CARNET DE BORD :
Big data et sciences sociales
Jean JACOB

SciencesPo

I- DES FOURMIS DANS NOS ALGORITHMES ?

Découper de manière optimale un tissu afin d'en tirer un nombre maximal de chemise, déterminer le trajet optimal pour se rendre d'un point A à un point B, ces problèmes peuvent être approchés par une même branche des mathématiques appliquées : l'optimisation. L'optimisation consiste facilement en la recherche du minimum (ou du maximum) d'une certaine quantité. Comment la nature peut-elle nous apporter dans la recherche d'extremum ?
La réponse se trouve dans une littérature déjà fournie, celles des *ant colonies optimization algorithm* (ACO). Dans les années 90, Marco Dorigo fut le premier à évoquer des translations entre les comportements observables des fourmis pour résoudre des problèmes d'optimisation, en essor suites aux nouvelles pratiques informatiques et algorithmiques. L'une des transpositions les plus réussies est celle des ACO, s'inspirant des pratiques de recherche de nourriture des fourmis. Cette dernière est détaillée dans son ouvrage *ant colony Optimization* publié aux presses du MIT :

"while walking from food sources to the nest and vice versa, ants deposit on the ground a substance called pheromone, forming in this way a pheromone trail. Ants can smell pheromone and, when choosing their way, they tend to choose, in probability, paths marked by strong pheromone concentrations." *Marco Dorigo, Ant colonies optimization algorithm*

La recherche de nourriture d'une colonie de fourmie repose sur une communication indirecte par phéromones. Marco Dorigo, après plusieurs expériences met en évidence que les fourmis réponde à un phénomène de "rétroaction autocatalytique", selon lequel plus le chemin est court, plus l'accumulation de phéromones est importante, et par conséquent plus le chemin est désirable et est en probabilité emprunté. Le scientifique en arrive à conclure que les colonies de fourmis ont une capacité d'optimisation intégrée : en utilisant des règles probabilistes basées sur des informations locales (concentration de phéromones), elles peuvent trouver le chemin le plus court entre deux points de leur environnement. Le comportement de recherche de nourriture des colonies de fourmis peut ainsi être reproduit en simulation et inspire les ACO. Ces algorithmes de fourmi (à présent " artificielles) constituent la création d'une métaheuristique qui repose sur l'idée que des agents simples peuvent réaliser des tâches complexes

"Ant colonies can accomplish complex tasks that in some cases far exceed the individual capabilities of a single ant. [...] Ant colony optimization is a metaheuristic in which a colony of artificial ants cooperate in finding good solutions to difficult discrete optimization problems." *Ibid*

Une métaheuristique peut être considérée comme un cadre algorithmique général qui peut être appliqué à différents problèmes d'optimisation avec relativement peu de modifications pour les rendre adaptés à un problème spécifique. La métaheuristique ACO a premièrement été appliquée au *travelling salesman problem*, qui est le problème d'un vendeur qui, partant de chez lui, veut trouver la tournée la plus courte qui le fait passer par un ensemble donné de

villes clientes puis revenir chez lui, en visitant chaque ville cliente exactement une fois. Intuitivement, l'on voit que la métaheuristique ACO s'applique donc à des problèmes d'optimisation que peuvent rencontrer les applications de calculs d'itinéraires, typiquement Waze. Dans le cas de Waze, il est d'autant plus intéressant de constater que la métaheuristique des ACO s'applique parfaitement, où les utilisateurs sont identifiés aux fourmis et le système de "pouces", indiquant des perturbations (circulation arrêtée, travaux, polices), identifié aux phéromones.

L'ACO n'est pas le seul algorithme s'inspirant du vivant. l'optimisation par essaims particulaires propose une autre transposition de l'ingéniosité du vivant aux algorithmes d'optimisation.

La force du vivant, et notamment de la structuration du vivant reposant sur une force collective comme l'explique Dorigo, est à l'origine d'une métaheuristique remarquable en informatique, invitant à constamment questionner les passerelles entre nature et algorithmique. S'inspirer de la nature pour innover, c'est le crédo d'une méthode d'innovation et de recherche : le biomimétisme. Sa force : sa diversité d'applications aussi variées que les problématiques que l'on retrouve dans la nature, et son inhérente durabilité.

II- IA ART ET RÉFLEXION SUR LA CAPACITÉ CRÉATIVE DES ALGORITHMES

L'utilité et les effets majeurs du *machine learning* dans de nombreux secteurs apparaissent incontestables, de même que son intégration dans nos activités quotidiennes. Ces dernières années, cependant, le machine learning a progressivement conquis des domaines nouveaux, au-delà de la science en général. Les industries créatives ont commencé à utiliser des technologies basées sur l'IA, non seulement pour optimiser leurs processus, comprendre les préférences des clients ou rendre la fabrication et la logistique plus efficaces, mais aussi pour repousser les frontières créatives et ainsi ouvrir de nouveaux horizons. Les implications de l'intelligence artificielle dans la sphère créative n'ont pas laissé le monde de l'art indifférent et une pratique artistique connue sous le nom d'*AI Art* s'est peu à peu déployée.

Une des premières mentions littéraires faisant allusion à une relation possible entre l'intelligence artificielle et l'art informatique se trouve dans les travaux de l'artiste et scientifique George Chaikin et de ses collègues, remontant à 1977. L'*AI Art*, faisant l'objet de cette entrée, est une pratique artistique qui a émergé grâce aux avancées les plus récentes dans les domaines du *deep learning* et de la reconnaissance d'image (*computer vision*). Depuis le développement de DeepDream en 2015, beaucoup d'efforts ont été consacrés à comprendre les subtilités du fonctionnement des ordinateurs. Le désir d'en savoir plus sur les couches cachées des réseaux de neurones artificiels, ainsi que leur énorme potentiel de

génération d'images, ont été la force motrice des travaux pionniers de nombreux chercheurs et artistes, notamment Mike Tyka, Memo Akten, Mario Klingemann et Robbie Barrat, pour en nommer quelques-uns. Jusqu'à présent, la créativité a toujours été perçue comme un trait humain, mais l'*AI Art* pousse ce concept plus loin en ajoutant le regard de la machine. La tension qui s'établit entre notre capacité humaine à inventer et les propositions d'ordre créatif renvoyées par la machine soulève de nombreuses questions.

On est ainsi en droit de se demander si un ordinateur peut être considéré comme un « agent créatif » autonome, pour reprendre l'expression empruntée à l'artiste Robbie Barrat, formule assez récurrente dans le débat sur la créativité des machines. Plus encore, on peut se demander si l'on peut considérer une machine comme un artiste à part entière ou si l'ensemble du processus qui entre dans un pipeline d'intelligence artificielle peut constituer une œuvre d'art.

La conférence TEDTalk que la technologue et dramaturge Kat Mustatea a donnée en 2018 est fascinante à cet égard. À travers des comparaisons avec le cinéma, la photographie et le théâtre, Kat Mustatea offre une perspective critique sur la valeur de l'art à l'ère numérique. Mustatea avance l'hypothèse que notre peur d'une superintelligence empêche de voir les algorithmes, les réseaux de neurones et l'apprentissage automatique pour ce qu'ils sont vraiment : des outils et des médiums que nous avons créés, donnant une forme similaire à celle de l'humain. Par conséquent, dans le domaine artistique, la véritable œuvre d'art n'est pas l'artefact basé sur l'IA, mais l'ensemble du processus, le bot, le logiciel que nous créons. L'artiste a conclu en disant que l'IA «est le grand nouveau médium de notre ère numérique», mais c'est une sorte de marionnette entre nos mains, et nous sommes les marionnettistes qui la dirigent.

Le manque de compréhension du fonctionnement de la technologie peut être trompeur et occulter le rôle joué par l'agent humain. Ainsi, les œuvres d'art numériques ou générées par ordinateur sont souvent critiquées ou dépréciées en raison du manque d'intentionnalité de la machine, ou pour le rôle prétendument limité de l'artiste. Pour saisir l'essence de cette pratique, il faut comprendre comment elle fonctionne. l'*AI Art* repose sur le deep learning, une sous-discipline du machine learning qui utilise l'apprentissage de la représentation combiné à des réseaux de neurones artificiels. Parmi les outils à la disposition des artistes figurent ainsi les GAN , *Generative Adversarial Networks*, une classe de réseaux de neurones inventée par le chercheur en machine learning Ian Goodfellow en 2014. Les GAN sont basés sur deux réseaux de neurones en compétition l'un avec l'autre : un réseau appelé un «générateur» crée de nouvelles images avec lesquelles il essaie de tromper un deuxième réseau. Ces images sont ensuite transmises à ce second réseau, appelé «discriminateur», qui a pour mission de les évaluer et de les accepter comme authentiques ou non. Les artistes IA entraînent généralement une machine avec d'énormes quantités de données — ou *datasets*. Cela permet aux réseaux de neurones d'apprendre non seulement quelles images générer mais aussi, bien que de manière souvent limitée, les éléments qui doivent caractériser ces images. Les artistes sélectionnent un algorithme, ajustent finement ses paramètres et adaptent ainsi ce qui est produit en sortie du GAN. Bien que les résultats de l'ordinateur ne soient pas entièrement contrôlables ou prévisibles, l'implication humaine est fondamentale. Compte tenu du fonctionnement très spécifique d'un réseau de neurones profonds, le processus

lui-même semble contenir une dimension artistique qui va bien au-delà de simples productions d'un algorithme informatisé.

Dans les industries créatives, la diffusion du *deep learning* demeure sujette aux controverses, entre enthousiasme et doute autour de cette technologie ainsi que sa pertinence dans des procédés créatifs. Cependant ces controverses ne doivent pas détourner l'attention des mots de Kat Mustatea, invitant à concevoir l'IA au prisme de son créateur.

III- DANS L'UNIVERS DU STUDIO AURECE VETTIER

Pour la rédaction de ce carnet de bord, l'artiste Paul Mouginot a partagé avec nous ses réflexions et ses travaux, nous aidant à mieux saisir la richesse artistique et conceptuelle que peut produire la rencontre entre art et intelligence artificielle.

Dans l'espace de la création par intelligence artificielle, le travail du studio parisien aurèce vettier porte une empreinte personnelle. Le studio fait se rejoindre l'idée artistique avec la maîtrise de l'exécution en s'appuyant sur les connaissances scientifiques, l'artisanat et l'amour de l'histoire de l'art. Le fondateur Paul Mouginot envisage le studio comme une expérience visant à favoriser une collaboration plus approfondie avec d'autres artistes et artisans, pour donner vie à de nouveaux types d'œuvres.

Leurs pièces aspirent à l'hybridité et sont le résultat d'inlassables allers-retours entre l'espace virtuel des données et du calcul et l'espace réel, dans lequel évoluent matières, couleurs, silhouettes... L'espace des données, où les mathématiques et le *deep learning* sont activés, est celui de la recherche et de la découverte de formes impossibles à déployer.
Aurèce Vettier n'envisage pas les propositions faites par la machine comme l'œuvre finale en raison de la relative rigidité des technologies IA. Non perçus comme des agents créatifs, ces algorithmes sont donc considérés comme des outils pour mettre au jour, modifier et assembler des œuvres nouvelles et sophistiquées. Cette recherche computationnelle ajoute toutefois un niveau élevé de savoir-faire requis par les œuvres d'art basées sur le *deep learning*, en prenant en compte toutes les étapes qui mènent à la création d'une œuvre d'art : de la collecte de données aux modèles, des algorithmes à l'ajustement des paramètres ; de la curation et l'entraînement de la machine à l'échange inlassable entre la dimension computationnelle et le monde réel.
Depuis 2018, aurèce vettier entraîne des GAN dans le cadre de la série Potential Herbariums (annexe 1 &2, page de couverture), une œuvre qui interroge le destin de la nature. Un travail passionné et dévoué a conduit à la combinaison d'images d'herbiers anciens existants avec celles créées après la collecte, le séchage et le catalogage de milliers de fleurs et de plantes dans la région montagneuse française de Savoie. 22 peintures à l'huile sur toile de différentes tailles exécutées par le peintre académique et artisan chinois Jack Lee, donne à représenter les fleurs impossibles générées par ces GAN, amenant à une alchimie entre la création humaine, fort de son savoir faire et son artisan, et l'intelligence artificielle, valorisant le savoir technologique. À son travail, Paul Mouginot a ajouté sa signature personnelle au travers

d'éléments faits à la main, écrits ou sculptés dans les cadres en bois, qui ont leur propre histoire à raconter, et permettent également aux œuvres d'être stockées et codées sur une blockchain.

IV- DES ALGORITHMES POUR AUTHENTIFIER LES OEUVRES D'ARTS

Si l'*IA Art* construit un débat philosophique autour de la création à l'intersection entre machine et humain, les méthodes d'intelligence artificielle se fraient d'ores et déjà un chemin dans les industries créatives comme soutien aux expertises humaines. à l'image de l'économètre qui se sert de l'intelligence artificielle pour approcher ses modèles de la réalité factuelle, l'intelligence artificielle a investi les sphères de l'authentification d'oeuvres d'art, un milieu souvent décrié pour son opacité, et que l'intelligence artificielle pourrait faire sortir de l'ombre.

Dans le monde de l'art, la vente de faux résonne comme un fléau. Pour donner un ordre d'idée, l'association pour la défense de l'œuvre de Miro estime que 20% des œuvres en circulation pourraient s'avérer être des faux. Le *National Gallery* de Londres, lors d'une exposition de 2010, prête attention a cette industrie parallèle. Intitulée *Close Examination: Fakes, Mistakes and Discoveries,* le titre fait parfaitement allusion aux enjeux sous tendus. D'une part, *fakes* réfèrent aux peintures réalisées dans l'intention délibérée de tromper. Communément appelées "faux", ce sont des œuvres qui ont été faussement présentées comme le produit d'un autre artiste ou d'une autre époque. *Mistakes* met en lumière l'idée que l'identification de faux dans le monde de l'art a toujours reposé sur l'appréciation des experts, fort de leurs connaissances, agissant selon le musée londonien avec "le meilleur savoir disponible du moment". Cette dépendance éveille tout de même les soupçons dès lors que des scandales éclatent au grand jour, à l'image de l'affaire Bill Pallot, expert du style XVIIIe siècle ayant délibérément introduit des fausses chaises Marie-Antoinette au château de Versailles . Finalement les *discoveries,* point de départ de notre entrée, évoquent les contributions vitales des sciences appliquées à la compréhension et l'authentification de tableaux anciens à des grands maîtres. En effet, la science joue un rôle primordial et croissant dans l'authentification d'œuvres. A titre illustratif, l'entreprise *Orion Analytical,* utilisant des méthodes de rayons X et microscopes électroniques poussés, a pu identifier qu'un portrait de Frans Hals d'une valeur de 10 millions de dollars, vendu en 2011 dans une vente privée orchestrée par Sotheby's était un faux. La maison de vente a dû rembourser intégralement l'acheteur, et a par la suite racheté la compagnie et engagé son fondateur James Martin.

L'intelligence artificielle peut devenir l'un des moteurs de l'authentification d'œuvres, en s'imposant comme un outil pratique, rentable et assez précis pour évaluer l'authenticité d'une œuvre. L'entreprise suisse *Art recognition* s'inscrit dans cette démarche et à des résultats concluants. L'algorithme conçu par la société s'appuie sur les méthodes de *Convolutional Neural Networks,* un algorithme de *deep learning* particulièrement performant lorsque les entrées, ou *imputs,* dans la boite noire de l'algorithme sont des images. S'inscrivant dans la lignée des algorithmes utilisant des méthodes inductives, le principe est de voir une image

non pas comme un tout, mais comme une matrice de *pixels*, pour en extraire des propriétés qui lui sont propres (aspect convolutif), à l'image de notre exercice de la séance 3 sur les présidents américains. L'algorithme d'*Art recognition* est entraîné à partir de d'oeuvres de l'artiste issu du catalogue raisonné (l'œuvre d'un grand peintre est traditionnellement définie par un groupe d'expert qui tient un "catalogue raisonné" d'œuvres authentiques de l'artiste) ainsi que de faux (pour matière à comparaison). L'entraînement repose sur l'hypothèse communément admise selon laquelle le motif des coups de pinceau constitue une empreinte propre à chaque artiste. L'algorithme cherche à apprendre les caractéristiques de coup de pinceau d'un artiste. Le résultat final de l'algorithme, ou *output,* consiste en une *heat map* relatant les zones où la divergence avec le coup de pinceau appris par intelligence artificielle de l'artiste sont les plus importantes (annexe 3).
Retour au National Gallery, et plus précisément au 2ème étage, salle 18, où loge *Samson et Delilah,* chef d'œuvre baroque attribué à Rubens. L'œuvre fut acquise par le prestigieux musée londonien à Christie's en 1980 pour un prix record à l'époque de 2,5 millions de livres sterling. Pourtant, nombreuses furent les critiques, mettant en doute l'attribution à Rubens du *Samson et Dalila* du musée londonien. Les critiques les plus exacerbées furent écrites sous la plume de l'artiste et spécialiste indépendante Euphrosyne Doxiadis, qui a affirmé dans plusieurs articles et interviews que certains détails ne collent pas. Parmi eux, l'expert estime que le tableau est peint d'un coup de pinceau plus lourd que le style réputé du maître. En 2021, l'algorithme d'*Art recognition* a également conclu que la *masterpiece* du National Gallery serait un faux. Dans un article paru dans The Guardian, *Was famed Samson and Delilah really painted by Rubens? No, says AI,* la CEO d'*Art recognition*, Dr. Carina Popovici livre les doutes qu'ont émis son algorithme à propos de l'authenticité de l'œuvre : "*The AI System evaluates Samson and Delilah not to be an original artwork by Rubens with a probability of 91.78%.*"

Malgré les risques que les technologies basées sur l'IA peuvent encore poser, la diffusion du *deep learning* dans les industries créatives a d'ores et déjà montré sa pertinence, comme l'atteste le récent rôle de l'intelligence artificielle dans l'authentification d'œuvres d'art. Si, Aujourd'hui l'ADAGP (équivalent de la SACEM pour les oeuvres d'art) et l'Office central de lutte contre le trafic des biens culturels réalisent des travaux considérables dans la lutte contre les faux, notamment en instaurant un cadre normatif pénalisant l'usage et la reproduction de faux, imaginer une technologie permettant d'authentifier une oeuvre pourrait résoudre le problème des faux d'une manière absolue pour nos oeuvres contemporaines. La blockchain et le stockage de NFT sur cette dernière (qui pourrait avoir une liaison physique avec une œuvre à l'image des œuvres de Paul Mouginot) pourrait faire percevoir un tel horizon.

Dans son discours présidentiel lors de la réunion annuelle de la Société d'histoire de la technologie (SHOT) de 1985, Melvin Kranzberg expose les "lois de Kranzberg", une série de six "axiomes" sur le rôle de la technologie dans la société. La première loi de Kranzberg s'articule de la façon suivante :

> "Technology is neither good nor bad; nor is it neutral".
> Melvin Kratzberg

Les interactions entre technologie et l'environnement social sont telles que les développements techniques ont souvent des conséquences environnementales, sociales et humaines qui vont bien au-delà des objectifs immédiats des dispositifs et pratiques techniques eux-mêmes, et la technologie peut avoir des résultats très différents lorsqu'elle est introduite dans des contextes différents ou dans des circonstances différente. Cette loi s'inscrit dans le prolongement de ce qui a été dit dans le cours sur une inscription sociale des pratiques scientifiques. L'historien de la technologie rejette alors le déterminisme technologique, cet axiome assurant que la maximisation des opportunités et la minimisation des risques liés au développement de technologies sont des processus qui sont socialement construits. Cette loi est le reflet d'un doxa indispensable sur la technologie selon lequel l'intelligence humaine reste toujours le guide de ces dernières. Aussi rassurant que déroutant pour certains que cela puisse paraître, c'est à la conscience humaine d'amener les technologies vers une direction de développement souhaitée, une idée qui n'est pas sans rappeler celle de Kat Mustatea sur la dépendance des machines à leur créateur. Le Big Data n'y échappe donc pas.

V- LES ÉTATS UNIS SE POSITIONNENT SUR LA RÉGULATION DES DONNÉES.

Les préoccupations sur la vie privée et le contrôle de l'Etat ou des acteurs de la *tech* grâce aux données remontent à l'origine même de la gestion de base de données électroniques. La prise d'empreintes digitales pour les détenus ou même le contrôle de l'impôt sur le revenu des travailleurs ont été parmi les premières bases de données mises en place aux États-Unis avant les années 20. Elles posaient d'ores et déjà des questions relatives à la vie privée et au contrôle d'ordre éthique et juridique. Le Big Data a porté ces questions à un tout autre niveau.

En référence au scandale de surveillance de masse de la *NSA* [Agence nationale de sécurité des États-Unis] porté à ébullition par Edward Snowden en 2013, l'éditeur de l'influente revue informatique "Communications of the ACM" commente : " La *NSA* est en train de réduire la technologie de surveillance d'Orwell [*1984*] à une technologie plutôt primitives en comparaison ". Si la panoptique *Big Brother* offrait un mode de surveillance dystopique aux yeux des lecteurs contemporains d'Orwell, force est de constater que nos activités numériques laissant des traces qui peuvent toujours être traquées et analysées, constituent un terreau très fertile à la surveillance. Cette traçabilité intangible de nos données se heurte à

fortiori à des préoccupations sur la vie privée. Dès lors, l'instauration de réglementations s'impose-t-elle dans la collecte de données numériques ?

En ce qui concerne la réglementation autour des données, un rapport de la Maison Blanche de 2014 suggère de renoncer pour l'essentiel à la tentative de réglementer les données qui peuvent ou non être collectées :

"L'attention des politiques devrait se concentrer davantage sur les utilisations réelles du big data et moins sur sa collecte et son analyse. Par utilisations réelles, nous entendons les événements spécifiques où quelque chose se produit qui peut causer une conséquence négative ou un préjudice à un individu ou à une classe d'individus "

big data: seizing opportunities, preserving values, Executive Office of the President 2014.

Cela correspond à la philosophie juridique qui consiste à ne pas réglementer la possession d'un objet, mais sanctionner son utilisation illégale. Par exemple, le cas des armes à feu aux États-Unis suit cette philosophie juridique. Les conséquences négatives mentionnées seraient alors sévèrement punies par la loi. Seulement, aux Etats unis, ne disposent pas d'une loi unique couvrant la confidentialité de tous les types de données au niveau fédéral. Dès lors, il semble difficile de contrôler de tels engagements sans règles juridiques plus contraignantes. Dans un *memorandum*, la NSA assume avoir violé certaines lois relatives aux protections des données. (Memorandum of the Chief of SID Oversight & Compliance (OC-034-12). Cette auto-déclaration témoigne dès lors que même lorsque des règles et des règlements sont en place (aussi incomplètes soient-elles), les avantages liés à la violation de ces règles sont si tentants que les gouvernements ont admis avoir entrepris des démarches illégales pour en tirer profit.

Outre les défis pratiques de la réglementation de la possession de données, autoriser la collecte de données consiste à s'assurer de ne pas limiter les avantages potentiels de son utilisation. En effet, "Savoir, c'est pouvoir" pour gouverner, les États et leurs administrations ont besoin de connaissances et de données, si bien que le big data offre un exercice du pouvoir sans précédent.
Une position de la maison blanche dans le même rapport de 2014 expose l'ambiguïté autour du recours au Big Data à des fins de contrôle :

"When wrestling with the vexing issues big data raises in the public sector, it can be easy to lose sight of the tremendous opportunities these technologies offer to improve public services, grow the economy, and improve the health and safety of our communities. These opportunities are real and must be kept at the center of the conversation about big data."

White House, Ibid

En Europe, le RGPD nouvellement entré en vigueur protège les données personnelles des utilisateurs. Contrôler et réguler l'usage de données pose de nombreuses questions: *quid* du

contrôle par le numérique ? Le positionnement de la CNIL sur le pass sanitaire, alertant des risques d'une banalisation d'outils numériques à des fins de contrôle, devrait ouvrir un débat particulièrement intéressant sur l'instauration de ce type de mesures.

VI- NE PAS SUCCOMBER À LA PANIQUE *FAKE NEWS*

De Récents événements sociaux et politiques, sans citer pour une énième fois l'élection présidentielle américaine de 2016, ont été marqués par la dissémination sur le web de "*fake news*", c'est-à-dire des informations fabriquées qui diffusent des contenus trompeurs, ou déforment grossièrement les informations réelles, partagées sur les plateformes de médias sociaux. Si la désinformation et la propagande existent depuis l'Antiquité, leur importance et leur influence à l'ère des médias sociaux suscite des interrogations plus exacerbées. En effet, la désinformation numérique massive a été désignée comme un risque technologique et géopolitique majeur par le rapport 2013 du Forum économique mondial.
Au-delà de la cristallisation du rôle politique et social des *fakes news* par les médias, le consensus académique est à la dédramatisation du phénomène. De nombreux papiers cherchent à réévaluer la place des *fakes news* dans l'opinion publique numérique, et ce dès 2016 (Gentzkow, 2016*, Les fake news ont-elle fait élire Trump*)

Cette entrée propose un détail des résultats publiés par Jennifer Allen and al. *Evaluating the fake news problem at the scale of the information ecosystem*. Le groupe de recherche d'Allen s'inscrit dans une littérature contestant la prévalence des fake news sur nos différents médias d'informations. L'étude met en évidence plusieurs résultats. Premièrement, si les Américains ne sont pas informés sur l'actualité, c'est parce qu'ils choisissent tout simplement de ne pas l'être. L'ère numérique signe l'apologie du divertissement. La consommation d'informations en ligne (même avec la consommation passive d'informations) ne représente que 4,2% de la consommation moyenne totale en ligne de la population ; et la consommation d'informations à la télévision représente 23% de la consommation totale de contenus télévisés. La télévision reste le principal médium d'information de la population américaine. Les informations en ligne ne représentent qu'une très faible part de la consommation médiatique de la population, minoritaire en comparaison des chaînes d'informations télévisées. Couplée à ce décalage, la proportion de *fake news* est alors quantitativement très faible : les fake news ne représentent que 0,15 % du régime médiatique quotidien des Américains.

Face à ces résultats, les auteurs ne peuvent pas considérer les *fake news* comme responsables d'une désintégration politique et de la polarisation de la société : l'impact quantitatif des *fake news* est trop faible. L'importance des fake news et la place qui leur est accordée, tant dans la société que dans les travaux de recherche, doivent donc être reconsidérées, selon les auteurs.

Pour arriver à de tels résultats, les auteurs ont travaillé à partir de deux *datasets :*
Les panels individuels de *Nielsen* sur la télévision et le web, un outil réputé pour sa représentativité, si bien qu'il est utilisé afin de répartir les budgets publicitaires aux Etats Unis. Ensuite, les auteurs se basent sur le *Comscore* sur le trafic des applications et des

navigateurs (de bureau et mobiles). Ces deux bases de données fournissent des données très représentatives des habitudes de consommation des médias aux États-Unis.
comment évaluent-ils les proportions de *news* et *fake news* à partir de ces bases de données ? D'une part, les auteurs prennent le parti pris de dire que la télévision ne relaie pas de *fake news*. En ligne, un aspect de la consommation qu'il est facile de mesurer est le temps passé sur les sites de médias. Mais il faut aussi tenir compte du temps passé sur les plateformes de consultation de contenus liés à l'actualité, notamment et essentiellement sur les médias sociaux.
Leur méthode pour évaluer la proportion de fake news sur les plateformes est la suivante : "Pour chaque visite sur un site de médias sociaux, nous examinons le site que l'utilisateur a visité immédiatement (dans une fenêtre d'une minute) après la session." Cette méthode par *referrals* qui est la manière la plus simple de mettre en évidence la part d'exposition à des contenus médiatiques sur les réseaux sociaux rejette les contenus directement produits sur la plateforme, qui peut donc constituer une certaine limite aux résultats. Le déroulé demeure tout de même convaincant pour démystifier l'impact des *fake news,* ou du moins ne pas cédér à la panique au regard des vagues médiatiques les diabolisant.

VII- FACEBOOK ET GILETS JAUNES : GENESE D'UN ESPACE DE DÉLIBÉRATION PUBLIC ?

Dans le cadre du cours de culture du numérique, dispensé par Dominique Cardon, mon groupe avions enquêté sur le rôle de facebook dans la montée fulgurante du mouvement des Gilets Jaunes. Cette entrée restitue certains de nos résultats empiriques, dont certains extraits d'interview avec Maxime Nicolle, l'une des figures médiatiques du mouvement Gilets Jaunes, ou Vincent Glad, journaliste.

Dans le numéro de la revue Pouvoirs intitulé La datacratie, Fabien Granjon développe le concept de « mouvement en crise », auquel le phénomène des Gilets jaunes semble correspondre. Forme moderne de contestation sociale, ces mouvements répondent aux espoirs de démocratie participative et de nouvel espace public qu'a constitué internet dès sa naissance, et plus nettement depuis les années 90. Ces mouvements en crise sont alimentés par une large contestation de l'ordre social et politique établi, régi par des élites politiques en place, jugées déconnectées des localités propres des citoyens.
Facebook représente une *présupposée* opportunité inédite de représentativité et de prise de décision horizontale, si bien que ce réseau social s'installe comme principale plateforme d'échanges et de communications du mouvement, notamment via des groupes. Que ce soient des journalistes, des meneurs du mouvement, des sociologues, tous reconnaissent le rôle central de Facebook aussi bien dans la genèse que dans l'évolution du mouvement.
La trajectoire *on-line* du mouvement semble avoir été d'une certaine manière un miroir de la trajectoire *off-line* : au départ la mobilisation était locale. Dans une étude publiée dans la revue economique des presses de sciences po, Pierre Boyer et al. (2020) met en évidence l'importance montre bien que les départements les plus actifs dans les mobilisations

physiques étaient aussi ceux où la mobilisation en ligne sur Facebook a été la plus importante (annexe 5). Si ces résultats confirment l'existance d'interconnextion entre les bulles gilets jaunes en ligne et celles actives physiquement ils témoignent également de l'origine locale du mouvement.

La mobilisation locale en ligne a pu être accrue par l'algorithme de recommandation facebook et notamment une variation dans l'ordre de priorisation de ce dernier (bien qu'aucune données quantitatives n'aient encore été produites sur le sujet. Dans un post Facebook du 12 janvier 2018, Mark Zuckerberg annonce deux changements dans la manière dont l'algorithme Facebook dresse le *news feed* de chacun : ce dernier compte rendre prioritaire les posts des amis, des familles ainsi que les groupes facebook de l'utilisateur, tout en favorisant une information locale. Cette nouvelle priorisation pourrait jouer un role dans la propagation d'informations et de contenus gilets jaunes sur Facebook, et expliquerait la montée en popularité de certains groupes facebook locaux de revendications communes.

L'évolution locale, ainsi que les travaux de modérations et d'homogénéisation des groupes par les modérateurs ont permis une concertation au niveau nationale sur facebook et sur les ronds points pour des journées entières de blocages routiers généralisés partout en France. En effet, le journaliste Vincent Glad qui s'est infiltré dès le début du mouvement sur des groupes Facebook nous a indiqué que tel qu'il avait pu le sentir, les groupes et pages nationales étaient devenues bien plus actives que les groupes locaux au fil des manifestations. Les groupes locaux servaient surtout à organiser une mobilisation à l'échelle locale, tandis que les groupes nationaux étaient devenus les sphères et lieux virtuels de discussion et de débat politique pour l'avenir du mouvement.
Il est nécessaire d'accentuer sur l'aspect virtuel de ces discussions qui constitue une nuance entre démocratie directe en ligne et idéal de démocratie directe « à l'athénienne ». Le mode de dialogue en ligne obéit à un schéma particulier, notamment à travers la culture du troll ou du clash. Conséquences directe de l' « ère du commentaire », une réflexion progressive, longue et argumentée, assimilable à un discours, retient moins l'attention qu'une réaction plus immédiate, brève et à caractère polémique (humoristique, effrayante, indignante,…). Ce faisant, le débat et la délibération commune font moins progresser un point de vue, une opinion particulière ou encore une ligne d'action (politique ou non) qu'ils ne confortent nombre de membres de ces groupes dans leurs certitudes sur une position généralement commune, effet de bulle de filtre.

"Facebook est un outil; il nous a été et nous est encore utile, mais c'est certain que sa portée nous échappe."

Maxime Nicolle, gilet jaune influent connu sous le pseudonyme *Fly rider*

VIII- PENSÉE A TOUT CEUX QUE MES ARTICLES NE CONCERNENT PAS : RÉFLEXION SUR L'INCLUSION NUMÉRIQUE.

Les coûts d'une société est constante transformation, c'est la construction d'une société à plusieurs vitesses. Dans le cas de la transition numérique, certains sont bloqués à la vitesse 0, tandis que d'autres atteignent la vitesse de la lumière.
À l'heure où le code du travail instaure un droit à la déconnexion, assurant aux employés un droit à ne pas être joignable à tout moment via les canaux de transmissions numériques, l'enquête « Conditions de vie et aspirations » rend paradoxalement compte qu'un tiers de la population française serait potentiellement en difficulté face à la transition numérique. La vitesse 0 de la transition numérique constitue un facteur d'exclusion croissant au siècle de la dématérialisation.
L'ère numérique offre aussi bien de formidables opportunités qu'elle exclut certains de cette dynamique. Jean Deydier dédie sa contribution à la revue Pouvoirs en 2018 à l'inclusion numérique. Dans *les exclus de la Datacratie*, le fondateur d'Emmaüs connect et *We Tech Care* (deux des association d'inclusion numérique les plus influentes de France) met en évidence que la difficulté face à la transition numérique de certains pans de la société résultent de deux problématiques : "des difficultés d'accès, et des limitations liées aux usages". Si les problèmes d'accès au numérique, ayant été la pierre angulaire des politiques d'inclusion numérique ces dernières années, tendent à se résoudre aussi bien géographiquement que économiquement, l'on voit une explosion de difficultés sur les usages, d'autant plus probant à l'heure d'une dématérialisation de nombreux services. Marie Alauzen, qui a consacré sa thèse à la modernisation des services publics en France, étudie la mise en place d'un système tel que France Connect, facilitant la lourdeur des procédures administratives françaises. Certaines procédures n'ont même plus d'équivalents hors ligne. Pourtant, si cette dématérialisation cherche à simplifier ces procédures, 40% des Français se déclarent inquiets à l'idée d'accomplir des démarches administratives en ligne. La non connectivité peut également susciter des formes de discriminations dans le secteur scolaire.
L'approche intersectionnelle permet également de mettre en évidence que l'exclusion numérique se construit à l'intersection d'autres formes d'exclusions. l'exclusion numérique est la plus prégnante pour des personnes en précarité sociale de sorte que plus de la moitié des personnes non connectées possèdent des bas revenus ou un niveau d'études faible. (non pas l'âge, mettant de côté l'hypothèse d'une inclusion numérique totale de la société au fil du renouvellement de la population) On assiste à une double peine, les services numériques visant à faciliter le quotidien, et rendre plus accessible les démarches, ne profite pas aux personnes qui en ont le plus besoin.

Par ailleurs, les règles de distanciation sociale liées à la crise sanitaire ont renforcé la nécessité à la connexion en ligne, nous coupant pendant plusieurs mois de toute interactions physiques. Dans ce contexte, la démultiplication des échanges et procédures en ligne (télétravail, scolarité sur zoom, etc.) a propulsé l'espace en ligne vers un statut primordial, si bien que le plan de relance France relance considère l'inclusion numérique comme l'un de

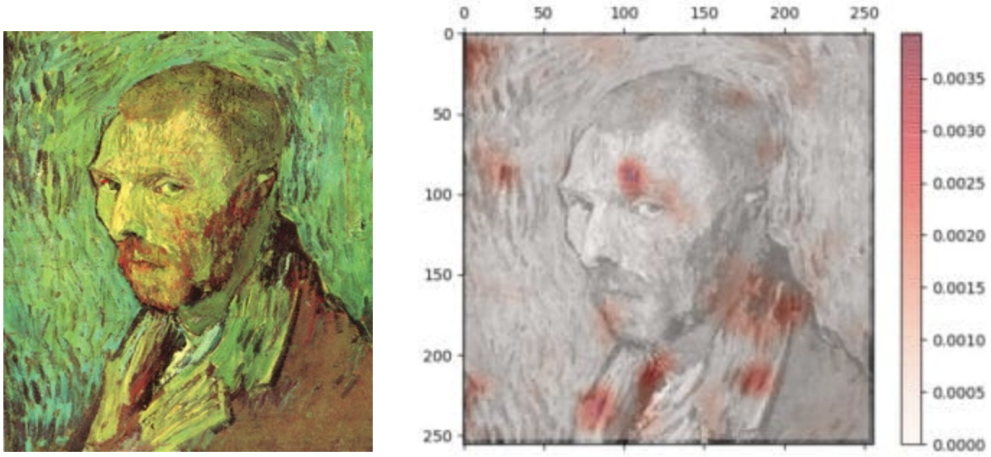
ses fers de lance. Une certaine conscience collective, pour laquelle Jean Deydier plaidoyait, semble se mettre en place face à ce problème.

La transition numérique bouleverse l'ensemble de la société. De nos jours, le paradigme du Big Data, qui correspond à l'utilisation massive de données afin d'établir de nouvelles formes de priorités et de nouvelles techniques d'estimations pour informer toutes sortes de décisions, s'immisce dans l'ensemble des procédés de production de connaissances, une tendance qui n'est pas prête de décroître. Seulement, à l'instar des anciens moyens de productions de sciences, il est primordial de conserver un "œil habillé" pour nous préserver des chemins dystopiques d'une telle transformation sociétale.

Annexes

aurèce vettier, *areal.collect(128)*
AV-2021-U-143
Huile sur Toile
Exemplaire unique, 283 × 185 cm — Courtesy of Romain Darnaud

aurèce vettier, territoires de la guêpe et son fardeau somptueux, Huile sur Toile



“L'une de nos analyses les plus précises. Notre IA a classé l'Autoportrait d'Oslo comme authentique avec une probabilité de 97 % – l'une des évaluations les plus précises que notre algorithme n’ait jamais produites ! Pour fournir des preuves visuelles de l'évaluation de l'algorithme, nous avons également généré une carte thermique qui est interprétée comme suit : les zones les plus importantes pour la décision de l'algorithme sont celles surlignées en rouge, et l'importance diminue progressivement proportionnellement à la nuance de rouge ; à l'autre extrémité du spectre, les régions non colorées ont le moins d'influence sur la décision. Par conséquent, les zones marquées ont la plus grande importance pour la décision finale.”
Source : Site web Art recognition

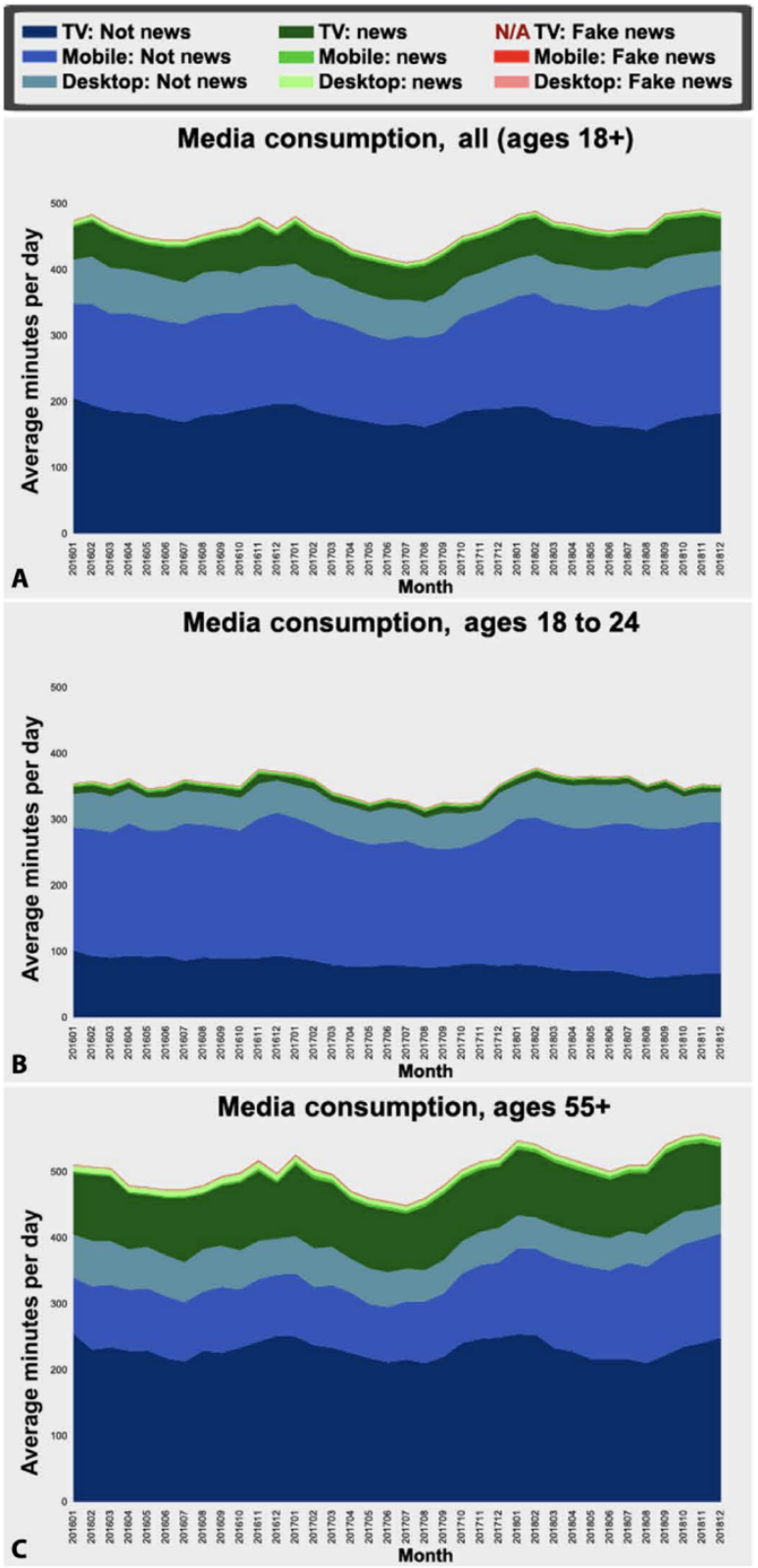
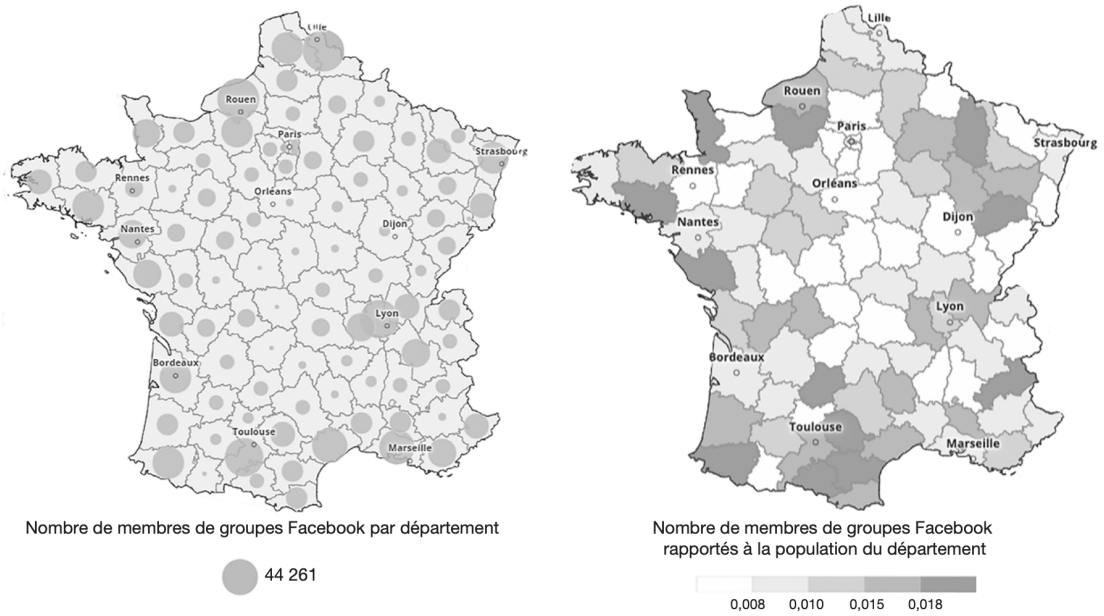


Fig. 1. Overall information consumption by category and platform over time, from January 2016 to December 2018. Breakdown of consumption for (**A**) the entire adult sample, 18 years and older, (**B**) 18 to 24 years old, and (**C**) 55 years and older. See table S6 for numerical values.

extrait de

Figure 5. *La mobilisation en ligne par département*



Extrait de : Boyer, P. C. et al. (2020) Les déterminants de la mobilisation des Gilets jaunes. *Revue économique*. [Online] 71 (1), 109–138.

Bibliographie

Références littéraires :

- Allcott, H. & Gentzkow, M. (2017) Social Media and Fake News in the 2016 Election. The Journal of economic perspectives. [Online] 31 (2), 211–235.

- Allen, J. Watts, J.D. & al. (2018) *Evaluating the fake news problem at the scale of the information ecosystem*, Sci Adv 6 (14),

- Andrews, L. (2012). *I Know Who You Are and I Saw What You Did: Social Networks and the Death of Privacy,* Simon and Schuster

- Deydier, J. (2018) Les exclus de la datacratie. Pouvoirs. [Online] 164 (1), 137–148.

- Dorigo, M. & Stützle, T. (2004) Ant colony optimization. Cambridge, Mass: MIT Press.

- Glover, F. and M. Laguna (1997) Tabu Search, Kluwer Academic Publishers, Boston.

"A metaheuristic refers to a master strategy that guides and modifies other heuristics to produce solutions beyond those that are normally generated in a quest for local optimality."

- Hilbert, M. (2016). *Big Data for Development: A Review of Promises and Challenges.* Development Policy Review, 34(1), 135–174.

- Kranzberg, M. (1995) 'Technology and History: "Kranzberg's Laws"', *Bulletin of Science, Technology & Society*, 15(1), pp. 5–13

- Vardi, M. Y. (2013). *The end of the American network.* Communications of the ACM, 56(11), 5–5.

TedTalk :

- Kat Mustatea, *What is the value of art in an age of thinking machines.* URL

Websites :

- ADAGP, *mémento à l'usage des artistes*
https://www.adagp.fr/sites/default/files/faux-contrefacon-auteurs-web.pdf

- Art recognition
https://art-recognition.com/fr/case-studies/vincent-van-gogh-self-portrait/

- Gouvernement FR
https://www.economie.gouv.fr/plan-de-relance/conseillers-numeriques-engagement-acteurs-prives

- National Gallery of London :
https://www.nationalgallery.org.uk/research/research-papers/close-examination

- *Society for the history of technology*
https://www.historyoftechnology.org/

- Successio Miro
https://www.successiomiro.com/fr/authentification

- *White House*
big data: seizing opportunities, preserving values, Executive Office of the President 2014.

Medias :

- Connaissance des arts
https://www.connaissancedesarts.com/artistes/pierre-paul-rubens/une-intelligence-artificielle-remet-en-cause-lauthenticite-dun-rubens-de-la-national-gallery-de-londres-11164559/

- The guardian
https://www.theguardian.com/artanddesign/2021/sep/26/was-famed-samson-and-delilah-really-painted-by-rubens-no-says-ai

- The economist
https://www.economist.com/books-and-arts/2021/01/07/new-tools-are-making-it-easier-to-authenticate-paintings