces notions, nous pouvons faire un détour par l'histoire de l'IA, qui longtemps opposa ces deux approches. La première, dite « reproductrice », cherche, non pas à copier le processus d'apprentissage humain mais ses résultats, en adoptant une approche programmatrice. On parle d'intelligence artificielle faible (*light AI*). L'apprentissage supervisé appartient à cette catégorie. La seconde, dite productrice, prend le parti inverse et se concentre sur le processus, selon une approche cognitive. On parle d'intelligence artificielle forte (*strong AI*). D'un côté, on cherche à mimer les résultats de l'intelligence (émulation) et de l'autre, son processus (simulation). Étant donné que l'intelligence humaine reste assez obscure (nous sommes à ce jour toujours en attente d'une définition satisfaisante), et surtout extrêmement plurielle, l'approche simulatrice de l'intelligence forte n'a pour l'instant pas porté de bons résultats. Mais comme le rappelle Luciano Floridi, l'IA est une « science artificielle », et tant que telle, n'est attachée à aucune des deux approches en particulier.³

Modèle

« [Les modèles mathématiques] font toujours des erreurs, car ils sont, par nature, des simplifications. Aucun modèle ne peut inclure toute la complexité du monde réel ou la nuance de la communication humaine. »⁴

3. Luciano FLORIDI, *The Fourth Revolution: How the Infosphere is Reshaping Human Reality*, Oxford: Oxford University Press, 2016

4. « [Mathematical models] would always be mistaken, however, because they are, by nature, simplification. No model can include all of the real world complexity or the nuance of human communication. » Cathy O'NEIL, *Weapons of math destruction: how big data increases inequalities and threatens democracy*, New York: Crown, 2016

Un modèle est une représentation abstraite d'un phénomène, une manière de synthétiser la réalité, qui permet de l'étudier et plus largement de tenter de la prévoir. Ils sont une architecture de causes, une recette qui permet de comprendre un phénomène. Les plus familiers sont les modèles météorologiques à partir desquels sont effectuées les prévisions. Dans le cas du temps, il s'agit d'algorithmes très complexes qui analysent des données météo en temps réel et prophétisent de possibles évolutions. Ils sont construits d'après des centaines de paramètres, qui allient théories physiques, probabilités et données historiques. Le rôle du hasard (ou de l'ignorance, suivant si vous considérez le hasard comme absolu ou comme le seul le résultat d'une méconnaissance humaine) dans le comportement des éléments rend la modélisation difficile, faisant baisser la validité des prévisions au fur et à mesure du temps.

Cette notion ne se limite pas au seul champ de la science naturelle. Comme l'explique Cathy O'Neil, qu'ils soient programmables sur un ordinateur ou non, nous avons tous des modèles en tête que nous utilisons au quotidien pour définir la réalité et cerner l'avenir. Par exemple, pour choisir nos vêtements. Différent pour chacun, nous obéissons tous à un modèle informel qui comprend des dizaines paramètres personnels, qui peuvent être la température, la saison, l'image de soi, l'âge, la taille, la situation géographique, la mode, le fait d'être ou non en prison etc... Un bon modèle doit maintenir «un aller retour constant entre ce qu'il essaie de comprendre ou prévoir et la réalité». Car «les conditions changent, donc le modèle aussi.»⁵ Mon modèle de vêtements à 15 ans ne sera pas le même qu'à 50. Ma passion démesurée pour les baskets rose flash aura probablement disparue entre temps.

Ordinateur

Un ordinateur est :

1. Une machine qui peut être programmée. Ce qui signifie qu'elle est capable de lire des programmes, ce qui induit, de prendre des décisions sans intervention humaine manuelle, selon ce qu'ils spécifient.

5. «*They would always be mistaken, however, because models are, by nature, simplification. No model can include all of the real world complexity or the nuance of human communication.*» Cathy O'NEIL, *Weapons of math destruction: how big data increases inequalities and threatens democracy*, New York : Crown, 2016

2. Une machine qui ne différencie pas conceptuellement programme et données. Ils sont tous deux codés sous la même forme (0 et 1), et lisibles indifféremment par la machine.

Programme

Les termes de « programme » et « d'algorithme » englobent chacun des significations extrêmement proches, qui se distinguent par le contexte. Un algorithme est une recette générique, le « mécanisme conceptuel de calcul systématique » alors que le programme se définit comme « l'écriture précise de l'algorithme dans des langages appropriés [langages dit de programmation] » qui communiquent avec une machine en vue d'effectuer une action. Le programme est donc un algorithme mais pas l'inverse.

On doit le premier programme à une femme, la britannique Ada Lovelace (1815-1852), qui travailla conjointement avec le mathématicien Charles Babbage sur la machine analytique.¹ Babbage la chargea de traduire un texte qui décrivait son fonctionnement vers le français, pour une publication dans un journal suisse. Ce faisant, elle l'annota personnellement, ajoutant des explications. L'un de ses appendices contient un algorithme détaillé, permettant de calculer des nombres de Bernoulli à l'aide de la machine. Rétrospectivement, il fut désigné comme le premier programme informatique.

Proxy

Un proxy est une donnée qui va servir à traduire la réalité en termes calculables par un algorithme. Le *WorldWideWeb* en est rempli. On peut citer simplement le bouton *like* sur Facebook qui est un proxy pour l'approbation, ou la fréquence à laquelle vous interagissez avec une personne qui peut être un proxy pour le concept d'"amitié". Une donnée qui représente un concept abstrait, une sorte de député numérique du réel. Le proxy est symbolique de la différence entre donnée et information. Faire sens avec une donnée, c'est la transformer en information.

Bibliographie

Ⓐ Article

Ⓒ Conférence

Ⓓ Documentaire

Ⓕ Film

Ⓖ Livre

Ⓕ Série

○ Cultes

Algorithmes

- (L) ABITEBOUL Serge, DOWEK Gilles, *Le Temps des algorithmes*, Paris : Le Pommier, 2017
- (D) BENOÎT Fabien, *Les Algorithmes au pouvoir*, Dig It!, Arte, 7.6.2017 [<http://creative.arte.tv/fr/episode/dig-it-les-algorithmes-au-pouvoir>] — Vu : juillet 2017
- (C) BERRY Gérard, *Les Algorithmes, cœur de l'informatique*, Collège de France : Paris FR, 25.1.2008 [<https://www.college-de-france.fr/site/gerard-berry/course-2008-01-25-10h30.html>] — Vu : avril 2017
- (L) CARDON Dominique, *À Quoi rêvent les algorithmes ? Nos vies à l'heure des big data*, Paris : Seuil, 2015
- (L) FINN Ed, *What Algorithms Want*, Cambridge : MIT Press, 2017
- (L) O'NEIL Cathy, *Weapons of math destruction: how big data increases inequalities and threatens democracy*, Crown : New York, 2016
- (A) ROUVROY Antoinette, Des Données et des hommes. Droits et libertés fondamentaux dans un monde de données massives, Académia [academia.edu], 2016 [http://www.academia.edu/15730565/DES_DONN%C3%89ES_ET_DES_HOMMES_DROITS_ET_LIBERT%C3%89S_FONDAMENTAUX_DANS_UN_MONDE_DE_DONN%C3%89ES_MASSIVE_S_version_augment%C3%A9e_]
- (A) ROUVROY Antoinette, La Vie n'est pas donnée, Académia [academia.edu], 2016 [http://www.academia.edu/31846143/La_vie_nest_pas_donn%C3%A9e]
- (A) ROUVROY Antoinette, Les Algorithmes, entre tyrannie du réel et exigence de justice, Académia [academia.edu], 2017 [https://www.academia.edu/20138015/Les_algorithmes_entre_tyrannie_du_r%C3%A9el_et_exigences_de_justice]
- (A) ROUVROY Antoinette, Le Régime de vérité numérique, débat entre Antoinette Rouvroy et Bernard Stiegler, Académia [academia.edu] 1.3.2015, [https://www.academia.edu/11937915/Le_r%C3%A9gime_de_v%C3%A9rit%C3%A9_num%C3%A9rique_Antoinette_Rouvroy_et_Bernard_Stiegler_article]
- (C) SALVIN Kevin, *How algorithms shape our world*, TEDGlobal : Édimbourg UK, 7.2011 [https://www.ted.com/talks/kevin_slavin_how_algorithms_shape_our_world#t-10828] — Vu : avril 2017

Big data

- (A) ANDERSON Chris, The End of Theory: The Data Deluge Makes the Scientific Method Obsolete, *Wired*, 23.6.2008 [<https://www.wired.com/2008/06/pb-theory/>] — Lu : avril 2017
- (A) GRAHAM Mark, Big data and the end of theory? *The Guardian*, 9.3.2016 [<https://www.theguardian.com/news/datablog/2012/mar/09/big-data-theory>] — Lu : décembre 2017
- (A) MATHHEWS Robert, Storks deliver Babies, *robertmatthews.org*, 6.2000 [<http://robertmatthews.org/wp-content/uploads/2016/03/RM-storks-paper.pdf>] — Lu : novembre 2017

Fake news

- (A) ANGWIN Julia, VARNER Madeleine, TOBIN Ariana, Facebook Enabled Advertisers to Reach ‘Jew Haters’, *ProPublica*, 14.9.2017 [<https://www.propublica.org/article/facebook-enabled-advertisers-to-reach-jew-haters>] — Lu : novembre 2017
- (A) DOJCINOVIC Stevan, Hey, Mark Zuckerberg: My Democracy Isn’t Your Laboratory, *New York Times*, 15.11.2017 [<https://www.nytimes.com/2017/11/15/opinion/serbia-facebook-explore-feed.html>] — Lu : novembre 2017
- (A) FINLEY Klint, Google Wants to Solve the Fake News Problem, *Wired*, 26.4.2017 [<http://www.wired.com/2017/04/google-wants-crowd-solve-fake-fact-problem/>] — Lu : novembre 2017
- (A) GARDINER Bryan, You’ll Be Outraged at How Easy It Was to Get You to Click on This Headline, *Wired*, 18.12.2015 [<https://www.wired.com/2015/12/psychology-of-clickbait/>] — Lu : avril 2017
- (A) HERN Alex, Facebook moving non-promoted posts out of news feed in trial, *The Guardian*, 23.10.2017 [<https://www.theguardian.com/technology/2017/oct/23/facebook-non-promoted-posts-news-feed-new-trial-publishers>] — Lu : novembre 2017
- (A) LAPOWSKY Issie, To Fix Its Toxic Ad Problem, Facebook Must Break Itself, *Wired*, 18.9.2017 [<https://www.wired.com/story/to-fix-its-toxic-ad-problem-facebook-must-break-itself/>] — Lu : novembre 2017
- (A) LORDON Frédéric, Politique post-vérité ou journalisme post-politique ?, *Le Monde Diplomatique*, 22.11.2016 [<http://blog.mondediplo.net/2016-11-22-Politique-post-verite-ou-journalisme-post>] — Lu : avril 2017
- (A) MITCHELL Amy, GOTTFRIED Jeffrey, KATERINA Eva Matsa, Millennials and Political News: Social Media – the Local TV for the Next Generation?, *Le Monde Diplomatique*, 1.6.2015 [<http://www.journalism.org/2015/06/01/millennials-political-news/>] — Lu : avril 2017
- (A) OREMUS Will, Facebook Is Telling News Outlets How to Do Their Job, *Slate*, 24.10.2017 [http://www.slate.com/articles/technology/technology/2017/10/facebook_is_telling_news_outlets_how_to_do_their_job.html] — Lu : novembre 2017
- (A) SUBRAMANIAN Samanth, The Macedonians Teens who Mastered Fake News, *Wired*, 15.2.2017 [<https://www.wired.com/2017/02/veles-macedonia-fake-news/>] — Lu : novembre 2017
- (C) TUFEKCI Zeynep, We’re building a dystopia just to make people click on ads, TEDGlobal: New York US, 9.2017 [https://www.ted.com/talks/zeynep_tufekci_we_re_building_a_dystopia_just_to_make_people_click_on_ads] — Vu : novembre 2017
- (A) TYNAN Dan, How Facebook powers money machines for obscure political ‘news’ sites, *The Guardian*, 24.8.2016 [<https://www.theguardian.com/technology/2016/aug/24/facebook-clickbait-political-news-sites-us-election-trump>] — Lu : novembre 2017

Filter bubbles

- (A) DREYFUSS Emily, Coders Think They Can Burst Your Filter Bubble With Tech, *Wired*, 19.11.16 [<https://www.wired.com/2016/11/coders-think-can-burst-filter-bubble-tech/>] – Lu : mars 2017
- (A) EL-BERMAWY Mostafa M., Your Echo Chamber is Destroying Democracy, *Wired*, 18.11.16 [<https://www.wired.com/2016/11/filter-bubble-destroying-democracy/>] – Lu : mars 2017
- (A) HEMPLE Jessie, Eli Pariser Predicted the Future. Now He Can't Escape It, *Wired*, 24.5.17 [<https://www.wired.com/2017/05/eli-pariser-predicted-the-future-now-he-cant-escape-it/>] – Lu : novembre 2017
- (E) LA PORTE Xavier, *Dimanche soir, ma bulle (de filtres) Facebook a éclaté*, *La vie numérique*, France Culture 27.04.17 [<https://www.franceculture.fr/emissions/la-vie-numerique/dimanche-soir-ma-bulle-de-filtres-sur-facebook-eclate>] – Lu : avril 2017
- (L) PARISER Eli, *The Filter Bubble: What is the Internet Hiding from You*, Penguin: Londres, 2011

Machine learning

- (A) BELL Lee, Machine learning versus AI: what's the difference?, *Wired*, 1.12.16 [<http://www.wired.co.uk/article/machine-learning-ai-explained>] – Lu : novembre 2017
- (A) BURGESS Matt, Just like humans, artificial intelligence can be sexist and racist, *Wired UK*, 13.04.17 [<http://www.wired.co.uk/article/machine-learning-bias-prejudice>] – Lu : novembre 2017
- (C) CALISKAN-ISLAM Aylin, BRYSON Joanna J., NARAYANAN Arvind, *Semantics derived automatically from language corpora necessarily contain human biases*, Princeton University, 30.8.2016, [http://www.princeton.edu/~aylinc/papers/caliskan-islam_semantics.pdf] – Lu : novembre 2017
- (C) CALISKAN-ISLAM Aylin, A Story of Discrimination and Unfairness, Chaos Communication Congress: Hamburg DE, 27.12.16, [https://media.ccc.de/v/33c3-802-6-a_story_of_discrimination_and_unfairness#video&t=1527] – Vu : avril 2017

PageRank

- (C) EKSTRÖM Andreas, *The moral bias behind your search results*, TEDxOslo: Oslo, 01.2015, [https://www.ted.com/talks/andreas_ekstrom_the_moral_bias_behind_your_search_results#t-10646] – Vu : avril 2017
- (A) ERTZSCHEID Olivier, *La Recherche dans l'âge de l'assistance*, *affordance.typepad.com*, 30.3.2017 [http://affordance.typepad.com/mon_weblog/2017/03/assistance-ou-assistance.html] – Lu : avril 2017

- Ⓐ KAPLAN Frédéric, Vers un capitalisme linguistique – Quand les mots valent de l'or, *Le Monde diplomatique*, 01.11.11, [<https://www.monde-diplomatique.fr/2011/11/KAPLAN/46925>]
- Ⓐ LUCCHESI Vincent, Ma vie sans Google, *Usbek & Ricca*, n°20 – Nov/Déc 2017

Prédiction

- Ⓐ ANGWIN Julia, LARSON Jeff, MATTU Surya, KIRCHNER Lauren, Machine Bias, *ProPublica*, 23.05.2016 [<https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>] – Lu : décembre 2017
- Ⓐ BENSLIMANE Ismaël, Predpol : prédire des crimes ou des banalités ?, *Cortecs*, 10.12.2014, [<https://cortecs.org/mathematiques/predpol-predire-des-crimes-ou-des-banalites/>] – Lu : décembre 2017
- Ⓐ DUCHATELLE Victoria, Peut-on hacker son auto-prophétie ?, *Cnil*, 28.03.17 [<https://lin.cnil.fr/fr/peut-hacker-son-auto-prophetie>] – Lu : avril 2017
- Ⓔ [Réalisateur inconnu], *Police prédictive*, *Futurmag*, Arte, 19.11.2016 [<http://www.arte.tv/fr/videos/060180-002-A/futurmag>] – Vu : juillet 2017
- Ⓐ GUILLAUD Hubert, Police prédictive : la prédiction des banalités, *Blog internet du Monde*, 27.06.2015, [<http://internetactu.blog.lemonde.fr/2015/06/27/police-predictive-la-prediction-des-banalites/>] – Lu : novembre 2017
- Ⓐ POLLONI Camille, Police prédictive : la tentation de « dire quel sera le crime de demain », *Rue 89*, 27.05.15 [<https://tempsreel.nouvelobs.com/rue89/rue89-police-justice/20150527.RUE9213/police-predictive-la-tentation-de-dire-quel-sera-le-crime-de-demain.html>] – Lu : novembre 2017
- Ⓕ SPIELBERG Steven, *Minority Report*, 2002 – Revu : novembre 2017
- Ⓒ URBAN Karl, *Berechnete Welt*, Chaos Communication Congress : Hambourg, 29.12.2016 [https://media.ccc.de/v/33c3-7899-berechnete_welt#video&t=545] – Vu : juillet 2017

Recommandation

- Ⓐ PEPPERELL Martyn, How YouTube autoplay gave a lost Japanese classic new life, *Dazed and Confused*, 3.4.2017 [<http://www.dazeddigital.com/music/article/35465/1/midori-takada-through-the-looking-glass-interview>] – Lu : avril 2017
- Ⓐ PIERCE David, The Secret Hit Making of the Spotify Playlist, *Wired*, 5.3.2017 [<https://www.wired.com/2017/05/secret-hit-making-power-spotify-playlist/>] – Lu : avril 2017

Général

- Ⓐ ARENDT Hannah, *Condition de l'homme moderne*, traduit de l'anglais par George Fradier, Paris : Calmann-Lévy, 1961

- Ⓕ EKELAND Ivar, *Au hasard : La chance, la science et le monde*, Paris : Seuil, 1991
- Ⓕ DEBRAY Régis, *Cours de médiologie générale*, Paris : Gallimard, 1991

Informatique

- Ⓓ AMOUROUX Vincent, *Mémoire numérique : médias de sauvegarde futurs*, Arte, 01.01.17 [[<http://www.arte.tv/de/videos/050717-000-A/unser-digitales-gedachtnis>] — Vu : avril 2017
- Ⓕ BERRY Gérard, *L'hyperpuissance de l'informatique*, Paris : Odile Jacob, 2017
- Ⓒ BERRY Gérard, *Pourquoi et comment le monde devient numérique (leçon inaugurale)*, Collège de France : Paris, 6.7.2008 [<https://www.franceculture.fr/emissions/college-de-france-40-lecons-inaugurales/gerard-berry-pourquoi-et-comment-le-monde-devient>] — Vu : mars 2017
- Ⓕ DYSON George, *Turing's Cathedral: The Origins of the Digital Universe*, Penguin Random House : Londres, 2012
- Ⓐ EXTANCE Andy, How DNA could store all the world's data, *Nature*, 31.8.2016 [<https://www.nature.com/news/how-dna-could-store-all-the-world-s-data-1.20496>] — Lu : décembre 2017
- Ⓕ FIORIDI Luciano, *The Fourth Revolution: How the Infosphere is Reshaping Human Reality*, Oxford University Press : Oxford, 2016
- Ⓕ FULLER Mathew, *Software Studies, a Lexicon*, MIT Press : Cambridge, 2008 : Algorithm GOFFEY Andrew
- Ⓕ GANASCIA Jean-Gabriel, *Le mythe de la Singularité - Faut-il craindre l'intelligence artificielle ?*, Seuil : Paris, 2017
- Ⓕ GIRARD Jean-Yves, TURING Alan, *La Machine de Turing*, Points : Paris, 1991
- Ⓒ GRAHAM-CUMMING John, *The greatest machine that never was*, TEDxImperialCollege : Londres, 03.2012 — Vu : novembre 2017
- Ⓒ KRIESEL David, *Traue keinem Scan, den du nicht selbst gefälscht hast*, Chaos Communication Congress : Hamburg, 2.1.2015, [https://media.ccc.de/v/31c3_-_6558_-_de_-_saal_g_-_201412282300_-_traue_keinem_scan_den_du_nicht_selbst_gefaelscht_hast_-_david_kriesel] — Vu : juillet 2017
- Ⓒ LECHNER Marie, *Le bruits des bots*, Strasbourg : Hear 22.02.2017
- Ⓒ MALAN David J., *Lecture 1 — Hardware : Understanding Computers and the Internet*, Cambridge : Harvard, 2011, [<https://www.youtube.com/watch?v=PLigQUosV34&t=3565s>] — Vu : novembre 2017
- Ⓕ PETZOLD Charles, *Code, the hidden language of hardware and software*, Microsoft Press, 2000

- (A) SAMPLE Ian, Harvard scientists pioneer storage of video inside DNA, *The Guardian*, 13.07.17 [<https://www.theguardian.com/science/2017/jul/12/scientists-pioneer-a-new-revolution-in-biology-by-embedding-film-on-dna>] — Lu : décembre 2017
- (D) SELLERS Donald, *History of Computers*, The History Channel, 01.01.1995 [<https://www.youtube.com/watch?v=VPLdJHrKAIU>] — Vu : novembre 2017

Politique

- (L) ARENDT Hannah, *Du Mensonge à la violence*, Calman-Lévy : Paris, 1972
- (B) ARENDT Hannah, *La Crise de la culture*, Gallimard : Paris, 1963
- (C) BRIDDLE James, *A New Dark Age - Turbulence, Big Data, AI, Fake News, and Peak Knowledge*, Chaos Computer Congress : Hambourg, 29.12.2016 [https://media.ccc.de/v/33c3-8163-a_new_dark_age#video&t=3394] — Vu : juillet 2017
- (A) BURGESS Matt, Facebook's changing its algorithms. You'll never guess what happens next, *Wired UK*, 5.8.2016 [<http://www.wired.co.uk/article/facebook-clickbait-algorithm-change>]
- (A) CADWALLADR Carole, Cambridge Analytica affair raises questions vital to our democracy, *The Guardian*, 4.3.2017 [<https://www.theguardian.com/politics/2017/mar/04/cambridge-analytica-democracy-digital-age>] — Relu : novembre 2017
- (A) CADWALLADR Carole, The great British Brexit robbery: how our democracy was hijacked, *The Guardian*, 07.05.2017 [<https://www.theguardian.com/technology/2017/may/07/the-great-british-brexit-robbery-hijacked-democracy>] — Relu : novembre 2017
- (A) CARPENTIER Megan, How Facebook powers money machines for obscure political 'news' sites, *The Guardian*, 24.8.2016 [<https://www.theguardian.com/technology/2016/aug/24/facebook-clickbait-political-news-sites-us-election-trump>] — Lu : novembre 2017
- (L) CASSILLI Antonio, CARDON Dominique, *Qu'est-ce que le digital labour ?*, Paris : Éditions de l'INA, 2014
- (A) DREYFUSS Emily, Facebook: too Big to Delete, *Wired*, [<https://www.wired.com/story/facebook-too-big-to-delete/>] — Lu : décembre 2017
- (A) GARCÍA MARTÍNEZ Antonio, The Solution to Too Much Facebook isn't More Facebook, *Wired*, 30.10.2017 [<https://www.wired.com/story/the-solution-to-facebook-overload-isnt-more-facebook/>] — Lu : décembre 2017
- (C) KRIESEL David, *SpiegelMining – Reverse Engineering von Spiegel-Online*, Chaos Communication Congress : Hambourg, 28.12.2016 [https://media.ccc.de/v/33c3-7912-spiegelmining_reverse_engineering_von_spiegel-online] — Vu : juillet 2017
- (A) LAPOWSKY Issie, Facebook didn't Create Donald Trump Alone: the Click Economy Did, *Wired*, 11.12.2016 [<https://www.wired.com/2016/11/facebook-alone-didnt-create-trump-click-economy/>] — Lu : novembre 2017
- (D) LEPERS John Paul, *Data brokers : la grande manipulation ?*, Vox Pop, Arte, 12.1.2017 [<http://info.arte.tv/fr/data-brokers-la-grande-manipulation>] — Vu : juillet 2017

- Ⓐ LEVITT Steven, DUBNER Stephen J., *Freakonomics: A Rogue Economist Explores the Hidden Side of Everything*, New York :William Morrow, 2005
- Ⓐ PIKETTY Thomas, *Le capital au XXI^e siècle*, Paris : Éditions du Seuil, 2013
- Ⓢ SIMON David, *The Wire*, Série (HBO), États-Unis, 2002
- Ⓐ SIMONITE Tom, When Government rule Software, Citizens are left in the Dark, *Wired*, 17.8.2017 [<https://www.wired.com/story/when-government-rules-by-software-citizens-are-left-in-the-dark/>] — Lu : décembre 2017
- Ⓐ SWARTZ Aaron, *Aaron Swartz, celui qui voulait changer le monde*, traduit par Marie-Mathilde Bortolotti et Amarante Szidon, Paris : B42, 2017

Surveillance

- Ⓐ BURGESS Matt, *GDPR will change data protection – here's what you need to know*, *Wired UK*, 7.11.2017, [<http://www.wired.co.uk/article/what-is-gdpr-uk-eu-legislation-compliance-summary-fines-2018>] — Lu : décembre 2017
- Ⓐ FOUCAULT Michel, *Surveiller et punir*, Paris : Gallimard, 1975
- Ⓐ HARCOURT Bernard, *Exposed*, *Cambridge : Harvard University Press*, 2015
- Ⓐ MEIER Patrick, Perils of Crisis Mapping: Lessons from the Gun Map, *iRevolutions*, 23.1.2013, <https://irevolutions.org/2013/01/23/perils-of-crisis-mapping/>, — Lu : décembre 2016
- Ⓓ POITRAS Laura, *Citizen Four*, 2015

Mise en page :

HTML / CSS

Avec l'aide du logiciel PrinceXml
pour l'export PDF

Version en ligne sur

memoire.marianneplano.fr

Fonts :

Archivo

Linux Libertine

Linux Libertine Bold

Linux Libertine Italic

Bande son d'écriture :

Kings of Convenience

Bande son de code :

J. Cole

Merci à :

Anne Laforet

Odile Sorlin

Luuse