
Le nouvel esprit critique¹

Pierre Lévy

University of Montreal

<https://orcid.org/0000-0003-1599-221X>

1. Le programme

La grande majorité de la population mondiale est – ou sera bientôt – connectée à l'Internet, c'est-à-dire à un appareillage global de communication, de mémoire et de calcul encore inimaginable il y a un siècle et qui contient désormais la vie de la cité, de ses cimes scientifiques, artistiques et spirituelles à ses abîmes criminels. Du côté obscur de la force, les pouvoirs économiques, politico-militaires et culturels utilisent sans scrupules le nouveau médium algorithmique afin de combattre leurs adversaires et de séduire, exploiter ou asservir leurs sujets. Du côté lumineux, l'internet ouvre à une diversité d'information et à une liberté d'expression inconnues à l'âge des médias classiques. Nous commençons tout juste à explorer les possibilités de coordination à grande échelle et d'intelligence collective offertes par l'Internet.

Bien que nous ayons tendance à les opposer d'un point de vue moral, le pouvoir d'oppression et de destruction – négatif – et la puissance de création et de pensée – positive – sont deux aspects de la même augmentation cognitive. Et nous-mêmes – à notre corps défendant, et quoique l'obscurité semble toujours venir de l'autre – nourrissons les deux figures de la force et leur donnons prise, y compris par la servitude volontaire si bien décrite par La Boétie et par la participation à la foule de micro-processus et de rapports sociaux qui nous traversent et que Michel Foucault nous a invité à analyser en détail². J'ajoute que

¹ A versão traduzida deste artigo compõe o capítulo especial do livro "A TÉCNICA COMO POTÊNCIA DO HUMANO: CONEXÃO DE SABERES EM SAÚDE E EDUCAÇÃO" de organização de Francisco Milton Mendes Neto, Nize Maria Campos Pellanda e Maria de Fátima de Lima das Chagas e publicado pela Edufersa, 2024. A Revista Eletrônica Kuab agradece a cessão do texto ao seu autor e aos organizadores do livro.

² *Surveiller et punir*, Gallimard, 1975, *Histoire de la sexualité*, I, II, III, Gallimard, 1976- 1984

la censure par les plateformes et leurs systèmes d'intelligence artificielle ne nous protégeront jamais de manière infaillible contre les «*fake news*» et les manipulations en tous genres. Le seul remède sérieux aux menées du côté obscur et la meilleure manière d'actualiser les virtualités positives du nouveau médium est l'éducation des populations et tout particulièrement leur entraînement à la communication réfléchie et à la pensée critique.

Le dernier siècle a été marqué par des guerres épouvantables, des génocides, des totalitarismes implacables et mille formes d'oppression, coloniales et autres. Pourtant, tout n'est pas sombre. Entre 1900 et 2020, la population humaine est passée d'1,7 milliards à 7,6 milliards d'individus. Gagnant plus de vingt ans, l'espérance de vie a bondi. La proportion de la population alphabétisée est passée de quinze à plus de quatre-vingt pour cent et la grande pauvreté a reculé. A comparer avec le XIX e siècle, nous avons affaire à une nouvelle humanité: urbaine, éduquée, en meilleure santé et dont la moitié féminine a gagné en puissance. Je ne prévois pas pour demain un monde sans conflits ni relations de pouvoir. Mais pourquoi ne pas s'appuyer sur la montée de l'alphabétisation et les nouvelles ressources du médium algorithmique pour équilibrer les tendances totalitaires de l'état plateforme, subvertir ses évaluations statistiques et favoriser l'apprentissage collaboratif autonome?

Il faut bien distinguer *l'esprit critique*, c'est-à-dire l'usage réfléchi du dialogue et de la raison, de la *théorie critique* (bien illustrée par³ et ses successeurs) pour qui la mécanisation des fonctions communicatives et intellectuelles humaines relève uniformément d'une dystopie totalitaire. Certes, comme je l'ai montré au chapitre précédant, ces tendances totalitaires existent. Mais les possibilités de jeu sont aussi présentes. Les avions – plus lourds que l'air – ont utilisé les lois de la physique pour échapper à la gravité et les bons judokas retournent en leur faveur la force de l'adversaire. Dans le même esprit,

³ Notamment dans *La Dialectique de la raison*, écrit avec Max Horkheimer, particulièrement dans le chapitre sur l'industrie culturelle. La première publication date de 1944 : *Philosophische Fragmente*, Institute of Social Research, New York. L'ouvrage a connu de nombreuses versions successives et traductions. Les deux thèses principales de Adorno et Horkheimer sont que la civilisation inspirée par la philosophie des lumières s'est retournée contre ses principes humanistes originels et que les industries culturelles mécanisées (presse, magazines, disques, cinéma, radio, télévision) exercent une domination totalitaire sur les esprits.

pourquoi ne pas utiliser la puissance de mémoire et de calcul des centres de données pour soutenir une insurrection tranquille de l'esprit au service de l'intelligence collective... avec pour horizon un développement humain qualitatif en harmonie avec la biosphère?

La puissance des algorithmes devrait être confiée à une population cultivée, formée au travail du sens, instruite de la nature de la vie symbolique, initiée à la complexité de la communication humaine et capable de ruser avec les algorithmes de l'état plateforme. Nous avons du mal à imaginer une civilisation dont l'éducation mettrait la mémoire de l'humanité, les programmes d'analyse automatique, l'apprentissage collaboratif et l'esprit critique à la portée des populations. Cela nous apparaît comme une utopie irréalisable. Mais souvenons-nous que les scribes ont été pendant des siècles une petite minorité dominant une population illettrée. Qui aurait pu prévoir, il y a trois mille ans, que quatre-vingt cinq pour cent de la population humaine saurait un jour lire et écrire? Pourtant c'est aujourd'hui le cas.

Le nouvel esprit critique ne porte pas tant sur la définition des bons ou des mauvais contenus que sur la méthode par laquelle les personnes et les communautés, quel que soit leur contexte, peuvent gagner en autonomie et en puissance intellectuelle. Comment transformer les torrents de données en fleuves de connaissances? Il nous faut d'abord prendre conscience de la nature stigmergique de la communication contemporaine afin de s'en servir de manière responsable. Nous devons ensuite mettre en oeuvre les grands principes du logiciel libre, des communs de la connaissance et de la souveraineté des individus sur leurs données. Ouvrons à tous et gratuitement les ressources d'apprentissage et d'analyse. Sur la base des communs, exerçons-nous au filtrage, à l'indexation et à l'évaluation collaborative des données pour une meilleure gestion des connaissances. L'analyse des flux de données doit devenir une compétence essentielle enseignée à l'école parce qu'elle conditionne désormais la compréhension du monde. Enfin, n'abandonnons pas l'organisation et l'interprétation des données aux algorithmes statistiques mais cultivons dès l'école par des méthodes actives la capacité à construire – dans et pour l'action – des modèles causaux de la réalité.

2. La communication stigmergique

Depuis l'émergence d'*homo sapiens*, la communication et la pensée humaine sont supportés par les symboles. Or les symboles ont pris la forme de *données*, partout présentes, interconnectées, calculables. Limité à un cercle très étroit de spécialistes il y a encore cinquante ans, l'univers logiciel sert désormais d'intermédiaire et de terrain de jeu à nos rapports sociaux, économiques et politiques, ainsi qu'à l'éducation et à l'apprentissage⁴. Même si nous n'en n'avons pas clairement conscience, notre société est devenue datacentrique: nos relations passent par les données. Cela ne signifie certainement pas que les formes socio-politiques et les modes de subjectivation antérieurs soient effacés ni même obsolètes, mais plutôt qu'ils se réorganisent en fonction des nouveaux rapports sociaux qui se nouent dans le médium algorithmique. Notre vie est-elle désormais «asservie aux données?»⁵ Sans doute, s'il est vrai que la vie humaine est asservie aux symboles. Mais n'oublions pas le sens inverse de la boucle de rétroaction. L'humanité se reproduit et elle évolue à partir de la *symbiose* entre des écosystèmes d'idées et les populations de primates parlants qui les entretiennent, s'en nourrissent et s'y réfléchissent...

L'internet a distribué le pouvoir de lire, d'écrire, de publier, de critiquer, de classer et d'organiser l'information. Du texte statique sur papier, nous sommes passé à l'hypertexte ubiquitaire. Qu'une information se trouve en un point du réseau et la voici partout. Dès lors, si l'échange de messages point à point a toujours lieu, le gros de la communication sociale s'effectue de manière *stigmergique*. L'étymologie grecque de ce mot étrange explique assez bien son sens: des signes (*stigma*) sont déposés dans l'environnement par l'action ou le travail (*ergon*) de membres d'une collectivité, et ces signes guident en retour – et récursivement – leurs actions⁶. Le cas classique est celui des fourmis qui laissent

⁴ Voir sur ce point *Software Takes Command* de Lev Manovich, Bloomsbury, London, 2013.

⁵ Dans un ouvrage récent Facebook. *L'école des fans*. Atelier Sens Public, Montréal, 2018, Gérard Wormser évoque « le spectre d'une vie humaine asservie aux données ». Le fondateur de la revue Sens Public a sans doute bien perçu l'ampleur de la mutation, mais il faut dépasser l'effroi initial et regarder la réalité en face.

⁶ Selon Wikipedia, qui reprend ici un concept créé par les éthologues « la stigmergie exprime la notion que les actions d'un agent laissent des signes dans l'environnement, signes perçus par lui-même et

une traîne de phéromones sur leur passage lorsqu’elles ramènent de la nourriture à la fourmilière. L’odeur des phéromones incite d’autres fourmis à remonter leurs traces pour découvrir le butin et ramener des vivres à la ville souterraine en laissant par terre à leur tour un message parfumé. Les fourmis communiquent donc par l’intermédiaire des traces qu’elles laissent dans leur environnement.

Plongés dans le réseau – la tête dans les nuages, les yeux, les oreilles et les doigts aux interfaces – nous communiquons par l’intermédiaire de la masse océanique des données que nous produisons, transformons et consommons. Chaque lien que nous créons, chaque tag ou hashtag que nous apposons sur une information, chaque acte d’évaluation ou d’approbation, chaque «j’aime», chaque requête, chaque achat, chaque commentaire, chaque partage, toutes ces opérations modifient subtilement la mémoire commune, c’est-à-dire la masse inextricable des *rappports* entre les données. Notre comportement en ligne émet un flux continu de messages et d’indices qui transforment la structure de la mémoire et contribuent à orienter l’attention et l’activité de nos contemporains. Il en est ainsi parce que l’information que nous produisons individuellement est traitée par des algorithmes afin d’être transformée en information utile pour la collectivité. Nous communiquons donc de manière indirecte, en ajoutant des données à la mémoire et en modifiant la structure des relations entre les données qui tissent notre contexte commun. Nous déposons dans l’environnement virtuel des phéromones électroniques qui déterminent en boucle l’action des autres et entraînent par dessus le marché les neurones formels des intelligences artificielles. En fin de compte, la communication s’établit toujours entre des personnes mais, dans le médium algorithmique, elle a lieu principalement sur un mode *stigmergique*.

On communique dès qu’on est en ligne, sans même le savoir ni le vouloir. Une grande partie de la communication algorithmique est *implicitement stigmergique*. Mais nous pouvons aussi communiquer par l’intermédiaire de notre environnement commun de manière consciente: c’est la différence entre l’humain et la fourmi. Il existe en effet des formes d’intelligence collective *explicitement*

les autres agents et qui déterminent leurs prochaines actions » voir <https://fr.wikipedia.org/wiki/Stigmergie> et <https://en.wikipedia.org/wiki/Stigmergy>

stigmergiques. Sur Github⁷ et dans le mouvement du logiciel libre, la programmation collaborative se coordonne par l'intermédiaire d'entrepôts communs. Une encyclopédie en ligne, comme Wikipedia, rassemble d'immenses communautés internationales de rédacteurs et d'éditeurs. Les dispositifs d'apprentissage collaboratif en réseau et les cours ouverts de type connexionniste transcendent l'éducation à distance classique. Les étudiants des MOOC (*massive open online course*) connexionnistes contribuent à produire leur matériel pédagogique commun sur une multitude de plateformes interconnectées. La plupart des formes de *crowdsourcing* (création de contenu par les foules) le partage de fichiers «pair à pair» (*P2P*), les jeux en ligne massivement multi-joueurs: tous illustrent ce modèle de relation sociale dans lequel une multitude d'individus se rassemble en essaim autour d'un trésor de données communes pour – simultanément – l'enrichir et l'exploiter.

Mais le trafic humain grouillant sur les données entreposées dans les nuages n'est perçu et coordonné aujourd'hui que de manière locale, compartimentée et selon les perspectives partielles et intéressées des plateformes. Le nouvel esprit critique doit favoriser la prise de conscience globale de la communication stigmergique et des nouvelles responsabilités qu'elle implique. Elle doit aussi encourager l'usage de modes d'analyse, de visualisation, d'exploration et de recherche trans-plateforme, qui n'enfermeraient plus les gens dans des bulles cognitives mais qui leur permettrait de paramétrer leurs explorations et leurs réseaux de contacts dans le champ ouvert de la mémoire collective. On peut même rêver d'un nouveau type de plateforme qui adresserait ses utilisateurs sur l'espace du savoir et leur renverrait l'image explorable de leur intelligence collective.

⁷ *GitHub* est un médium social spécialisé dans le partage et la programmation collaborative de logiciel.

3. Les communs logiciels

Nous communiquons par l'intermédiaire de la mémoire commune et pensons dans le cours de sa transformation collaborative. C'est pourquoi il n'y aura de pensée libre que par la libération des données et des algorithmes. Énonçons quatre impératifs qui recueillent aujourd'hui les suffrages des gens bien informés : maximiser les communs de la connaissance; ouvrir les formats, les métadonnées, les données, les modèles, les algorithmes et les logiciels en général; optimiser la transparence des processus en ligne; assurer autant que possible la souveraineté pratique et légale des individus sur les données qu'ils produisent ou qui les visent.

Maximiser les communs de la connaissance signifie qu'on ne devrait pas avoir à payer pour accéder à des bases de connaissance. Cela comprend les dictionnaires, encyclopédies et manuels scolaires, les résultats de la recherche, les données produites par les institutions ou les organismes gouvernementaux, les informations nécessaires à la prévention des maladies et à l'amélioration de la santé publique, les bases de données juridiques et les ressources d'apprentissage de toutes sortes. Une bonne part de ce programme est déjà réalisé ou en cours de réalisation. Constatons toutefois que beaucoup de données qui sont officiellement publiques sont inaccessibles dans la pratique faute d'outils de recherche, de navigation et d'exploitation adéquats. On préfère d'ailleurs maintenant parler de données FAIR (*findable, accessible, interoperable and re-usable*⁸) plutôt que de données publiques. Outre qu'elle favorise directement l'apprentissage et l'accès à l'information, la "mise en commun" promeut la transversalité : le décroisement et la pollinisation croisée si utiles à la croissance de l'intelligence collective. Les communautés de création et de recherche peuvent ainsi interconnecter et recombinaison les informations disponibles, ce qui nous amène au principe d'ouverture.

Le tropisme vers l'ouverture a d'abord été expérimenté dans la communauté scientifique (comme la pré-publication rapide dans arXiv.org), le

⁸ C'est-à-dire facile à trouver, accessible, interopérable et réutilisable. Voir <https://www.force11.org/group/fairgroup/fairprinciples>

mouvement du logiciel libre, la licence *creative commons*, Wikipedia, Git⁹ (notamment utilisé par GitHub) et bien d'autres entreprises. On ajoute ici au principe de l'accès gratuit la possibilité pour des volontaires de participer de manière collaborative à la croissance des communs. L'ouverture a été adoptée par de nombreux gouvernements et grandes industries. "Ouvert par défaut" devient la norme. Donnons en exemple Microsoft (le M final des GAFAM), qui a fait sa fortune sur la commercialisation de logiciels propriétaires fermés mais dont les principaux revenus viennent maintenant de services nuagiques, et qui a annoncé dès 2018 que tout son portefeuille de brevets logiciels basculait dans l'open source¹⁰. Pourquoi ce tropisme vers l'ouverture? Parce qu'il maximise l'amélioration des biens et des services, favorise la confiance et soutient l'engagement collaboratif. Il peut être appliqué aux formats de données, aux systèmes d'exploitation, aux modèles, aux algorithmes et même au matériel¹¹. L'ouverture s'applique également aux métadonnées, taxinomies, ontologies et architectures d'information.

Enfin, la transparence des acteurs et des processus forme la base de la confiance et conditionne l'authenticité d'un dialogue sans lequel s'étirole la vie démocratique d'une cité. Afin de limiter la corruption et les manipulations – qui s'alimentent évidemment de l'opacité – les gens et les données doivent être retraçables et auditables, y compris dans le cas des administrateurs des plateformes et des responsables des gouvernements. Dans la continuité de la tradition d'émancipation des lumières, la transparence ne doit pas être unidirectionnelle (les sujets dominés sous le regard panoptique d'une caste dominante qui reste dans l'ombre) mais bel et bien réciproque¹², sans distinction entre dirigeants et dirigés. Comment pourrions-nous viser une intelligence collective réflexive – permettant aux équipes et communautés de toutes tailles

⁹ Git est un logiciel de gestion de version décentralisé mis au point en 2005 par Linus Torvalds, auteur du noyau LINUX. Il est principalement utilisé pour gérer la production collaborative de logiciel mais peut être utilisé pour n'importe quel type de texte.

¹⁰ Voir <https://www.zdnet.com/article/microsoft-open-sources-its-entire-patent-portfolio/>

¹¹ Voir https://en.wikipedia.org/wiki/Open-source_hardware

¹² David Brin avait attiré l'attention sur ce point dès 1998 avec son livre *The transparent Society*, paru chez Perseus books.

d'observer et de comparer leurs activités cognitives – si nous n'affirmons pas la transparence comme une des valeurs principales du nouvel esprit critique?

A ces arguments de bons sens et d'éthique en faveur de la transparence j'ajoute, pour leur faire contrepoids, deux remarques dont l'une relève du réalisme politique et l'autre de la banalité anthropologique. Premier point, selon la théorie¹³ et la pratique de Julian Assange, les pouvoirs contemporains ont une dimension "conspirationniste": leurs réseaux d'échanges d'informations secrètes leur permettent de penser et d'agir à l'insu de leurs adversaires et du public. Pour les vaincre, leurs opposants doivent interrompre ou manipuler leur communication interne et surtout organiser des fuites vers l'extérieur. Illustrée par l'action de Wikileaks et d'autres réseaux de journalistes et de lanceurs d'alerte, la guérilla informationnelle mène une guerre d'attrition contre une conspiration ennemie qu'elle force à épuiser ses ressources pour maintenir la sécurité de son réseau. Le secret devenant de plus en plus coûteux, les pouvoirs adoptent des contre-stratégies qui minimisent l'occultation classique au profit de modes de coordination interne et de stupéfaction de l'adversaire (la communication trumpienne en est un exemple) qui sont d'autant plus efficaces qu'elles se déroulent au grand jour. Second point, la pression à la transparence sur les individus et les organisations les plus puissants est un phénomène aisément observable. Les grandes compagnies cotées en bourse doivent publier régulièrement leurs résultats et sont scrutées par une multitude d'analystes et d'activistes. Au sommet des états-nations riches et démocratiques, les présidents, les ministres, voire l'ensemble des membres de la classe politique et journalistique exposent leurs idées et sentiments sur les médias sociaux et vivent dans une sorte de télé-réalité permanente où leurs actes sont passés au crible par le public¹⁴. Tandis que les faibles et les démunis restent dans l'obscurité et que les pouvoirs maffieux ou totalitaires cultivent le secret, la transparence assumée pourrait signaler l'éclat des puissances montantes.

¹³ https://en.wikipedia.org/wiki/Julian_Assange_-_Writings

¹⁴ La chose n'est pas nouvelle. Agamemnon dit à la scène 5 de l'acte I de l'Iphigénie de Racine: "Triste destin des rois! Esclaves que nous sommes _ Et des rigueurs du sort et des discours des hommes! _ Nous nous voyons sans cesse assiégés de témoins _ et les plus malheureux osent pleurer le moins!" Je souligne. Racine reprend d'ailleurs ici l'Iphigénie en Aulide d'Euripide

Les nécessités de toutes sortes et le désir d'attention qui valorisent la transparence doivent être tempérés par le respect de la vie privée des citoyens. Luciano Floridi¹⁵ a bien remarqué que si la transmission d'information était encore coûteuse aux XIXe et XXe siècles, c'est au contraire la limitation de la communication qui devient onéreuse dans l'espace numérique. Si bien que l'équilibre entre vie publique et vie privée qui s'était stabilisé durant quelques siècles est remis en question à l'âge des téléphones intelligents, des médias sociaux, de l'Internet des choses, du trafic des données dans les nuages... et de la collusion entre les plateformes et les services de renseignement. Depuis quelques années, l'ancienne problématique de la vie privée fait d'ailleurs place à la nouvelle revendication de la souveraineté des individus sur les données qu'ils produisent et qui les concernent. A qui nos messages sur les médias sociaux sont-ils en fin de compte diffusés? Savons-nous exactement qui aura le droit d'exploiter nos données, et comment, lorsque nous acceptons les conditions d'utilisation des plateformes? Pouvons-nous nous prévaloir du droit à l'oubli et demander à Google d'effacer de ses résultats des sites qui diffusent des informations calomnieuses sur notre compte ou qui révèlent des fautes depuis longtemps passées et expiées? A ces craintes s'ajoutent de nouveaux espoirs. Les informations cruciales qui concernent notre état civil, notre situation financière ou juridique, nos titres de propriété, nos contrats, nos diplômes ou notre santé sont aujourd'hui confisquées par des institutions cloisonnées et opaques datant de l'ère du papier et face auxquelles l'individu se trouve en situation de dépendance. Pourquoi ne pas placer le citoyen du nouvel espace politique au centre et – une fois les informations garanties par une procédure acceptée de tous – lui accorder l'autogestion de ses données personnelles tout en fluidifiant les transferts d'information. Fini le harcèlement bureaucratique des innombrables formulaires incompatibles issues d'administrations voisines qui réclament les mêmes informations sous d'autres rubriques et les mêmes pièces à fournir pour la millième fois! C'est toute la promesse de "l'identité souveraine" appuyée sur la chaîne de blocs ou d'autres crypto-technologies¹⁶. Cette identité

¹⁵ *The fourth révolution, how the infosphere is reshaping human reality*, Oxford UP, 2014.

¹⁶ Voir par exemple: « The time for self-sovereign identity is now » <https://medium.com/learning-machine-blog/the-time-for-self-sovereign-identity-is-now-222aab97041b>

numérique autogérée nous libèrerait du poids de prouver sans cesse à nouveau que nous sommes nous-mêmes. Elle nous défendrait aussi bien contre l'oppression issue de la fragmentation bureaucratique des vieux états que contre l'exploitation informationnelle des nouvelles plateformes centralisées.

En somme, les nouvelles techniques d'autogestion de l'identité personnelle pourraient redéfinir la relation privé-public au profit d'une maîtrise directe de ses données par le citoyen qui déciderait à qui, quand et pour quels usages ses informations crypto-certifiées sont transmises. Selon le même esprit, mais avec d'autres techniques, Tim Berners-Lee, le concepteur du World Wide Web, a récemment lancé le protocole "Solid" qui a précisément pour objectif de rendre aux individus le contrôle sur leurs données et les applications qu'ils utilisent¹⁷. On retrouve le même souci de décentralisation et de réappropriation de leurs données par les utilisateurs dans des projets comme Holochain (une plateforme pour des applications décentralisées)¹⁸, Scuttlebutt (un réseau social décentralisé pair-à-pair)¹⁹ et bien d'autres. Mais dans la perspective d'intelligence collective réflexive qui est la nôtre, il ne faudra pas seulement accorder aux personnes un contrôle sûr et pratique sur leurs propres données mais aussi leur renvoyer l'image mouvante que leurs données agrégées et anonymisées dessinent ensemble – selon les communautés et les réseaux sociaux – grâce à des techniques d'analyse et de visualisation enfin démocratisées.

Notons que l'augmentation des communs de la connaissance, la prime à l'*open source*, l'exigence de transparence des processus et des algorithmes comme la souveraineté des individus sur leurs données font déjà partie de la culture politique en émergence. Même s'il ne sont pas encore réalisés, ces principes sont aujourd'hui largement acceptés et – sauf catastrophe ou régression – s'imposeront de plus en plus dans l'avenir. Le nouvel esprit critique encourage tous les participants de l'espace public numérique à créer, commenter, catégoriser, évaluer et analyser son contenu en connaissance de cause et avec pleine capacité d'agir.

¹⁷ Voir : <https://www.inrupt.com/blog/one-small-step-for-the-web>

¹⁸ <https://holochain.org/>

¹⁹ <https://www.scuttlebutt.nz/>

4. La curation des données

La communication stigmergique, comme la mise en commun des algorithmes et des données, ne sont pas des buts en soi. Un des objectifs de la nouvelle pensée critique est précisément de mettre entre les mains du plus grand nombre des outils et des méthodes pour *extraire de la connaissance* à partir des données.

Or il est impossible d'apprendre des données sans les avoir filtrées, classées et évaluées au préalable. Bien mieux, ces activités constituent en elles-mêmes une puissante méthode d'apprentissage collaboratif.

Comme Monsieur Jourdain faisait de la prose, tout le monde fait aujourd'hui de la curation de données²⁰ sans le savoir. Sur les plateformes de médias sociaux comme Facebook, Pinterest, Instagram ou Twitter, mais aussi dans une multitude d'applications en ligne plus spécialisées comme Evernote, Scoop.it ou Diigo ou Pocket, les utilisateurs font référence à des données (textes, images, vidéos, musique...) qu'ils accompagnent de commentaires, de hashtags classificateurs et de diverses formes d'évaluations et d'émoticons. Ces *posts* s'accumulent dans des collections personnelles ou communautaires, apparaissent sur les fils d'autres utilisateurs et sont réexpédiées *ad libitum* avec d'éventuels changements de commentaires, de hashtags et d'appréciations émotionnelles. Les *posts* deviennent eux-mêmes des données qui peuvent à leur tour faire l'objet de références, de commentaires, de marquage affectif, de recherche et d'analyse. Les médias sociaux nous proposent des outils perfectionnés de gestion de base de données, avec des algorithmes de fouille, d'apprentissage machine, de reconnaissance de forme et de filtrage collaboratif qui nous aident à naviguer parmi la masse du contenu et les foules d'utilisateurs. Mais l'alimentation de la base tout comme la catégorisation et l'évaluation des données sont aujourd'hui encore principalement à notre charge d'utilisateurs humains.

Le mot *curation*, employé d'abord en anglais pour désigner l'activité d'un commissaire d'exposition dans l'univers des galeries d'art et des musées, a été

²⁰ On dit aussi de la curation de contenu.

récemment généralisé à toutes les activités de collection d'information. L'étymologie latine du mot évoque le soin médical (la cure) et plus généralement le souci. Nous entrons dans une société datacentrique, où le souci des données, l'activité qui consiste à collectionner, organiser et exploiter des données pour soi et pour les autres devient cruciale. Et l'enjeu ultime de la curation de données n'est autre que la production et le partage de connaissances.

Je vais maintenant évoquer un certain nombre de sphères d'activité dans lesquelles la maîtrise de la curation collaborative de données commence à s'imposer comme une compétence essentielle : la conservation des héritages, la recherche scientifique, l'apprentissage collaboratif, la production et la diffusion des nouvelles, le renseignement à sources ouvertes et la gestion des connaissances.

La plus grande part de la mémoire culturelle – les communs de la connaissance – deviendra bientôt accessible gratuitement. Les responsables des archives, bibliothèques, médiathèques et musées collectionnent depuis des siècles des artefacts porteurs d'information et les organisent de telle sorte que leur public puisse les retrouver et les consulter. C'est dans ce milieu professionnel qu'est d'abord apparu la distinction entre données et métadonnées. Du côté des données, les documents physiques sont posés sur des étagères. Du côté des *métadonnées*, un fichier permet de rechercher les documents par auteurs, titres, sujets, disciplines, dates, etc. Le bibliothécaire fabrique une fiche, voire plusieurs fiches, pour chaque document qui entre dans son royaume et le lecteur fouille dans les fiches pour en explorer le contenu et savoir où se trouvent placés les livres qu'il veut lire. Souvent, les documents contiennent leurs propres métadonnées, comme lorsque les livres spécifient leurs auteurs, titres, maison d'édition, date et lieu de publication, sujet, numéro ISBN, etc. Parmi les systèmes de métadonnées, les *langages documentaires* se présentent comme des normes de description des sujets. Sans l'appareillage des métadonnées et les principes d'organisation qui les sous-tendent il serait impossible d'exploiter les informations contenues dans une bibliothèque. Depuis la fin du XXe siècle, le monde des archives, des bibliothèques et des musées connaît une grande transformation. La numérisation fait converger toutes les

informations dans le médium algorithmique et cette unification met cruellement en évidence la disparité et l'incompatibilité des systèmes de classification en usage²¹. De plus, les principaux systèmes de métadonnées ont été conçus et utilisés à l'époque de l'imprimerie, ils n'exploitent donc pas les nouvelles possibilités de calcul automatique. Finalement, les flots d'information ont tellement crû qu'ils échappent à toute possibilité de catalogage classique par un petit nombre de professionnels. Depuis quelques années, les musées et bibliothèques numérisent et mettent en ligne leurs collections en faisant appel au *crowdsourcing*, c'est-à-dire à l'intelligence collective des internautes, pour catégoriser les données. Cette curation collaborative de données brouille la distinction entre curateurs et utilisateurs tout en manifestant la diversité des points de vue et des intérêts du public. Par ailleurs, une multitude de sites puisant leurs données dans le Web ouvert, et souvent indépendants des institutions classiques de préservation des héritages culturels, permettent aux amateurs d'art ou aux bibliophiles de partager leurs goûts et leurs trouvailles, de se regrouper par sensibilité et par centres d'intérêts.

De manière parallèle et complémentaire, le mouvement de la «science ouverte» vise à la constitution d'un bien commun scientifique mondial comprenant non seulement les publications traditionnelles (articles, livres, rapports) mais également les données brutes et les outils logiciels ayant servi à les exploiter²². La communauté scientifique internationale communique ainsi de manière de plus en plus transparente en se rassemblant autour du bien commun numérique délocalisé qu'elle produit et dont elle se nourrit, comme les universités d'antan se rassemblaient autour de leur bibliothèque²³. On commence à entrevoir de nouvelles formes de collaboration et d'évaluation scientifique qui ne passeraient plus par des publications dans des revues (qu'elles soient imprimées ou en ligne) mais par une communication beaucoup plus fluide au sein d'un

²¹ Je parle bien ici des systèmes de classification (conceptuels) et non du format de fichier pour les métadonnées RDF (Resource Description Framework) qui tend aujourd'hui à s'imposer, notamment dans sa version SKOS (Simple Knowledge Organization System) comme un standard.

²² Voir par exemple le rapport de 2012 de la Royal Society de Londres : Science as an open enterprise, dirigé par Geoffrey Boulton, <http://royalsociety.org/policy/projects/sciencepublic-enterprise/report/>

²³ Voir sur ce point, de Michael Nielsen, *Reinventing Discovery: The New Era of Networked Science*, Princeton University Press, 2012

nouveau médium réunissant les avantages de *Research Gate*²⁴ et de *Github*²⁵, orienté vers le travail d'équipe, l'évaluation sociale, le partage des algorithmes et des données. Le besoin est bien réel. Pendant la pandémie de 2020, on a vu des dizaines de milliers de médecins utiliser une technologie aussi primitive qu'un groupe Facebook pour partager rapidement les résultats scientifiques provenant d'une multitude de sources et améliorer leur service aux malades²⁶.

La numérisation des archives et des héritages culturels, l'accessibilité des données et statistiques compilées par les gouvernements et les institutions internationales, les communications et transactions des internautes recueillies par les grandes plateformes du Web, toutes ces nouvelles sources offrent aux sciences humaines une matière première dont l'abondance défie l'imagination. Par ailleurs les blogs de chercheurs, les plateformes collaboratives spécialisées dans la collection d'articles²⁷, ou les logiciels de gestion collaborative de références comme Zotero et Mendeley transforment profondément les pratiques de recherche. Enfin, une frange croissante des professionnels des sciences humaines s'initie à la programmation et à l'usage avancé des algorithmes, produisant et partageant le plus souvent des outils *open source*. L'édition scientifique traditionnelle est en crise puisque la communication entre chercheurs n'a plus besoin de journaux imprimés. Chaque plateforme en ligne propose ses propres méthodes d'appréciation des publications, basées sur un traitement automatisé des interactions sociales, ce qui remet en question les modes classiques de filtrage et d'évaluation des articles. Certes, le problème posé par l'incompatibilité des plateformes et des systèmes de catégorisation reste à résoudre. Il subsiste donc quelques obstacles à franchir, mais tout est en place pour que la curation collaborative de données s'impose comme l'activité centrale de la recherche en sciences humaines... et de son évaluation.

²⁴ Research Gate est un médium social spécialement conçu pour les chercheurs universitaires, comme AcademiaEdu ou Mendeley.

²⁵ GitHub est un médium social qui permet aux programmeurs de partager le code qu'ils produisent et de gérer ses versions successives.

²⁶ Kathy Peach et Ian Gray « Five ways collective intelligence can help beat coronavirus in developing countries » in The Conversation, 20 Avril 2020. <https://theconversation.com/fiveways-collective-intelligence-can-help-beat-coronavirus-in-developing-countries-136548>

²⁷ Comme *Research Gate* ou *Academia.edu*, déjà cités.

Dans le prolongement du mouvement de la science ouverte, la méta-discipline émergente des *digital humanities*²⁸ travaille à numériser et à mettre en ligne l'ensemble des archives, textes, documents divers et commentaires accumulés qui constituent le matériau des sciences humaines. Elle construit autour de cette masse documentaire une myriade d'équipes de recherche qui échangent leurs outils en vue de l'exploitation des données. Le virage numérique étant inévitable pour les sciences humaines et sociales, la notion de digital humanities a d'ailleurs de moins en moins de sens. *Humanities* ou sciences humaines devraient suffire puisque l'activité sociale contemporaine produit spontanément des torrents de données numériques et qu'il n'existe pas d'autre façon de traiter ces données qu'au moyen d'algorithmes. Les données et la puissance de calcul sont disponibles, de nombreuses études passionnantes sont publiées, mais les grandes percées scientifiques se font attendre. Les statistiques et les modèles de réseaux sociologiques utilisés en 2020 sont loin de décrire et d'expliquer la production collective de sens de manière convaincante. En effet les sciences humaines sont fragmentées en multiples disciplines concurrentes et théories incompatibles. D'autre part elles ne disposent pas de modèles mathématiques de la signification et encore moins de sa production sociale. Mais l'on peut prévoir pour les siècles qui viennent une révolution scientifique des sciences humaines comparable à la révolution qu'ont connue les sciences de la nature à l'époque moderne. Miroir des intelligences collectives, une mémoire numérique décloisonnée munie d'un système de coordonnées sémantique pourrait offrir à la communauté des chercheurs un terrain de jeu herméneutique inépuisable.

La curation collaborative de données émerge également comme une pratique essentielle dans le domaine de l'éducation. A l'époque du médium algorithmique, les connaissances évoluent vite, presque toutes les ressources d'apprentissage sont disponibles gratuitement en ligne et les étudiants sont déjà plongés dans le bain des médias sociaux. Le vieux modèle des communautés d'apprentissage s'organisant autour d'une bibliothèque ou d'un entrepôt

²⁸ Voir par exemple, sous la direction de David Berry, *Understanding Digital Humanities*, Palgrave Macmillan, 2012.

physique de documents est donc obsolète. L'apprentissage doit être de plus en plus pensé comme partiellement délocalisé, collaboratif et continu. L'ensemble de la pratique sociale acquiert une dimension d'apprentissage. Cela n'implique pas que les institutions d'enseignement classiques, école et université, ne soient plus pertinentes, bien au contraire. C'est précisément parce que l'apprentissage va puiser dans un stock de ressources pratiquement infini qu'aucune autorité transcendante ne peut plus organiser et hiérarchiser a priori que l'école a l'obligation d'entraîner les jeunes gens à l'apprentissage collaboratif et critique par le biais des médias sociaux. La fameuse littéracie numérique ne repose pas principalement sur l'acquisition de compétences techniques en informatique (qui changent rapidement), mais plutôt sur un savoir-faire socio-cognitif orienté vers l'intelligence collective et la curation collaborative de données : filtrer les contenus pertinents pour tel ou tel groupe, les catégoriser, les évaluer, consulter les données accumulées et y réfléchir, rédiger de courtes synthèses, etc. Ainsi les enseignants utilisent des plateformes de social bookmarking (partage de signets) comme Diigo pour animer leurs cours, les MOOCs connectivistes²⁹ font appel aux étudiants pour alimenter leurs ressources d'apprentissage, on trouvera une multitude de hashtags reliés à l'éducation et à l'apprentissage sur Twitter³⁰ et les groupes Facebook abritent de plus en plus de classes...

L'industrie de la nouvelle est en pleine mutation. Ce ne sont plus seulement les agences de presse et les journalistes professionnels qui produisent les *news* mais également les acteurs culturels, économiques, politiques et militaires par l'intermédiaire de leurs sites et de leurs agents dans les médias sociaux. N'oublions pas non plus les citoyens ordinaires qui prennent des photos et des vidéos grâce à leurs téléphones intelligents, qui diffusent ce qu'ils voient et ce qu'ils pensent sur toutes les plateformes et qui réagissent en temps réel aux nouvelles diffusées par les médias classiques. Quant aux journalistes engagés dans le *journalisme de données*, ils s'initient à l'exploitation statistique des bases de données ouvertes pour en retirer les synthèses et les

²⁹ Voir l'article fondateur de George Siemens « Connectivism: A Learning Theory for the Digital Age » E-learn space, décembre 2004 <https://web.archive.org/web/20100201232701/http://www.elearnspace.org/Articles/connectivism.htm>

³⁰ ...dont les miens, voir par exemple <https://twitter.com/plevy/moments>.

visualisations qui vont alimenter leurs articles. Ils repèrent les ensembles prometteurs de données fournies par les gouvernements, les organismes de production de statistiques ou d'autres institutions, voire par des *hackers* (Panama papers, Wikileaks...). Ils analysent ensuite ces données au moyen d'algorithmes appropriés, puis visualisent et expliquent les résultats à leurs lecteurs³¹. Le journaliste numérique doit posséder les mêmes compétences que l'analyste travaillant pour une agence de renseignement ou le chercheur en sciences humaines. La pandémie de 2020 a été l'occasion d'une débauche de visualisations de statistiques épidémiologiques. Elle a aussi mis en évidence l'absurdité des comparaisons lorsque les données ne sont pas recueillies et traitées de la même manière.

Du côté de la réception, la consommation des nouvelles se fait de plus en plus en ligne. La distribution passe entre les mains des objets intelligents, des médias sociaux, des moteurs de recherche et des plateformes de vente en ligne. L'entreprise californienne *deepnews.ai* propose même une sélection des «meilleures» nouvelles sélectionnée par des algorithmes d'intelligence artificielle. La manne financière de la publicité délaisse les journaux, la radio et la TV pour enrichir Google et Facebook. Les médias de masse traditionnels se sont transformés en producteurs de contenus pour les médias sociaux. L'extrait croustillant d'une interview sur Fox ou CNN est visionné sur Facebook et je lis sur Twitter tel article article du Washington Post. Journaux et chaînes de télévision ne sont plus que des sources de données comme les autres et le public a de moins en moins confiance en eux. Hélas, la distinction entre journaliste et «commentateur» de type *WuMao*, se brouille progressivement. Mais la fonction d'expression de l'opinion publique et d'autorité des journalistes professionnels devient moins nécessaire puisque chacun peut accéder directement aux sources, à savoir les messages émis par les acteurs eux-mêmes. Sur les sujets qui m'intéressent, je suis les experts de mon choix, j'écoute tous les sons de cloche et je me fais ma propre idée sans être obligé de m'en remettre à des synthèses journalistiques simplificatrices et forcément tributaires d'un agenda ou

³¹ On parle en anglais de data journalism ou data-driven journalism. Voir par exemple, libre et ouvert en ligne: The Data Journalism Handbook, Sous la direction de Jonathan Gray, Liliana Bounegru et Lucy Chandlers, 2012.

d'un récit-maître politique ou national. En somme, aussi bien les professionnels de l'information que le nouveau public critique en ligne pratiquent assidûment la curation collaborative de données.

Le domaine du renseignement économique («*business intelligence*»), politique ou militaire échappe progressivement à l'ancienne logique de l'espionnage. A part les noms et adresses des agents doubles et le détail des plans d'attaque, presque tout est désormais disponible sur internet. A qui sait chercher et lire entre les lignes, les images des satellites, les sites médiatiques, académiques, diplomatiques et militaires, sans oublier les rapports des «*think tanks*», toutes ces sources, bien utilisées, permettent de comprendre les situations et de prendre des décisions éclairées. Certes, les agents d'influence, trolls, utilisateurs masqués et robots logiciels tentent de brouiller les cartes, mais ils révèlent à la longue les stratégies des marionnettistes qui les manipulent. L'interception des signaux électroniques, l'abondance des sources d'information en ligne et les possibilités d'analyse automatique ne suppriment cependant pas la nécessité d'entretenir un personnel chargé de recueillir des informations sur place et d'interagir avec des humains. De plus, la communauté du renseignement n'a pas seulement besoin d'ingénieurs mais aussi de compétences linguistiques, culturelles et scientifiques, d'une érudition en sciences humaines, d'un certain flair pour extraire les renseignements pertinents du flot des données, d'une participation éclairée aux médias sociaux et d'un savoir-faire collaboratif. Dans le domaine en pleine expansion de l'*open source intelligence*, les agences de renseignement – comme la nuée de leurs fournisseurs d'information, d'analyse et de synthèse – coopèrent dans la production, l'échange et l'évaluation des données. Ici encore, la curation collaborative de contenu est à l'ordre du jour.

Une équipe de travail, une entreprise quelconque – qu'elle soit publique, privée ou associative – se trouve dans la nécessité de «gérer ses connaissances» pour atteindre ses buts. Le terme de *gestion des connaissances* a commencé à être utilisé vers le milieu des années 1990, au moment même où naissait le Web et alors que l'idée d'une économie basée sur les savoirs et l'innovation commençait à s'affirmer. L'un des principaux fondateurs de cette nouvelle discipline, Ikujiro Nonaka (né en 1935), s'est attaché à décrire le cycle

de création des connaissances dans les entreprises en insistant sur la phase d'explicitation et de dissémination des savoir-faire pratiques. A la suite de Nonaka, de nombreux chercheurs et praticiens ont tenté de déterminer les meilleures méthodes pour expliciter les savoirs tacites – nés de l'expérience – afin de les conserver et de les diffuser dans les organisations. Les premiers outils de gestion des connaissances étaient assez rigides et centralisés, à l'image de l'informatique de l'époque. On met en place aujourd'hui (en 2020) de véritables médias sociaux d'entreprise, dans lesquels les collaborateurs peuvent repérer mutuellement leurs compétences, créer des chaînes de nouvelles, des groupes de travail et des communautés de pratique³², accumuler des ressources et partager des données. Indépendamment des outils techniques utilisés, la gestion des connaissances est une dimension transversale de toute entreprise. Cette épistémologie appliquée inclut la conservation des savoirs et savoir-faire, le développement des compétences et des ressources humaines, l'art de créer et de diffuser les connaissances. Dans les médias sociaux d'entreprise qui supportent la gestion des connaissances, l'une des principales activités se trouve être justement la curation collaborative de données.

Dans le domaine de la communication organisationnelle et du «*knowledge management*»³³, une bonne gestion *sociale* des connaissances émerge, sur le mode *bottom-up*, des activités personnelles des membres de l'organisation pour gérer leurs propres connaissances. L'entreprise ou l'administration devient ainsi un médium facilitant trois processus complémentaires. Premièrement, ses membres, ses clients et ses partenaires accumulent des données communes, codant ainsi le savoir qui leur est utile. Deuxièmement, chacun peut accéder à ces données au moment opportun. Troisièmement, des conversations ouvertes entre les participants au réseau facilitent aussi bien la codification du savoir à destination du collectif que son appropriation personnelle en situation de travail.

³² https://fr.wikipedia.org/wiki/Communaut%C3%A9_de_pratique

³³ Le créateur de ce nouveau champ est sans conteste Ikujiro Nonaka. Une approche plus contemporaine valorisant le chaos créateur est défendue notamment par David Weinberger, voir par exemple son dernier livre: *Too Big to Know: Rethinking Knowledge Now That Facts Aren't Fact, Experts Are Everywhere, and the Smartest Person in the Room is the Room*, Basic Books, 2012.

Il existe une pratique commune à de nombreux secteurs de la culture mondiale contemporaine, pratique dont les cloisonnements sociaux et la disparité des jargons professionnels dissimulent l'unité et la transversalité. Je fais l'hypothèse que la curation collaborative de données est un des principaux supports technosociaux de l'intelligence collective à l'époque du médium algorithmique: écrire et lire... sur des flots de données. L'apprentissage machine augmente et prolonge déjà nos activités de curation collaborative: en évaluant et catégorisant des données, nous nourrissons du même coup des algorithmes neuro-mimétiques. En d'autres termes, nous enseignons peu ou prou à des programmes comment traiter les données « comme nous ». Cette automatisation de la curation va produire des flots de données catégorisées et évaluées qui alimenteront à leur tour des algorithmes d'analyse, de synthèse et de visualisation. La gestion des connaissances et l'apprentissage collaboratif prennent alors la forme d'une boucle rétroactive entre une conversation ouverte dans les médias sociaux et l'analyse automatique des données résultantes. L'augmentation de la curation collaborative par l'apprentissage machine et l'analyse des mégadonnées va inévitablement transformer la transmission des héritages culturels, la recherche en sciences humaines, la médecine, le journalisme, le renseignement et bien d'autres activités parmi lesquelles, au premier chef, l'éducation. Le nouvel esprit critique veut mettre l'intelligence artificielle au service de l'intelligence collective.

5. L'éducation à l'intelligence collective

Les médias de masse du XXe siècle – journaux imprimés, radio et télévision – diffusaient leurs messages vers un grand nombre de récepteurs passifs à partir d'un centre contrôlé par une petite minorité. En revanche, les médias sociaux du XXIe siècle permettent à leurs utilisateurs d'émettre à destination d'une audience globale. Ces relations de communication résultent en principe d'une transaction contractuelle et ouverte dans laquelle les usagers se

confèrent mutuellement des droits de lecture. Ce sont les usagers eux-mêmes qui choisissent qui reçoit leurs messages, ceux qu'ils suivent et quelles applications ils autorisent à fouiller dans leurs données³⁴. Malgré cette fiction juridique, comme nous l'avons vu au chapitre précédent, il existe une asymétrie entre les plateformes et leurs utilisateurs puisque ces derniers n'ont leur mot à dire ni sur les algorithmes changeants qui produisent leur fil de nouvelles, ni sur la censure des contenus, ni sur la sélection et la mise en valeur des commentateurs. Il reste que le citoyen de l'état plateforme se définit d'abord par ses *droits d'accès*.

Quelles données a-t-il le droit de consulter? A-t-il accès, par exemple, aux données qu'il produit par son propre comportement en ligne? A-t-il accès à des sources de nouvelles qui critiquent son gouvernement national? Quels algorithmes ses moyens financiers et ses compétences techniques le mettent-ils en mesure d'utiliser? Quels réseaux peut-il fréquenter, et qui se retrouve dans ces réseaux? Mais les droits ne suffisent pas à décrire la réalité. L'internaute se définit aussi par ses *usages effectifs*. On peut avoir le droit d'accéder à d'énormes masses de données ou à de nombreux logiciels libres sans savoir se servir des applications et sans avoir analysé le moins du monde les données. On peut se retrouver sur Twitter sans utiliser les fonctions de recherche avancée, sans savoir comment entretenir son réseau personnel ni constituer des sites d'experts par sujet, sans contribuer à l'intelligence collective et plus généralement sans exploiter les ressources de connaissance et de compétence qui fleurissent sur cette plateforme comme sur beaucoup d'autres. En démocratisant l'analyse des flots de données (*data analytics*), en éduquant à la curation, en favorisant l'interopérabilité sémantique, l'éducation du futur devra transformer un grand nombre de droits formels d'accès en exercices effectifs d'une nouvelle citoyenneté logicielle.

Chaque révolution dans la manipulation symbolique a stimulé de nouveaux développements de la connaissance et des méthodes

³⁴ Les analyses sociologiques publiées au début des années 2010 ne pouvaient pas encore observer la prise de pouvoir des plateformes. Voir par exemple Lee Rainie et Barry Wellman, *Networked: The New Social Operating System*, MIT Press, 2012. Voir également, de Manuel Castells, *Communication Power*, Oxford University Press, 2011.

d'apprentissage. *L'école* fut inventée par les scribes. A l'ombre des temples, ce n'était au début qu'une institution particulière vouée à l'entraînement professionnel d'une caste de spécialistes de l'écriture. En ces temps reculés, la pédagogie était stricte et répétitive et ressemblait à celle de l'école primaire jusqu'au milieu du XXe siècle : les châtiments corporels y étaient bien présents. Dans les cultures lettrées qui se sont développées entre le sixième siècle avant l'ère commune et la fin du Moyen-âge, *l'éducation libérale* visait des élites plus nombreuses que celles des premières écoles. Les jeunes gens étaient non seulement entraînés à lire et à écrire, mais aussi à connaître les classiques, à rédiger des poèmes, à construire des argumentations rationnelles et à produire des discours persuasifs. Les *éducations nationales* uniformisatrices de l'époque moderne ont rendu la scolarisation obligatoire, généralisé l'alphabétisation et développé l'enseignement supérieur. L'Internet et la mutation vers l'économie de l'information nous poussent vers un apprentissage tout au long de la vie, pluriculturel, collaboratif et ouvert qui s'appuie sur les biens communs de la connaissance disponibles en ligne.

Comme nous l'avons vu plus haut, les compétences en curation collaborative de données sont au coeur de la nouvelle littéracie. La figure 1 les présente de manière systématique et met en ordre les savoirs, savoir-faire et savoir-être qui supportent l'augmentation de l'intelligence collective en ligne.

	CONSCIENCE	SIGNIFICATION	MÉMOIRE
Intelligence personnelle	Gestion de l'attention - Prioriser les sujets - Choisir les sources	Interprétation - Produire des hypothèses - Analyser les données	Gestion de la mémoire - Maintenir la catégorisation - Conserver les données
Intelligence critique des sources	Critique externe - Diversifier les sources - Croiser les données	Critique interne - Identifier les catégories - Identifier les récits	Critique pragmatique - Evaluer la transparence - Identifier l'agenda
Intelligence collective	Communication stigmergique - S'adapter à la mémoire globale - S'adapter à la mémoire locale	Leberté - Faire preuve de responsabilité - Utiliser ses pouvoirs	Apprentissage collaboratif - Enternaliser le savoir tacite - Internalise le savoir explicite

Figure 1 – L'éducation à l'intelligence collective

La table est organisée en trois rangées et trois colonnes interdépendantes. La première rangée explicite les fondements de l'intelligence personnelle des données, la seconde rappelle l'indispensable travail critique sur les sources et la troisième détaille les compétences nécessaires à l'émergence d'une intelligence collective augmentée par les algorithmes. L'intelligence personnelle et l'intelligence collective travaillent ensemble et ni l'une ni l'autre ne peuvent se passer d'intelligence critique! Les colonnes évoquent trois dimensions complémentaires de la cognition: la conscience réflexive, la production de signification et la mémoire. Aucune d'elles ne doit être tenue pour acquise et toutes peuvent faire l'objet d'entraînement et de perfectionnement. Dans chaque case, l'item du haut pointe vers un exercice plus abstrait et intangible tandis que celui du bas indique une mise en oeuvre plus concrète et située de la compétence. Je vais maintenant commenter la figure 1 rangée par rangée.

La notion d'intelligence personnelle doit ici s'entendre au sens d'une *compétence cognitive* individuelle. Mais elle tire également vers la signification du mot «*intelligence*» en anglais. Dans ce dernier sens, elle désigne la capacité d'un individu à mettre en place son propre système de renseignement. La gestion de l'attention ne concerne pas seulement l'exercice de la concentration et l'art complémentaire d'éviter les distractions. Elle inclut aussi le choix réfléchi de priorités d'apprentissage et le discernement de sources d'information pertinentes. La personne elle-même doit décider de ce qui est pertinent et de ce qui ne l'est pas selon ses propres critères et en fonction des priorités qu'elle s'est donné. Quant à la notion de *source*, est-il besoin de souligner ici que seuls les individus, les groupes et les institutions peuvent être ainsi qualifiés. Seuls donc ils méritent la confiance ou la méfiance. Les médias sociaux ne sont en aucun cas des sources mais plutôt des plateformes de communication. L'interprétation des données relève également de la responsabilité des curateurs. Avec tous les algorithmes statistiques et les outils d'analyse automatique («*big data analytics*») du monde, nous aurons encore besoin d'hypothèses causales, de théories et de systèmes de catégorisation pour soutenir ces théories. Les corrélations statistiques peuvent suggérer des hypothèses causales mais elles ne les

remplacent pas. Car nous voulons non seulement prédire le comportement de systèmes complexes, mais aussi les comprendre et agir sur la base de cette compréhension. Or l'action efficace suppose une saisie des causes réelles et non seulement la perception de corrélations. Sans les intuitions et les théories dérivées de notre connaissance personnelle d'un domaine, les outils d'analyse automatique de données ne seront pas utilisés à bon escient. Poser de bonnes questions aux données n'est pas une entreprise triviale! Finalement, les données collectionnées doivent être gérées au plan matériel. Il nous faut donc choisir les bons outils d'entreposage dans les «*nuages*» et savoir manipuler ces outils. Mais la mémoire doit être aussi entretenue au niveau conceptuel. C'est pourquoi le bon curateur est capable de créer, d'adopter et surtout de maintenir un système de catégorisation qui lui permettra de retrouver l'information désirée et d'extraire de ses collections la connaissance qui lui sera utile.

Passons maintenant à l'intelligence critique. Cette dernière porte essentiellement sur la qualité des sources. Elle exige d'abord un travail de critique «externe». Nous savons qu'il n'existe pas d'autorité transcendante dans le nouvel espace de communication. Si nous ne voulons pas être trompé, abusé, ou aveuglé par des oeillères informationnelles, il nous faut donc autant que possible diversifier nos sources. Notre fenêtre d'attention doit être maintenue bien ouverte, c'est pourquoi nous nous abonnerons à des sources adoptant divers points de vue, récits organisateurs et théories. Cette diversité nous permettra de croiser les données, d'observer les sujets sur lesquelles elles se contredisent et ceux sur lesquelles elles se confirment mutuellement. L'évaluation des sources demande également un effort de décryptage des identités: c'est la critique «interne». Pour comprendre la nature d'une source, nous devons reconnaître son système de classification, ses catégories maîtresses et son récit organisateur. En un sens, une source n'est autre que le récit autour duquel elle organise ses données: sa manière de produire du sens. Finalement l'intelligence critique possède une dimension pratique. Cette «troisième critique» est la plus dévastatrice parce qu'elle compare le récit de la source avec ce qu'elle fait réellement. Je vise ici ce qu'elle fait en diffusant ses messages, c'est-à-dire l'effet concret de ses actes de communication sur les

conversations en cours et l'état d'esprit des participants. Je vise également les contributions intellectuelles et esthétiques de la source, ses interactions économiques, politiques, militaires ou autres telles qu'elles sont rapportées par d'autres sources. Grâce à cette mémoire nous pouvons noter les contradictions de la source selon les moments et les publics, les décalages entre son récit officiel et les effets pratiques de ses actions. Enfin, plus une source se montre transparente au sujet de ses propres sources d'informations, de ses références, de son agenda et de son financement et plus elle est fiable. Inversement, l'opacité éveillera les soupçons. En somme, l'intelligence critique – une part essentielle de l'esprit critique – comprend la capacité d'évaluer la transparence d'une source d'information, de vérifier son exactitude par recoupements et de décrypter ses présupposés.

L'intelligence collective dont il va être question maintenant n'est pas une solution miracle mais un savoir-faire à cultiver qui présuppose et renforce en retour les intelligences personnelles et critiques. On se souvient de la définition de la stigmergie: il s'agit d'un mode de communication dans lequel les agents se coordonnent et s'informent mutuellement *en modifiant un environnement ou une mémoire commune*. Dans le médium algorithmique, la communication tend à s'établir entre des pairs qui créent, catégorisent, critiquent, organisent, lisent, promeuvent et analysent des données au moyen d'outils logiciels. Il s'agit bien d'une communication stigmergique parce que, même si les personnes dialoguent et se parlent directement, le principal canal de communication reste une mémoire commune que les participants exploitent et transforment ensemble. Il est utile de distinguer entre les mémoires locale et globale. Dans la mémoire *locale* de réseaux ou de communautés particulières, nous devons prêter attention à des contextes et à des histoires singulières. Il est également recommandé de tenir compte des contributions des autres participants, de ne pas aborder des sujets non-pertinents pour le groupe, d'éviter les provocations, les explosions d'agressivité, etc. Quant à la mémoire *globale*, il faut se souvenir que chaque action dans le médium algorithmique réorganise – même de façon infinitésimale – la mémoire commune. Nous créons notre environnement symbolique de manière collaborative. Le bon agent humain de l'intelligence collective gardera

donc à l'esprit que ses actions en ligne contribuent à l'information des autres agents. La *liberté* dont il est question dans la figure 1 se présente comme une dialectique entre pouvoir et responsabilité. Le pouvoir recouvre notre capacité à créer, évaluer, organiser, lire et analyser les données, notre aptitude à faire évoluer la mémoire commune à partir de la multitude distribuée de nos actions. La responsabilité se fonde sur une conscience réfléchie de notre puissance collective, conscience qui informe en retour l'orientation de notre attention et le sens que nous donnons à l'exercice de nos pouvoirs.

Enfin, l'apprentissage collaboratif est le processus cognitif majeur de l'intelligence collective et le principal bénéfice social des habiletés en curation de données. Afin de bien saisir ce processus, souvenons-nous de la distinction entre savoirs tacites et savoirs explicites. Les savoirs tacites recouvrent ce que les membres d'une communauté ont appris dans des contextes particuliers, les savoir-faire internalisés dans les réflexes personnels à partir de l'expérience. Les savoirs explicites, en revanche, sont des récits, des images, des données, des logiciels ou d'autres ressources documentaires, qui sont aussi clairs et décontextualisés que possible, afin de pouvoir être partagés largement. L'apprentissage collaboratif enchaîne deux mouvements. Le premier consiste à traduire le savoir tacite en savoir explicite pour alimenter une mémoire commune. Dans un second mouvement, complémentaire du premier, les participants exploitent le savoir explicite et les ressources d'apprentissage disponibles dans la mémoire commune afin d'adapter ces connaissances à leur contexte particulier et de les intégrer dans leurs réflexes quotidiens. Les participants aux processus d'intelligence collective sont potentiellement des étudiants lorsqu'ils internalisent un savoir explicite et ils peuvent se considérer comme des enseignants lorsqu'ils mettent des savoirs explicites à la disposition des autres. Ce sont des *pairs* (voir la figure 2) qui travaillent dans un champ de pratique commun. Ils transforment autant que possible³⁵ leur savoir tacite en savoir explicite et travaillent en retour à traduire la partie des connaissances explicites qu'ils veulent acquérir en savoir pratique personnel. Dans le médium

³⁵ « Autant que possible » parce qu'il est impossible d'explicitier *tout* le savoir tacite, comme l'a bien montré Michael Polanyi.

algorithmique, le savoir explicite prend la forme de données catégorisées et évaluées. Quant à la spirale de transformation réciproque des savoirs tacites en savoirs explicites, elle prend place dans des médias sociaux, selon les règles d'une conversation créative et civilisée.

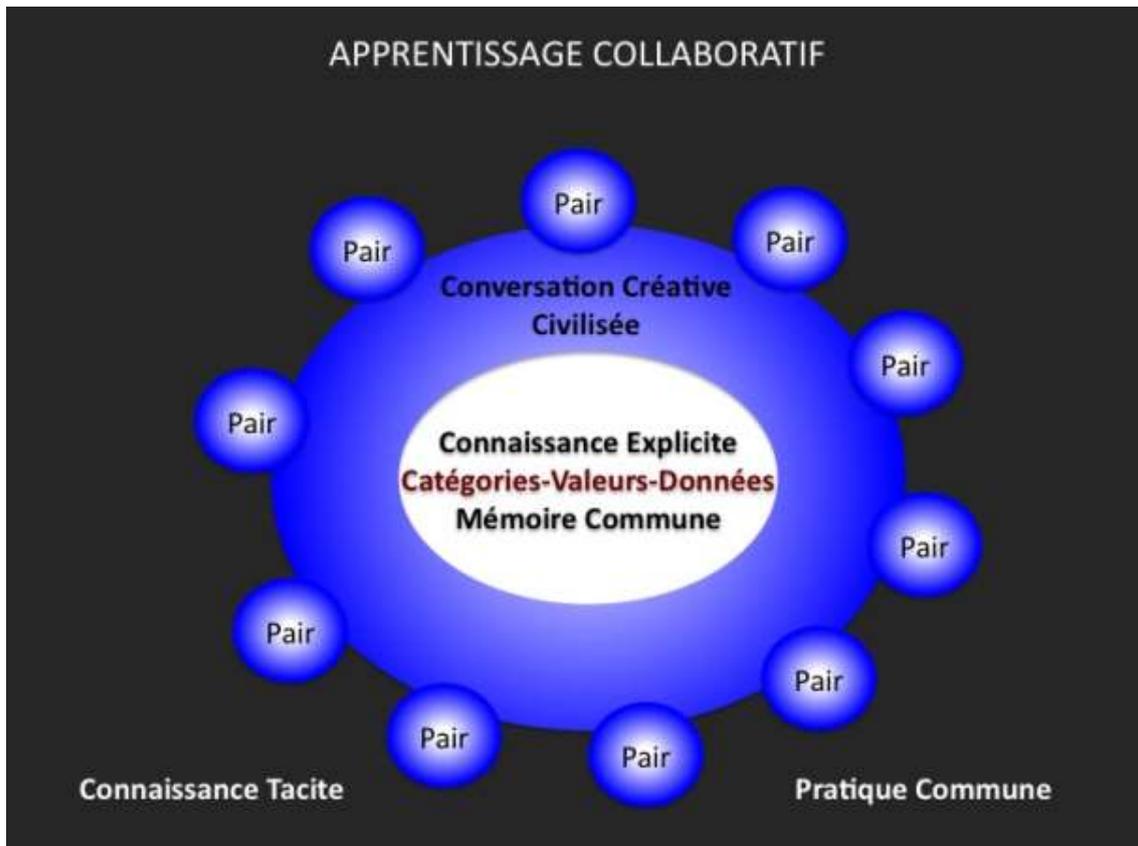


Figure 2 – La transformation de la connaissance tacite en connaissance explicite – et *vice versa* – passe par une conversation créative entre pairs dans les médias sociaux

L'école du futur se conçoit comme une entreprise de communication et d'intelligence collective soutenue par les médias sociaux. Les enseignants construisent avec leurs étudiants des communautés ouvertes de pratique, de dialogue et de réflexion utilisant des plateformes gratuites déjà utilisées par les élèves et le grand public. Les contacts transversaux pertinents avec des experts et d'autres communautés d'apprentissage se produisent alors de manière organique. Les élèves apprennent à apprendre, à dialoguer et à diffuser leurs connaissances et compétences de manière collaborative. L'alimentation en continu d'algorithmes d'analyse des mégadonnées et l'observation réflexive des

résultats font partie intégrante de leur expérience. La production des élèves guidés par leurs enseignants contribue à augmenter les données d'entraînement pour les machines. En apprenant, et en apprenant à apprendre, les communautés d'étudiants explorent et cartographient l'espace du savoir de leur génération.

6. Théoriser pour agir

Pour finir ce chapitre sur le nouvel esprit critique, je voudrais élaborer sur un thème dont j'ai déjà dit quelques mots plus haut: l'importance de la pensée théorique et de la construction de modèles causaux pour équilibrer la montée des traitements statistiques sur les mégadonnées.

Les données brutes n'ont jamais été aussi abondantes et disponibles, tout comme l'offre d'applications et de services de traitements statistiques. Que ce soit dans les sciences, l'économie ou la politique, les activités humaines sont de plus en plus fondées sur la curation et l'analyse d'énormes flots de données numériques³⁶. L'analyse des mégadonnées remplace les sondages et les enquêtes partielles par la prise en compte de l'ensemble des informations disponibles. Hier, seuls les gouvernements, les grandes entreprises et les centres de recherche académiques avaient les moyens financiers et les compétences pour extraire quelque information pertinente d'énormes flux de données. Aujourd'hui cette pratique s'est largement répandue. Peut-être que, dans le futur, des outils gratuits et manipulables à partir de simples tablettes permettront de produire, d'acheminer, de fouiller et d'analyser collaborativement l'océan des données. Quoiqu'il en soit, il est tentant de confier l'interprétation des données à un outil particulier, à un groupe de consultants ou à notre plateforme favorite sans trop se poser de questions.

³⁶ Je traduis ici par « flots de données » l'anglais « big data ». On lira par exemple sur ce thème l'ouvrage de Viktor Mayer-Schonberger, et Kenneth Cukier, qui faisait le point sur la question en 2013 : *Big Data: A Revolution That Will Transform How We Live, Work, and Think*, Eamon Dolan/Houghton Mifflin, 2013. Chaque année qui passe confirme l'importance de ce nouveau secteur d'activité.

Certains observateurs enthousiastes du traitement statistique des « *big data* », comme Chris Anderson, (l'ancien rédacteur en chef de *Wired*), d'ailleurs récemment rejoint par David Weinberger³⁷, se sont empressés de déclarer que les théories scientifiques – en général! – étaient désormais obsolètes³⁸. Nous n'aurions plus besoin que de mégadonnées et d'algorithmes statistiques opérant dans les centres de calcul: les théories – et donc les hypothèses qu'elles proposent et la réflexion dont elles sont issues – appartiendraient à une étape révolue de la méthode scientifique. Il paraît que les nombres parlent d'eux-mêmes. Mais c'est évidemment oublier qu'il faut, préalablement à tout calcul, déterminer les données pertinentes, savoir ce que l'on compte, et nommer les patterns émergents. Quelles sont les unités? Comment et pourquoi sont-elles découpées et catégorisées ainsi? Quant aux traitements statistiques, il en existe un grand nombre, dont chacun optimise un critère particulier et néglige plus ou moins d'autres traits. Un traitement statistique n'est jamais neutre, il résulte d'une série de choix. De plus, aucune corrélation statistique ne livre directement des relations causales. Celles-ci relèvent nécessairement d'hypothèses qui *expliquent* les corrélations mises en évidence par les calculs statistiques. Si l'analyse des données a été faite pour éclairer des décisions qui visent la transformation d'une situation, cette analyse doit nous révéler *des relations causales*, dont les corrélations ne sont au mieux que des indices. De simples corrélations peuvent être utiles pour repérer les tendances d'un marché, par exemple, mais elles ne seront jamais suffisantes pour fonder des décisions pertinentes en politique ou en médecine.

Sous couvert de pensée révolutionnaire, Chris Anderson et ses émules ressuscitent la vieille épistémologie positiviste et empiriste en vogue au XIXe siècle selon laquelle seuls les raisonnements inductifs (c'est-à-dire uniquement basés sur les données) sont scientifiques. Cette position revient à refouler ou à passer sous silence les théories – et donc les hypothèses risquées fondées sur une pensée personnelle – qui sont nécessairement à l'oeuvre dans n'importe

³⁷ « Alien Knowledge. When Machines justify knowledge ». In Backchannel, 18 avril 2017, https://backchannel.com/our-machines-now-have-knowledge-well-never-understand-857a479dcc0e?mbid=synd_digg

³⁸ Voir : de Chris Anderson « The End of Theory: The Data Deluge Makes the Scientific Method Obsolete », in *Wired*, 23 juin 2008.

quel processus d'analyse de données et qui se manifestent par des décisions de filtrage, d'évaluation et de catégorisation. De fait, on ne peut initier un traitement statistique et interpréter ses résultats sans aucune théorie. Le seul choix que nous ayons est de laisser les théories à l'état tacite ou bien de les expliciter. Expliciter une théorie permet de la relativiser, de la comparer avec d'autres, de la partager, de la généraliser, de la critiquer et de l'améliorer³⁹. Cela constitue même une des principales composantes de ce qu'il est convenu d'appeler «la pensée critique», que l'éducation secondaire et universitaire est censée développer chez les étudiants.

Outre l'observation empirique, la connaissance scientifique a toujours eu à voir avec le souci de la catégorisation et de la description correcte des données phénoménales, description qui obéit nécessairement à des théories plus ou moins formalisées. En décrivant des relations fonctionnelles entre des variables, la théorie offre une prise conceptuelle sur le monde phénoménal qui permet (au moins partiellement) de le prévoir et de le maîtriser. Les données d'aujourd'hui correspondent à ce que l'épistémologie des siècles passés appelait les phénomènes. Pour continuer de filer cette métaphore, les algorithmes d'analyse de flots de données correspondent aux instruments d'observation de la science classique. Ces algorithmes nous montrent des patterns, c'est-à-dire en fin de compte des images. Mais ce n'est pas parce que nous sommes capables d'exploiter la puissance du médium algorithmique pour «observer» les données qu'il faut s'arrêter en si bon chemin. Nous devons maintenant nous appuyer sur la puissance de calcul de l'Internet pour «théoriser» (catégoriser, modéliser, expliquer, partager, discuter) nos observations, sans oublier de remettre cette théorisation entre les mains d'une intelligence collective foisonnante.

Tout en soulignant la distinction entre corrélation et causalité dans leur livre de 2013 sur les *big data*, Viktor Mayer-Schonberger et Kenneth Cukier annoncent que nous nous intéresserons de plus en plus aux corrélations et de moins en moins à la causalité, ce qui les range dans le camp des empiristes. Leur livre fournit néanmoins un excellent argument contre le positivisme statistique. Ils racontent la belle histoire de Matthew Maury, un officier de marine

³⁹ Je renvoie sur ce point aux idées de Karl Popper et Michael Polanyi dont j'ai parlé à la section ??.

américain qui, vers le milieu du XIXe siècle, agrégea les données des livres de navigation figurant dans les archives officielles pour établir des cartes fiables des vents et des courants⁴⁰. Certes, ces cartes ont été construites à partir d'une accumulation de données empiriques. Mais une telle accumulation n'aurait jamais pu être utile, ou même simplement faisable, sans le système de coordonnées géographique des méridiens et des parallèles... qui est tout sauf empirique et basé sur des données. De la même manière, ce n'est sans doute qu'en adoptant un système de coordonnées sémantiques que nous pourrions organiser et partager les flots de données de manière utile. Une harmonisation internationale et trans-linguistique des systèmes de métadonnées aiderait à traiter et comparer les données et supporterait des formes d'intelligence collective plus puissantes que celles que nous connaissons aujourd'hui.

En 2020, la plupart des algorithmes qui gèrent l'acheminement des recommandations et la fouille des données sont opaques, puisqu'ils sont protégés par le secret commercial des grandes plateformes. Quant aux algorithmes d'analyse ils sont, pour la plupart, non seulement opaques mais aussi hors d'atteinte de la majorité des internautes pour des raisons à la fois techniques et économiques⁴¹. Or il est impossible de produire de la connaissance fiable au moyen de méthodes secrètes. Bien plus, si l'on veut résoudre le problème de l'extraction d'information utile à partir du flot diluvien des *big data*, on ne pourra pas éternellement se limiter à des algorithmes statistiques travaillant sur le type d'organisation de la mémoire numérique dont nous disposons en 2020. Il faudra tôt ou tard, et le plus tôt sera le mieux, implémenter une organisation de la mémoire conçue dès l'origine pour les traitements *sémantiques*. On ne pourra apprivoiser culturellement la croissance exponentielle des données – et donc transformer ces données en connaissance réfléchie – que par une mutation qualitative du calcul.

Retenons que la «science des données» devient une composante essentielle de la compréhension des phénomènes économiques et sociaux. Plus aucune organisation ne peut s'en passer. Au risque de marcher à l'aveugle, les

⁴⁰ In *Big Data: A Revolution...* (déjà cité) p. 73-77

⁴¹ Il existe des instruments *open source* pour l'analyse de grands flots données, mais il faut la compétence d'un informaticien de bon niveau pour s'en servir.

stratégies économiques, politiques et sociales doivent s'appuyer sur l'art d'analyser les mégadonnées. Mais cet art ne comprend pas seulement les statistiques et la programmation. Il inclut aussi ce que les américains appellent la «connaissance du domaine» et qui n'est autre qu'une *modélisation* ou une *théorie causale* de la réalité analysée, théorie forcément d'origine humaine, enracinée dans une expérience pratique et orientée par des fins. Ce sont toujours les humains et leurs récits producteurs de sens qui mobilisent les algorithmes.