N° 6 | 2024
Le numérique comme objet, le numérique comme système

Le numérique comme artefact technoscientifique : une « r-évolution » paradigmatique ?

Lise VERLAET *Professeur des universités*
Institut des Technosciences de l'Information et de la Communication
UR LHUMAIN
Université de Montpellier Paul-Valéry

Édition électronique :
URL :
https://intelligibilite-numerique.numerev.com/index.php/articles/revue-6/3724-le-numerique-comme-artefact-technoscientifique-une-r-evolution-paradigmatique
Date de publication : 15/07/2024

Cette publication est sous licence **CC BY-NC-ND** (Attribution - No commercial - No derivatives).

Pour **citer cette publication** : VERLAET, L. (2024) Le numérique comme artefact technoscientifique : une « r-évolution » paradigmatique ?. *Intelligibilité du numérique*, (6). https://doi.org/10.34745/

Alliant les approches constructiviste et de la pensée complexe, cet article entend explorer le numérique en tant qu'artefact technoscientifique, pour analyser les transformations paradigmatiques qu'il introduit dans la production, la validation et la diffusion des savoirs. Le numérique, au cœur de l'écosystème technoscientifique, ne se limite plus à son rôle d'outil, mais s'affirme comme un acteur central des dynamiques entre science, technique et humanité. Cette recherche s'appuie sur une analyse comparative des paradigmes et méthodologies inhérents à deux types d'artefacts technoscientifiques. D'une part, les technologies de l'intelligence ou IA symbolique, soit une instrumentation agrégatrice de connaissances qui soutient les capacités cognitives humaines par des cadres interprétables. Et d'autre part, les technologies intelligentes ou IA connexionniste, soit une instrumentation basée sur les données et la prédiction, plus opaque et reposant sur des systèmes auto-apprenants. L'hybridation entre IA symbolique et IA connexionniste est avancée somme une solution afin de concilier puissance technologique et science éthique et réflexive, tout en préservant transparence, contextualisation et humanité.

The digital as a technoscientific artifact: a paradigmatic "r-evolution"?

Combining constructivist approaches and the framework of complex thought, this article aims to explore the digital realm as a technoscientific artifact, analyzing the paradigmatic transformations it introduces in the production, validation, and dissemination of knowledge. Digital technologies, at the core of the technoscientific ecosystem, are no longer confined to their role as tools but emerge as central actors in the dynamics between science, technology, and humanity. This research relies on a comparative analysis of the paradigms and methodologies inherent to two types of technoscientific artifacts. On the one hand, technologies of intelligence or symbolic AI represent a knowledge-aggregating instrumentation that supports human cognitive capacities through interpretable frameworks. On the other hand, intelligent technologies or connectionist AI are data-driven and prediction-based systems, often opaque and relying on

self-learning mechanisms. The hybridization of symbolic and connectionist AI is proposed as a solution to reconcile technological power with ethical and reflective science, while preserving transparency, contextualization, and humanity.

Mots-clés :
Humanités numériques, Constructivisme, Pensée complexe, Artefact numérique, Digital Humanities, Technoscience, Technologies intelligentes, IA symbolique, Digital artifact, Transformation paradigmatique, Paradigme des données, Technologies de l'intelligence, IA connexionniste, Épistémologie computationnelle, Complex thought,, Paradigmatic transformation, Technoscience, Data paradigm, Technologies of intelligence, Intelligent technologies, Symbolic AI, Connectionist AI, Computational epistemology

---

1. Introduction

Le concept de technoscience va bien au-delà de la simple interaction entre science et technique. Comme a pu le définir Gilbert Hottois (2006) ou Bruno Latour (1989), la technoscience représente un mode de production du savoir où la science et la technique forment une entité unique, indissociable et interconnectée. Dans cette perspective, la science ne peut plus être pensée sans les technologies qu'elle génère et utilise, et inversement, la technique se nourrit directement des avancées scientifiques. La technoscience fait également écho à ce que Gaston Bachelard nommait la phénoménotechnique. « *La science contemporaine fait entrer l'homme dans un monde nouveau. Si l'homme pense la science, il se renouvelle également en tant qu'homme pensant [...]* » (Bachelard, 1953, p. 2). Leur omniprésence dans les pratiques scientifiques, mais aussi dans des champs qui ne se conçoivent qu'à travers elles, suscite autant d'enthousiasme que d'inquiétudes. Entre utopies et dystopies, elles questionnent les sociétés sur leurs impacts éthiques, sociaux et culturels. Mécanismes intellectuels et info-communicationnels salvateurs pour parvenir à une éthique technoscientifique. Parmi ces révolutions technologiques, le numérique s'impose comme une force centrale, en particulier dans les sciences du sensible, et redéfinit le cadre des interactions entre l'humain, la science et la technique.

Internet est devenu en quelques décennies le plus grand corpus multimodal jamais créé par l'Humanité. Pour étudier ces mégadonnées, il a été nécessaire de penser des technologies et techniques qui puissent les explorer, les collecter, les stocker, les traiter et les analyser. Le besoin d'un outillage technoscientifique propre à l'étude de corpus numérique a permis l'émergence des « humanités numériques », lesquelles peuvent

être définies comme une transdiscipline unifiée par un ensemble convergent de pratiques et de méthodes spécifiques au support d'inscription qu'est le numérique et qui intègre l'ensemble des Sciences Humaines et Sociales, des Arts et des Lettres. Les Humanités numériques désignent donc des dispositifs technoscientifiques, des technologies de l'intelligence par excellence, qui s'appuient sur l'ensemble des paradigmes, savoir-faire et connaissances propres à ces disciplines, mais également à même de pouvoir les faire évoluer. Comme nous l'analyserons de manière systémique tout au long de cet article, ces changements paradigmatiques et technologiques ne sont pas exempts d'incidences sur la science et la recherche scientifique, et *in fine* sur la triadique science-technique-humain.

Ces transformations, bien que porteuses de promesses, suscitent également des inquiétudes sur les incidences épistémologiques, méthodologiques et éthiques qu'elles impliquent. Le numérique, en tant qu'artefact technoscientifique, ne se limite plus à son rôle d'outil : il est devenu un acteur essentiel dans l'écosystème scientifique, influençant les processus de validation des savoirs, les pratiques disciplinaires et les choix technologiques. La question centrale de cet article porte sur la mutation des technologies de l'intelligence : ces technologies, initialement conçues pour servir l'humain, tendent-elles à inverser cette dynamique en plaçant l'humain au service des machines ? Cette inversion soulève de multiples enjeux, notamment en termes de contrôle, dépendance et appropriation des outils technologiques.

En décortiquant les conséquences systémiques de cette « révolution numérique », nous analyserons à quel point les artefacts technologiques façonnent les modes de production des savoirs. Nous explorerons les interdépendances entre science, technique et société en évaluant comment ces outils transforment les pratiques de recherche et redéfinissent les paradigmes disciplinaires dans les Sciences Humaines et Sociales. En particulier, nous mettrons en lumière l'équilibre précaire entre les potentialités offertes par ces artefacts et les défis qu'ils imposent, notamment dans la préservation d'une science réflexive et éthique.

Cet article s'articulera autour de plusieurs axes : d'abord, l'évolution historique des artefacts technoscientifiques et leur rôle dans l'émergence des Humanités Numériques ; ensuite, une analyse critique des impacts épistémologiques et sociaux du numérique sur la recherche scientifique ; enfin, une réflexion prospective sur la place de l'humain face à la montée en puissance des technologies intelligentes. Ce faisant, nous tenterons d'éclairer les débats contemporains autour du rôle du numérique dans la co-construction des savoirs et de la société.

2. Le numérique comme artefact technoscientifique

Il est évident que les évolutions techniques n'ont pas commencé avec le numérique (Rieffel, 2014). Il a fallu plusieurs millénaires pour passer de l'oral à l'écrit, dès son

apparition les modes d'échange et de partage d'informations se sont transformés, le savoir et la culture pouvaient enfin laisser une trace, être conservés et transmis aux générations suivantes pour construire, sur ces acquis, de nouvelles connaissances.

Comme le souligne Jack Goody (2007), l'écrit dépasse largement le rôle d'un simple « *mode d'enregistrement de la parole* ». Il permet non seulement à l'auteur de mobiliser une rhétorique adaptée pour structurer et articuler sa pensée, mais aussi aux lecteurs de l'assimiler et de s'en approprier les idées (Piaget, 1967), influençant ainsi leur manière de percevoir et de réfléchir sur le monde qui les entoure. Les travaux de L. Vygotski (1934) sur l'acquisition du langage écrit montrent que celui-ci exige un haut niveau d'abstraction, dépassant de loin celui de l'oralité. Citant Jackson et Head, l'auteur avance que le langage écrit constitue la clef du langage intérieur car cet écrit doit être développé au maximum afin d'en assurer l'intelligibilité par autrui. Le langage écrit accentue l'effort d'abstraction, d'analyse et de réflexivité et subséquemment participe à une auto-construction du savoir. « *L'écriture est la possibilité du jeu de l'intellect sur la langue* » (Goody, 1979). De plus, le travail collaboratif inhérent à la production-appropriation de l'écriture, cette intelligence collective, a indéniablement concouru à l'évolution de la connaissance et, ce faisant, à une (r)évolution de l'espèce humaine. L'intrication entre pensée-langage-écrit est indispensable à l'humain, or ces rapports risquent de se transformer profondément du fait des récentes évolutions technologiques et en particulier de l'anthropomorphisme des intelligences artificielles (IA).

L'épithète « scientifique » qualifie toute personne, tout objet, tout processus qui ont une part dans la construction sociale de la science (Meyriat, 2006), lesquels constituent de ce fait une partie des actants du récit scientifique. Au sein de cette co-construction scientifique, l'auteur et le lecteur se confondent, à la fois producteur et utilisateur d'informations et résultats de la recherche scientifique. Autour de cette production intellectuelle s'est organisé un écosystème assurant préservation et la diffusion des connaissances et des documents, notamment les métiers de l'édition de la recherche (Chartier & Martin, 1990) et de l'information scientifique et technique (IST), dont l'influence a configuré les pratiques et usages info-communicationnels de la documentation scientifique (Callon, 1989 ; Juanals, 2003 ; Gardiès & al., 2010). Les normes et standards de l'édition et de l'IST sont d'ailleurs devenus des composants structuraux de la publication scientifique.

Comme le retrace Blair (2020), dès l'Antiquité, la surabondance de manuscrits d'érudition a entrainé des problématiques de gestion de l'information dont les 4 points cardinaux consistent à stocker, sérier (ou classer), sélectionner, et synthétiser (ou résumé). Se développent alors des méthodes (différents types d'index, de tables, de renvois, de listes, de bibliographies, etc.) pour sérier les informations et consulter la documentation scientifique ainsi que pour construire des ouvrages de références (dictionnaires, encyclopédies, thésaurus, miscellanées, collections de citations, anecdotes historiques, etc.). Ces formes de technologies de l'intellect (Goody, 2006) pour créer, gérer et utiliser ces documents et informations se sont perfectionnées au

gré des évolutions des supports techniques, de l'inscription matérielle des connaissances (Bachimont, 2004). L'artefact numérique a révolutionné le stockage et l'accès aux ressources via de multiples systèmes d'information et de recherche, les modalités de partage et de collaboration, ou encore l'émergence de nouvelles disciplines dont les Humanités Numériques (HN). Toutefois, les méthodes intellectuelles employées par l'IST pour traiter les informations et documents n'ont pas connu de transformations aussi significatives (Legroux, 2008). Qui plus est, cela n'a pas résolu la question de la réception et de la compréhension de l'information (même si de nombreux projets en HN travaillent activement en ce sens), et a même complexifié les pratiques informationnelles et les usages des SI recherche (Gardiès & al., 2009). Là encore, les récentes avancées de l'IA sont en train de bousculer l'ensemble de cet écosystème.

La science en tant que construction sociale est dépendante des communautés de chercheurs, de leur champ disciplinaire et des théories des connaissances mobilisées, lesquels apportent une dimension pragmatique et cognitive à leurs activités communes reconnaissables à travers l'analyse des produits de leur recherche (Latour, 1987 ; Wenger, 1998) mais aussi à travers les artefacts technoscientifiques utilisés pour conduire leur recherche, traiter leurs corpus, leurs données... Rappelons que ces artefacts sont à la fois des objets techniques, mais aussi et peut-être surtout des produits de la recherche scientifique, lesquels jouent un rôle central dans la manière dont les connaissances sont produites, matérialisées, et utilisées dans différents contextes. Les artefacts co-évoluent avec les humains qui les conçoivent et les utilisent, l'individuation technique (Simondon, 1958), en ce qui concerne la science l'on pourrait transposer à une individuation technoscientifique. De plus en plus, se développent des infrastructures technoscientifiques, qui au-delà des outils, configurent les systèmes sociaux. La sphère numérique a non seulement décuplé le pouvoir de l'écrit, mais nous assistons également à une véritable « *textualisation des pratiques sociales* ». Ce phénomène a été soulevé par Emmanuel Souchier, Etienne Candel et Gustavo Gomez-Mejia (2019, p.7), qui affirment que les outils technologiques opèrent un recensement du monde humain, conduisant à une naturalisation de l'ensemble du social. En cela, les artefacts technoscienfiques constituent une textualisation des pratiques scientifiques.

3. Technologie de l'intelligence

Comme le rappelle Bruno Bachimont (2017), les systèmes numériques se caractérisent par la propriété d'une double coupure. Une coupure sémantique d'une part, c'est-à-dire que les entités manipulées n'ont pour la machine aucun sens. Ce qui importe pour que les entités soient manipulées c'est qu'elles soient distinguables. Et d'autre part, une coupure matérielle dans le sens ou le numérique est un support virtuel qui accueille tous les autres types de supports. La virtualisation numérique est clairement non dépendante du support technique. Dit de manière triviale, l'artefact numérique est le nec plus ultra de la machine à calculer, il ne sait pas ce qu'il calcule mais permet toujours un résultat ! Ces potentialités sont à la fois prodigieuses et effrayantes.

Prodigieuses car cet artefact est potentiellement capable de faire tout ce que l'on veut, et effrayantes exactement pour les mêmes raisons. D'autant plus que dans l'imaginaire collectif, il existe une suprématie du chiffre et du calcul, lequel serait exempt d'erreur, conduisant de fait à une véracité du nombre. On veut des chiffres pour prouver que… D'ailleurs, étymologiquement, le numérique est la consécration de cette suprématie. Mais ceux qui manient les nombres, savent à quel point cette véracité est illusoire et surtout résolument contextuel. Élément contextuel que l'on omet parfois pour une montée en généralité… Ce qui est sûr, c'est que normativement le nombre fait loi, pour ne pas dire fait la loi, mais la loi est différente de la véracité et plus encore de la vérité. Force est néanmoins de constater comme le précise Serge Bouchardon[1] que la culture numérique tend à valoriser les processus de simulation et d'approximation au détriment des pratiques réflexives. Or en LLASHS, le numérique s'entend davantage comme une technologie de l'intelligence.

Sans refaire l'historicité du concept comme a pu le dresser Pascal Robert (2000), il convient néanmoins de préciser ce que nous entendons par technologie de l'intelligence, que nous retrouvons également dans la littérature scientifique sous la terminologie de « technologie intellectuelle ». Ce concept a été largement popularisé par les travaux de Jack Goody (2007), qui considère l'écriture comme une véritable technologie de l'intellect. En tant qu'innovation des moyens de communication, l'écriture repose sur la « raison graphique », laquelle a joué un rôle central dans l'évolution culturelle de l'humanité. Par ailleurs, Pierre Lévy (1991 ; 2011) s'inscrit dans cette réflexion en mettant en avant la « raison hypertextuelle », essentielle dans le nouvel espace d'information et de communication qu'il désigne sous le terme de cyberespace. L'écriture hypertextuelle comme nouvelle métaphore de la communication et dont les principes, méthodes et outils sont mis à l'honneur au sein des conférences internationales H2PTM conduites par un consortium de laboratoires en Sciences de l'Information et de la Communication et en particulier du laboratoire Paragraphe (UR 349) depuis la fin des années 80. Et nous ne pouvons que partager le point de vue de Jean Clément (2007) lorsqu'il souligne que l'hypertexte est une technologie intellectuelle particulièrement pertinente à l'ère de la complexité (Verlaet *et al.*, 2013). Les préceptes de l'hypertexte ne sont d'ailleurs pas sans rappeler le modèle des réseaux de neurones artificiels particulièrement plébiscités en *machine learning*. Les travaux de Bruno Bachimont (2015 ; 2020a) abordent également les technologies intellectuelles, notamment à travers la théorie des supports. Selon cette théorie, tout objet technique constitue l'inscription matérielle d'une connaissance, et toute connaissance a une origine technique, définissant la connaissance comme une capacité à agir de manière téléologique (Bachimont, 2004). Dans cette perspective, le support numérique engendre de nouvelles formes et dynamiques de représentation de l'information et des connaissances, reposant sur des manipulations calculatoires. Bruno Bachimont (2007) désigne ce phénomène comme la « raison computationnelle », qui s'appuie sur les modèles propres à l'ingénierie des connaissances.

En d'autres termes, les technologies de l'intelligence sont des solutions techniques et technologiques permettant à l'humain d'accroître ses propres facultés de raisonnement,

de mieux comprendre la complexité de nos sociétés (Morin, 1996). Pour de nombreux courants théoriques, l'intelligence serait l'aptitude d'un acteur à résoudre les situations problématiques en les appréhendant sous de nouvelles perspectives à travers la réorganisation de son système de pertinence (Schütz, 1987), de son cadre de l'expérience (Goffman, 1991). Or, comme Madeleine Akrich, Michel Callon et Bruno Latour (2006) l'ont souligné dans leurs travaux, il est nécessaire d'anticiper les comportements et les usages lors de la conception d'artefacts techniques. Les technologies de l'intelligence ne visent en aucun cas à se soustraire à l'intelligence humaine, auquel cas il s'agirait de travailler sur l'intelligence artificielle. La frontière entre technologies de l'intelligence et technologies intelligentes est mince. L'une permet à son utilisateur de construire de nouveaux schèmes de pensée, la seconde se substitue à la raison humaine par le développement d'une intelligence artificielle et, de fait, une indépendance décisionnelle sur laquelle nous nous déchargerions de toute activité mémorielle (Dyens, 2008). L'écart entre les deux concepts est toutefois subtil car permettre aux machines d'assister l'humain, consiste à leur octroyer les paramétrages nécessaires à cette fin. Une large partie des travaux menés tentent donc de retranscrire le comportement humain en algorithmes. Nous pouvons et devons donc nous interroger sur les limites à ne pas dépasser dans la délégation décisionnelle à la machine. Pour reprendre les conclusions de l'ouvrage de Bruno Bachimont « *Ingénierie des connaissances et des contenus, le numérique entre ontologie et document* » :

> *« Par conséquent, l'enjeu n'est pas de raisonner en imitant l'utilisateur pour assurer à sa place la relation au monde dans lequel résoudre le problème, mais de construire la représentation permettant à l'utilisateur d'effectuer lui-même les tâches qui lui reviennent, en abordant le réel à travers la médiation des représentations construites par les outils de l'ingénierie des connaissances. L'enjeu est de passer d'une représentation formelle des connaissances à une médiation formelle des expressions non formelles des connaissances. [...] Dans l'ingénierie des contenus, on s'intéresse à la modélisation de la forme d'expression pour permettre leur manipulation et exploitation informatiques » (Bachimont, 2007, p.243).*

Bruno Bachimont souligne que pour l'ingénierie des contenus, dans la plupart des cas, ces modélisations n'abordent pas la question du sens mais tentent d'appréhender la question de la forme en reproduisant numériquement les expressions. Or les technologies de l'intelligence doivent essentiellement être conçue en fonction des signifiances pour les utilisateurs de ces artefacts. Pour Sylvie Leleu-Merviel, la « signifiance » entend dépasser le sens rationnel généralement attribué à la « signification » tout en y intégrant la subjectivité des acteurs-interprétants, considérant que l'« *on ne peut tenir pour négligeable l'intelligence ''émotionnelle'' dont la portée est affective ou sensible, liée à un vécu et des souvenirs personnels* » (Leleu-Merviel, 2004, p.127). En ce sens, les technologies de l'intelligence sont l'émanation de co-constructions entre une intentionnalité forte portée sur l'accroissement des connaissances humaines, la prise en considération de la dimension sociale et bien sûr les dimensions techniques et technologiques. Il est également indispensable de

respecter la mémoire de ces expressions à travers une posture « généalogique » (Bachimont, 2016) laquelle assure leur traçabilité. Cette posture généalogique entend réintroduire la notion de preuve et de légitimité, qui tendent à disparaître de par la virtualisation des données qui caractérise les ressources numériques (Leleu-Merviel, 2002, p.104). Car s'il est admis que le numérique obéit à une raison computationnelle et manipule un langage conventionnel (Bachimont, 2007 ; Jeanneret, 2019), il est d'autant plus prégnant de pouvoir, par les processus d'éditorialisation, réinstaurer la confiance dans l'écrit, l'auctorialité ou d'imaginer de nouvelles formes (Verlaet & Chante, 2017).

La multimodalité des corpus voire leur transmédialité nécessite de réinterroger de manière profonde la façon dont ils sont ou pourraient être utilisés par différents publics. Les pratiques actuelles ou en devenir sont déterminantes pour concevoir des modèles de métadonnées adaptés aux usages afin que ceux-ci ne soient pas « hors sol ». Il ne s'agit pas seulement de penser à leur édition numérique (numérisation ou nativement numérique), laquelle nécessite de considérer les pratiques de production, de diffusion, de circulation, de partage, d'utilisation des ressources, d'utilisation des dispositifs, mais aussi d'envisager leur éditorialisation, leur réutilisation dans différents contextes et, ce faisant, d'explorer la reliance entre des informations hétérogènes et éparses dans la sphère numérique compte tenu de leur signifiance.

Nous le savons, Internet est devenu en quelques décennies le plus grand corpus multimodal jamais créé par l'humanité, or pour étudier les mégadonnées en découlant, il a été nécessaire de penser des technologies et techniques qui puissent les explorer, les collecter, les stocker, les traiter et les analyser. Le besoin d'un outillage technoscientifique propre à l'étude de corpus numériques a permis l'émergence des « Humanités numériques », lesquelles, si l'on mêle les définitions posées dans « *A Digital Humanities Manifesto* » de 2008 (Digitalhumanities, 2008) puis le « *Manifeste des Digital Humanities* » de 2010 (Dacos, 2011), peuvent être définies comme une transdiscipline unifiée par un ensemble convergent de pratiques et de méthodes spécifiques au support d'inscription qu'est le numérique et qui intègre l'ensemble des Sciences Humaines et Sociales, des Arts et des Lettres. Les Humanités numériques désignent donc des dispositifs technoscientifiques, des technologies de l'intelligence par excellence, qui s'appuient sur l'ensemble des paradigmes, savoir-faire et connaissances propres à ces disciplines, mais également à même de pouvoir les faire évoluer. La « révolution numérique » est intrinsèquement liée à l'essence même du numérique qui « *n'est plus seulement un outil au service de la recherche mais aussi un objet de recherche à part entière. Il est à la fois instrument, méthode, terrain et objet de recherche.* » (Bourdeloie, 2013, p.9). Ceci n'étant pas sans incidences épistémologiques, méthodologiques, éthiques et bien entendu techniques (Bourdeloie, 2013 ; Boullier, 2015 ; Bastin & Tubaro, 2018 ; Salvador, 2018 ; Bachimont, 2019 ; Milette & al., 2020). De plus, le paradigme de la mégadonnée corrélatif au numérique tend à supplanter les paradigmes précédents voire à les occulter en particulier dans les disciplines LLASHS.

Les initiatives sont nombreuses pour développer des artefacts technoscientifiques,

toutes porteuse de potentialités et d'avantages mais aussi de limites et de risques. Comme nous l'évoquions précédemment, il nous semble à cet égard intéressant de considérer la frontière fugace entre technologies de l'intelligence et technologies intelligentes (Gallot & Verlaet, 2016). Car, comme le souligne Bruno Bachimont (2020a), si les premières sont « *à rapprocher des révolutions comme celle de l'écriture* », les secondes laissent le calcul, soit « *la manipulation aveugle de symboles vides de sens* », traiter des problèmes complexes qu'elles ne peuvent saisir.

4. Une tradition reposant sur l'agrégation des connaissances

Il nous semble important de faire une distinction entre au moins deux types d'artefacts technoscientifiques. Le premier, que nous aborderons dans cette section, fait référence à une instrumentation « agrégatrice de connaissances », laquelle vise à construire des bases de données ou de connaissances spécialisées généralement à partir de corpus (re-)connus et scientifiquement validés afin d'en explorer et analyser plus aisément les contenus grâce aux fonctionnalités offertes par le numérique. Ce sont les instrumentations de type « agrégatrice de connaissances » que nous qualifions de technologies de l'intelligence puisqu'elles entendent permettre à l'homme d'accroître ses propres facultés de raisonnement, de mieux comprendre la complexité de nos sociétés (Morin, 1994). Cette instrumentation « agrégatrice de connaissances » n'est pas sans rappeler l'intelligence artificielle symbolique. Fondée sur des bases de règles explicites et des modèles logiques formels, l'IA symbolique permet de structurer, organiser et exploiter des corpus complexes à partir de leur contenu sémantique. Contrairement aux approches connexionnistes qui reposent sur des modèles statistiques et des processus d'apprentissage implicites, l'IA symbolique s'attache à la formalisation de connaissances interprétables par les humains. Cette approche épouse les ambitions du Web sémantique en facilitant l'annotation, la recherche et la mise en relation des données, tout en respectant les standards de transparence et de lisibilité essentiels aux disciplines LLASHS. Elle représente donc une alliance entre raison computationnelle et signification humaine, permettant aux chercheurs de s'approprier des outils technoscientifiques sans perdre de vue la pertinence disciplinaire et contextuelle des corpus manipulés.

D'un point de vue technique, ces technologies s'appuient notamment sur les principes du Web sémantique (Berners-Lee & Al., 2001) et ses recommandations et standards (World Wide Web Consortium), lesquels proposent d'annoter les ressources numériques à l'aide de métadonnées permettant à la fois leur caractérisation, leur interopérabilité et leur manipulation. Cependant intégrer dans un projet d'Humanités numériques la dimension Web sémantique doit reposer sur un référentiel technique et normatif complexe et chronophage à mettre en œuvre qui nécessite l'implication de professionnels de l'information scientifique et technique et/ou d'informaticiens. Or, comme le relate Gilles Bastin et Paola Tubaro (2018) ces difficultés techniques et

relationnelles sont trop souvent minimisées par les chercheurs. De fait, nombre de projets ne les ont pas pris en compte. Le « bricolage » technoscientifique en découlant, loin d'être inintéressant, empêche cependant de capitaliser à grande échelle sur l'ensemble des recherches et travaux conduits. Comme le pointe très justement le rapport EPRIST (2020), la révolution du Web sémantique peine à faire ses preuves, d'une part car ses principes sont peu voire pas connus des porteurs de projets, que les normes largement instituées de la diffusion scientifique sont largement bousculées par ces principes, et enfin, qu'il existe un manque de coordination au niveau national voire international des (grandes) infrastructures de recherche dont les tenants et aboutissants manquent de lisibilité au niveau méso ou ne sont pas encore adaptés aux projets souvent mis en œuvre au niveau micro. Le projet DEF19, acronyme de Dictionnaire des Éditeurs Français du XIXème siècle, est représentatif de ce constat (Geslot & Rebolledo-Dhuin, 2020). Outre le travail important de collecte via une approche empirique des données nécessaires à l'établissement d'une base de données exhaustive, la dimension technologique pour gérer cette base de données a très vite fait surface. DEF19 s'est alors tourné vers la TGIR Huma-Num pour trouver des solutions « clés en main » et ainsi contrer les difficultés rencontrées. Mais les solutions apportées, et par ailleurs créées dans d'autres cadres de recherche, ne correspondaient pas totalement aux enjeux propres à ce projet. Après de longs tâtonnements technologiques d'appropriation-accommodation, le projet DEF19 a fini par se plier aux potentialités offertes par ces solutions techniques, se fondre dans les contraintes imposées par l'architexte (Jeanneret & Souchier, 2002). Les acteurs de DEF19 ont toutefois trouvé d'autres fonctionnalités non envisagées, intéressantes et surtout nécessaires pour une pérennisation du projet, mais au détriment de certains de ses objectifs. Face à l'évolution constante des technologies et à l'émergence - somme toute récente - des technologies dédiées aux LLASHS, il est indéniable que tous les cas de figure sont très loin d'avoir été envisagés et que ces «bricolages» *in situ* concourent aux développements d'un nouvel outillage technoscientifique fécond même si insatisfaisant à l'heure actuelle tant les initiatives sont riches en apport de connaissances et plurielles dans leurs visées, chaque discipline souhaitant interroger un aspect particulier voire singulier des corpus et interfère ce faisant sur le traitement, la diffusion et l'exploitation de ceux-ci (Galleron & Idmhand, 2020). C'est là toute la difficulté d'une instrumentation « agrégatrice de connaissances » : répondre à tous les usages souhaités et/ou que peuvent en faire les scientifiques selon la signifiance de leur spectre disciplinaire, et sous-tend de s'interroger sur l'éditorialisation et sur ce que Gérald Kembellec (2019) nomme la sémiotique du code.

Comme le souligne Bruno Bachimont (2016, 2020b), le Web sémantique a en quelque sorte mis fin au paradigme inhérent aux bibliothèques et aux recherches documentaires pour s'intéresser aux recherches d'informations tout en mettant l'accent sur les techniques liées à l'éditorialisation des contenus. L'éditorialisation des contenus permettant à notre sens de penser les différents usages disciplinaires des corpus. Mais si le paradigme du Web sémantique (et donc de l'information) a supplanté celui des bibliothèques (et des documents), il en reste le digne représentant dans le sens où ils partagent plusieurs desseins notamment la capitalisation, l'organisation et la diffusion

des connaissances et savoirs. Le paradigme du Web sémantique tend lui aussi à être mis à défaut par le paradigme de la donnée ou plus précisément de la mégadonnée (ou *big data*). Pour Bruno Bachimont, avec le paradigme de la (méga)donnée nous avons affaire à des enregistrements infra-informationnels. Or la donnée en elle-même n'apporte que le sens dont elle est porteuse, d'ailleurs, elle ne se lit pas à l'échelle de la donnée en elle-même, mais à partir de la masse de données à laquelle elle semble appartenir. De fait, avec le Web de données, nous faisons marche arrière puisqu'*in fine* nous revenons en quelque sorte à une indexation de niveau zéro – à laquelle le Web sémantique tendait à se soustraire – couplé au principe de proximité informatique (non nécessairement sémantique) propre à la clusterisation. Rétropédalage flagrant face à l'incommensurable tâche de sémantiser le Web et la facilité arrogante du chiffre et du calcul pour essayer d'y apporter une réponse.

5. Une modernité reposant sur les données et la prédiction

C'est dans le paradigme de la donnée que s'inscrit le second type d'artefact technoscientifique, lequel correspond à une instrumentation « exploratrice de traces » soit un outillage qui va permettre de collecter, traiter et analyser les données. Ces données peuvent relever d'une démarche de *big data* ou de *smart data*[2]. Dès lors que l'on parle de mégadonnées, qui est composante fondamentale des systèmes d'IA contemporains, il est nécessaire de se remémorer le précepte des 5V.

1. Le Web a généré un **V**olume de données et d'information jusqu'alors jamais atteint. En 2018, cette masse était estimée à 33 zettaoctects et devrait être multipliée par 5 d'ici 2025 selon l'étude Data Age 2025[3]. Ce volume constitue une base essentielle pour entraîner les modèles d'IA.
2. Les sources et données numériques ne sont pas stables dans le temps, ce que l'on nomme la **V**élocité des données numériques en référence à la vitesse à laquelle elles sont générées, déplacées, voire supprimées. La vitesse de génération et de transformation des données oblige les systèmes d'IA à s'adapter en temps réel, renforçant ainsi leur pertinence dans des applications dynamiques comme la prédiction et l'analyse instantanée.
3. Les sources et données numériques se caractérisent également par une grande **V**ariété, tant de par leur hétérogénéité que leur multimodalité : mesures, données textuelles, signes, traces, images, sons… Cette hétérogénéité exige des systèmes d'IA, comme le *deep learning*, capables de traiter différents types de formats.
4. La **V**éracité des données collectées, soit la crédibilité que l'on peut accorder aux données. Derrière le volume se cache également des *fakenews*, des faux profils… L'authentification des données reste un point problématique, mais néanmoins considérée comme une marge d'erreur acceptable. Ceci reste toutefois très préoccupant pour des applications IA dans des contextes sensibles comme des

diagnostics médicaux.
5. Enfin, la **V**aleur des données, c'est-à-dire la valeur ajoutée, l'apport réel de la donnée en fonction du contexte de la recherche. L'IA est utilisée pour transformer les données brutes en informations utiles, en identifiant les relations cachées et en proposant des interprétations contextualisées.

Si l'on peut compenser ou limiter les défis sous-jacents aux « 5 V » : la vélocité par un enregistrement et stockage des données, la variété par le recours à différentes méthodes de traitements, la véracité par une vérification systématique et la valeur en sélectionnant avec attention les sources et données collectées ; c'est évidemment la notion de volume qui rend les autres tâches ardues si ce n'est irréalisable.

Pour traiter cette masse de données, les outils statistiques et métriques peuvent apparaître comme des méthodes d'analyses toutes indiquées pour les traiter (semi)-automatiquement… Or – faut-il le rappeler – le sens n'est pas lié à une donnée, la donnée en elle-même n'a que peu d'intérêt, ce n'est que lorsqu'elles sont rassemblées, organisées et interreliées que le sens apparaît, que ces données deviennent une information (Mazza, 2009). Concernant les démarches liées aux instrumentations « exploratrice de traces », l'on va collecter des données brutes (*raw data*) de manière parfois très arbitraire en faisant fi de leur nature voire en les compilant (*derived data*) ou en procédant à diverse agrégations pour construire des jeux de données (*dataset*). Nous pouvons à cet égard nous interroger, à l'instar de Françoise Paquienséguy (2019), sur la pertinence de l'exploitation de ces données « brutes » puisque celles-ci ont d'ores et déjà été manipulées - parfois à plusieurs reprises - afin d'en permettre le traitement statistique. La méthodologie inhérente aux mégadonnées repose généralement sur 4 étapes (Gonzalez, 2019) :

1. **Identification et extraction des données** : Les données à analyser peuvent être brutes, primaires ou secondaires. Les données brutes, directement collectées sur le terrain ou via des dispositifs numériques, sont non structurées et nécessitent un travail de préparation. Les données secondaires, quant à elles, sont issues de bases existantes et collectées à d'autres fins. Par exemple, des données issues de capteurs IoT ou de réseaux sociaux illustrent les sources possibles dans les mégadonnées. Cette étape est essentielle pour garantir que l'échantillon de données est représentatif des objectifs de l'étude.
2. **Nettoyage et transformation des données** : Le nettoyage des données vise à éliminer les erreurs (valeurs aberrantes, doublons) et à corriger les incohérences, telles que des formats non standardisés ou des données manquantes. Par exemple, dans une analyse des comportements des consommateurs, il est courant d'imputer des valeurs manquantes en utilisant des méthodes comme la moyenne ou la médiane des données similaires. Cette étape garantit la fiabilité de la base de données et influence directement la qualité des résultats finaux. Elle inclut également des processus de normalisation ou d'encodage pour préparer les données à des analyses automatisées.
3. **Analyse des données** : Les données nettoyées sont ensuite soumises à des

analyses complexes, souvent basées sur des approches comme l'analyse statistique multivariée, l'analyse de réseaux, ou encore des techniques lexicométriques pour explorer des corpus textuels. Par exemple, l'analyse des réseaux sociaux peut révéler des schémas d'interactions au sein d'une communauté en ligne. Dans certains cas, des techniques avancées comme le machine learning ou le deep learning sont utilisées pour détecter des patterns ou pour prédire des comportements futurs à partir des données.

4. **Présentation des données** : Une fois les analyses effectuées, les résultats doivent être présentés de manière compréhensible et visuellement attrayante. Les techniques de visualisation des données, comme les graphiques interactifs, les tableaux de bord ou les cartes de chaleur, sont de plus en plus utilisées pour rendre les résultats accessibles à un public non spécialiste. Par exemple, des outils comme Tableau ou Power BI permettent de transformer des analyses complexes en visualisations dynamiques et interactives, facilitant la prise de décision.

Ces manipulations sont particulièrement chronophages (Flach, 2012) et l'on cherche ce faisant à en automatiser les processus (Berti-Équille, 2019), ajoutant ainsi une couche algorithmique potentiellement néfaste à l'intelligibilité des situations analysées. L'on peut également distinguer deux approches concernant le traitement des mégadonnées : le *machine learning* (ou apprentissage machine) et le *deep learning* (apprentissage profond). Le *machine learning* entend automatiser le caractère prédictif de tout comportement (Reigeluth, 2018) grâce aux variables détectées par le *data-scientist*, en repérant des *patterns* d'observations. Dans ce cas, nous restons dans une démarche alliant puissance algorithmique à l'intelligence humaine. En ce qui concerne le *deep learning*, la démarche vise à entraîner la machine sur des bases de données massives afin qu'elle puisse identifier par elle-même les variables par computation. Nous laissons ainsi les pleins pouvoirs aux algorithmes de ces technologies intelligentes ou intelligences artificielles. Ces technologies intelligentes manipulent ainsi des données dont elles ignorent le sens et dont les résultats parfois obscurs nécessitent que les *data-scientists* aient recours au *storytelling* pour les interpréter, les rendre compréhensibles et valider par là-même les attentes de l'auditoire (Bachimont, 2019).

Au-delà des polémiques épistémologiques entourant des démarches inhérentes au paradigme de la donnée (Bigot & Mabi, 2017 ; Salvador, 2019), de nombreux auteurs mettent en garde les communautés scientifiques quant à leurs usages et aux résultats ainsi produits. Ainsi Guillaume Carbou (2017) fustige l'utilisation des outils statistiques pour l'étude des grands corpus de données textuelles, lesquels permettent au mieux d'avoir une lecture renouvelée pour le chercheur qui connaît son corpus. Mais il va sans dire que ces méthodes conviennent à d'autres usages. A ce titre, le dossier coordonné par Gilles Bastin et Paola Tubaro (2018) se révèle précieux en termes d'exemples sur l'exploitation des *big data* en Sciences Sociales. Nous noterons à cet égard la référence faite à O'Reilly *et al.* (2012) quant à la pertinence du *machine learning* appliqué aux données de recherches cliniques, les algorithmes permettent grâce aux données patients d'avoir une lecture plus fine quant à l'efficacité des médicaments via l'analyse des caractéristiques sociodémographiques ou de santé. Ceci n'étant pas sans incidence d'un point de vue éthique et conforte notamment l'importance d'une législation sur les

données personnelles (RGPD).

6. Changement de paradigme

Prenant appui sur Kant et la science classique, Bruno Bachimont argue que la « *loi scientifique n'appartient pas à la même construction scientifique que la donnée* » (Bachimont, 2015, p.391). En effet, bien que le traitement des mégadonnées soit statistique et devrait légitimement s'inscrire dans les approches hypothético-déductives, il n'existe pas toujours d'hypothèse préalable à ce type d'étude qu'il s'agirait de vérifier. Dans certains cas, les démarches scientifiques liées aux mégadonnées s'apparenteraient même davantage aux approches empirico-inductives, il existe une problématique scientifique, un questionnement, que l'on va essayer sinon de solutionner tout au moins d'expliquer par l'observation et l'analyse, soit à travers un cadrage théorique et méthodologique, les faits et les phénomènes. Ce n'est qu'après traitements métriques et statistiques de cette masse de données que l'on va pouvoir extraire une information que l'on espère exploitable et intelligible. Au regard de la démarche scientifique inhérente aux mégadonnées, celle-ci relèverait d'une hybridation méthodologique, s'appuyant sur l'empirie pour la collecte et sur le déductif pour le traitement et l'analyse. En d'autres termes, le paradigme de la mégadonnée initierait les approches « empirico-déductives », du moins pour les formes les moins avancées de traitements algorithmiques. En effet, elles ne peuvent à notre sens être véritablement inductives car même si l'on part bien de faits et données observables, l'on ne part pas du particulier pour induire une généralité. Au contraire on se sert de la masse des redondances pour en déduire une « explication ».

> *« L'enjeu et la révolution attendue des mégadonnées sont donc bien la compréhension nouvelle promise des activités humaines. Les mégadonnées pourraient constituer une révolution des sciences de la culture à l'instar de la révolution scientifique qui a permis, à l'orée de notre modernité, de transformer notre relation à la nature d'une description fondée sur le langage à un rapport de mesure expérimentale et de formalisation calculée. À l'instar de la nominalisation des sciences de la nature, les mégadonnées proposent la nominalisation des sciences de la culture » (Bachimont, 2015, p.395).*

Cette nominalisation des sciences de la culture est poussée à son paroxysme avec l'apprentissage machine ou le *machine learning,* lequel relève d'une autre forme d'approche scientifique. L'on parle de *machine learning* dès lors que l'algorithme repose sur les variables (*features*)[4] détectées par le *data scientist* pour influer sur la prédiction. Le *deep learning* quant à lui traite des données non-structurées, l'algorithme aura été entrainé, sur une base de données massives, à identifier par lui-même les variables par computation. À savoir que selon les experts-analystes des mégadonnées, environ 80 à 90% des données captées seraient non structurées. Ces variables s'apparentent à celles de la statistique classique mais s'opèrent sur des jeux de données hétérogènes lesquels

n'ont pas nécessairement de relation de dépendance. Les *features* du *machine learning* vont plus loin et tentent à travers ces mégadonnées de repérer des *patterns* d'observations. Car comme le souligne Tyler Reigeluth

> *« Il n'est plus seulement question, avec le machine learning, d'automatiser le geste du travail manuel, ou le traitement cognitif de l'information, mais d'automatiser le caractère prédictif de tout comportement. En tant que régime de production de prédictions, le machine learning nous intéresse en ce que son efficacité normative ne se réduit pas à l'efficacité d'une certaine forme discursive, mais s'insère dans une activité sociale » (Reigeluth, 2018, p.40).*

Cette notion de *pattern* nous renvoie aux modèles de la découverte (*patterns of discovery*) de Nordwood Russel Hanson (1958), c'est-à-dire des modèles structurants-organisateurs-configurateurs, d'organisations intelligibles révélées par la reliance[5] (Morin, 1986), utilisés comme le précise J.-L. Le Moigne (2008) pour décrire les démarches de la modélisation des processus complexes. Nous retrouvons également cette idée à travers la liction (Leleu-Merviel & Useille, 2008) et les schèmes organisateurs qu'Émilie Flon et Yves Jeanneret définissent comme suit :

> *« Il s'agit d'une construction structurelle qui assure un lien entre les dimensions technique, sémiotique et phénoménologique de la représentation, définissant ainsi certaines conditions déterminantes de la construction du sens des informations au sein d'un dispositif. Le schème ne se réduit ni à une modalité sémiotique (comme l'écrit), ni à une forme (comme la liste), ni à une procédure cognitive (comme le classement) mais intègre ces différents niveaux en une construction singulière » (Flon & Jeanneret, 2010, p.4)*

En l'occurrence le *machine learning* s'intéresse à ces schèmes organisateurs produits par les lictions régulières de données en situation et à un moment T, ces *patterns* sont donc issus d'un cadre qui les contextualise. Les algorithmes sont programmés pour identifier les *patterns* issus de nos habitus (Bourdieu, 1986) repérables à travers nos traces d'activités dans le cyberespace. Comme le souligne Dominique Cardon, « *la plupart du temps, les prédictions algorithmiques ne font que confirmer, en leur donnant une amplitude plus ou moins grande, des lois sociales bien connues* » (Cardon, 2015, p.65). Ce qui rappelle le phénomène de la « *textualisation des pratiques sociales* », lequel consiste au recensement quasi exhaustif du monde notamment à travers la collecte et l'analyse des traces et des données qui permettrait d'aboutir à « *une sorte de naturalisation de l'ensemble du social, de sorte qu'il existe désormais des calculs et des services pour tout* » (Souchier *et al.*, 2019, p.7).

Ces traitements algorithmiques permettent ainsi à partir de mégadonnées de faire des prédictions, parfois sans que ceux-là même qui les ont programmés ne comprennent comment les machines en sont arrivées à de tels « résultats », en particulier dans le cas du d*eep learning*. C'est ce qui est appelé la « boîte noire » du programme et se veut être le reflet de l'auto-apprentissage machine, d'une certaine Intelligence (Artificielle).

De fait, le *machine learning* et la quête de *pattern* ne coïncident pas véritablement avec les approches « empirico-déductives » énoncées plus avant, mais s'apparenterait à ce que N.R. Hanson appelle la « rétroduction »

> *« Décrivant, en historien des scientifiques (plutôt que des sciences) le comportement cognitif des scientifiques transformant leurs théories par des observations et leurs observations par de nouvelles théories, N. Hanson souligne qu'ils ne procèdent guère par déduction ou par induction comme le prétendent les traités d'épistémologie. Ils adoptent une stratégie cognitive plus tâtonnante, moins linéaire, plus récursive, plus téléologique (la quête de quelque forme d'intelligibilité), qu'il appellera plus volontiers la 'retroduction' que 'l'abduction' empruntant les deux termes à C.S. Peirce. Ceci en conceptualisant plus audacieusement me semble-t-il, le sens du mot retroduction que ne le faisait Pierce : Le sens aristotélicien du terme 'abduction' (qu'utilisait souvent Peirce, privilégiant 'le raisonnement vraisemblable', alors que le mot 'retroduction' chez Hanson privilégie l'interprétation par le raisonnement récursif ou auto référentiel ou chiasmatique : 'la théorie engendre le fait qui engendre la théorie, qui...'). » (Le Moigne, 2003b).*

Ainsi existe-t-il d'ores et déjà deux paradigmes de la mégadonnée correspondant à des programmations algorithmiques différentes. Le premier, s'appuyant sur des traitements statistiques et métriques classiques, pourrait constituer les approches « empirico-déductives ». Le second, inhérent à l'apprentissage machine et à l'identification de *patterns*, pourrait former les approches « empirico-retroductives ».

7. Conclusion : vers une hybridation ?

Les artefacts technoscientifiques, et en particulier le numérique, s'imposent aujourd'hui comme des catalyseurs de transformation dans nos pratiques scientifiques et sociales. En tant qu'artefact central dans l'écosystème technoscientifique, le numérique a profondément redéfini les modes de production, de validation et de diffusion des savoirs, tout en reconfigurant les paradigmes épistémologiques et méthodologiques. Cette révolution, loin d'être un simple ajustement technique, révèle une reconfiguration profonde de notre rapport au monde, où l'intelligence humaine et la puissance algorithmique cohabitent dans une tension créative, mais non sans risques. La double coupure caractérisant les systèmes numériques (Bachimont, 2017), nous rappelle que ces artefacts demeurent, pour l'instant, dépourvus de sens propre. Le calcul, bien qu'extraordinairement efficace pour traiter des données massives, repose encore sur des interprétations humaines pour générer des insights exploitables. Les techniques de data storytelling, en rendant ces données intelligibles, montrent que l'humain reste au cœur des processus décisionnels. Pourtant, la montée en puissance des technologies intelligentes, notamment le *deep learning*, soulève des questions cruciales sur l'autonomie algorithmique et la délégation des processus cognitifs. Sommes-nous en

train de céder à une logique où la machine, loin d'être un prolongement de l'intellect humain, impose sa propre rationalité ? Face à cette évolution, il est essentiel de rappeler que les technologies de l'intelligence ne doivent pas se substituer à l'intelligence humaine, mais plutôt la compléter et l'enrichir. L'approche computationnelle, bien que puissante, ne saurait remplacer la capacité humaine à interpréter les signaux faibles, à contextualiser, et à porter un regard critique sur les résultats obtenus. C'est ici que réside toute la pertinence des humanités numériques, qui cherchent à harmoniser les apports technologiques avec une réflexion éthique et épistémologique sur les savoirs qu'elles produisent.

Rappelons également la distinction entre technologies de l'intelligence et technologies intelligentes. Les premières, principalement basées sur l'IA symbolique, renforcent les capacités de raisonnement humain en fournissant des outils pour naviguer dans la complexité. Ces technologies s'appuient sur des modèles logiques explicites et des règles formelles, permettant à l'humain de comprendre, d'interagir et de s'approprier les processus sous-jacents. Elles peuvent être considérées comme des prolongements intellectuels, facilitant l'analyse et la prise de décision tout en respectant la transparence et l'interprétabilité. Tandis que les secondes fondées sur l'IA connexionniste, en se basant sur des modèles auto-apprenants, tendent à opacifier les processus cognitifs en raison de leur caractère souvent non interprétable. Cet équilibre fragile entre assistance et autonomie technologique doit guider nos choix pour préserver la primauté de l'intelligence humaine dans un monde de plus en plus structuré par les artefacts numériques. Cependant, pour tirer pleinement parti de ces artefacts, il nous apparaît nécessaire de conjuguer les forces des approches symboliques et connexionnistes de l'intelligence artificielle. L'IA symbolique, avec sa transparence et sa logique explicite, offre un cadre structurant et interprétable pour renforcer la capacité de raisonnement humain. À l'inverse, l'IA connexionniste, grâce à sa puissance d'apprentissage implicite et sa capacité à détecter des *patterns* dans de vastes ensembles de données, permet d'explorer des relations complexes souvent inaccessibles par des approches traditionnelles. En associant ces deux paradigmes, il devient possible de maintenir une interprétabilité et une cohérence conceptuelle tout en exploitant la puissance analytique des algorithmes d'apprentissage profond. Cette synergie entre IA symbolique et connexionniste contribuerait à concevoir des artefacts technologiques à la fois performants et compréhensibles, renforça
nt ainsi leur intégration éthique et méthodologique dans les pratiques scientifiques et sociales. En équilibrant ces deux approches, nous pouvons envisager un avenir où technologies de l'intelligence et technologies intelligentes s'entrelacent pour mieux répondre aux défis posés par la complexité croissante de nos environnements numériques et en particulier nos artefacts technoscientifiques.

Enfin, les implications sociales et culturelles de cette textualisation généralisée des pratiques scientifiques et sociales ne peuvent être ignorées. Comme le précises Dominique Cardon, les calculateurs, loin d'être neutres, produisent des cadres cognitifs et culturels qui orientent nos représentations du réel.

> *« Les calculateurs fabriquent notre réel, l'organisent et l'orientent. Ils produisent des conventions et des systèmes d'équivalence qui sélectionnent certains objets au détriment d'autres, imposent une hiérarchisation des valeurs qui en vient progressivement à dessiner les cadres cognitifs et culturels de nos sociétés » (Cardon, 2015, p.14).*

Dès lors, il est impératif de continuer à interroger ces artefacts à travers une perspective critique et complexe, en intégrant les dimensions éthiques, sociales et disciplinaires dans leur conception et leur usage.

Ainsi, la révolution numérique, bien qu'ambivalente, ouvre des perspectives fascinantes pour repenser notre rapport au savoir et à la société. Ce n'est pas seulement une question d'innovation technologique, mais une véritable redéfinition des relations entre science, technique et humanité. En affirmant une approche humaniste des artefacts technoscientifiques, nous pouvons espérer préserver et renforcer les valeurs fondamentales de la science : la quête de sens, la réflexion critique et la responsabilité sociale.

8. Références citées

Akrich, M. (2006). Les objets techniques et leurs utilisateurs de la conception à l'action. *Sociologie de la traduction. Textes fondateurs*, p.179-199.

Bachimont, B. (2004). *Arts et sciences du numérique : ingénierie des connaissances et critique de la raison computationnelle*. Mémoire de HDR. Université Technologique de Compiègne.

Bachimont, B. (2007). *Ingénierie des connaissances et des contenus, le numérique entre ontologies et documents.* Paris, Hermès Lavoisier.

Bachimont, B. (2015). Le numérique comme milieu : enjeux épistémologiques et phénoménologiques. Principes pour une science des données. *Interfaces numériques*. Vol.4 n°3, 385-402.

Bachimont, B. (2016). De l'éditorialisation à l'éditorialisation (2007-2017). *Séminaire Sens Public « Écriture numériques et éditorialisation »*, Montréal, 17 novembre 2016. [En ligne] https://seminaire.sens-public.org/spip.php?article62

Bachimont, B. (2017). Le numérique comme milieu: enjeux épistémologiques et phénoménologiques.: Principes pour une science des données. *Interfaces numériques*, *4*(3), 385-402.

Bachimont, B. (2019). Comment faire confiance aux données ? Questions épistémologiques et philosophiques. *Culture et IA. Intelligence Artificielle Graph Alpes*, La Clusaz, 21 au 24 mars 2019. [En ligne] https://youtu.be/JSEnI_FmFfM

Bachimont, B. (2020a). L'IA, le brin d'herbe, la caresse et le regard. *Interfaces numériques.* Vol.9, n°1. DOI : 10.25965/interfaces-numeriques.4135

Bachimont, B. (2020b). Rhétorique de l'intelligibilité, vers des humanités du numérique. *Séminaire NumeRev*, MSH-Sud, Montpellier, 28 février 2020. [En ligne]

https://youtu.be/Z9q5pijXp-E
Bastin, G., Tubaro, P. (2018). Le moment big data des sciences sociales. Revue française de sociologie, vol. 59, n°3. 375-394. [En ligne] https://www.cairn-int.info/revue-francaise-de-sociologie-2018-3-page-375.htm.
Blair, A. (2020). *Tant de choses à savoir*. Paris : Seuil.
Berners-Lee, T., Hendler, J., et Lassila, O. (2001). Le web sémantique. *Scientific American , 284* (5), 34-43.
Berti-Équille, L. (2019). Machine Learning – Based Data Cleaning : Current Solutions and Challenges. *Actes du colloque CIDE21 "Économie(s) du document : pratiques et prospectives liées au numérique"*, co-organisé par l'ESCT et GERIICO, Djerba, 3-4 avril 2019. Editions Europia
Bigot, J. & Mabi, C. (2017). Une instrumentation numérique des sciences humaines et sociales: Enjeux épistémologiques et communicationnels. *Les Cahiers du numérique*, vol. 13(3-4), 63-90.
Boullier, D. (2015). *Les trois humanités numériques et le positivisme algorithmique*, Toulouse, 26 Nov 2015, Colloque Franco-québecois. URL : https://www.canal-u.tv/video/fmsh/humanites_numeriques_dominique_boullier.20859
Bourdeloie, H. (2013). Ce que le numérique fait aux sciences humaines et sociales. *Tic&société*, vol. 7, n° 2. [En ligne] http://journals.openedition.org/ticetsociete/1500.
Bourdieu, P. (1986). Habitus, code et codification, *Actes de la recherche en sciences sociales*, vol. 64, n° 1, 40-44.
Callon, M. (1989). *La science et ses réseaux : Genèse et circulation des faits scientifiques*. Paris : Éditions La Découverte.
Carbou, G. (2017). Analyser les textes à l'ère des Humanités numériques: Quelques questions pour l'analyse statistique des données textuelles. *Les Cahiers du numérique*, vol. 13(3-4), 91-114.
Cardon, D. (2015). *À quoi rêvent les algorithmes. Nos vies à l'heure des* big data. Paris : Éditions Seuil.
Chartier, R., & Martin, H.-J. (Dir.). (1990). *Histoire de l'édition française. Tome 2 : Le livre triomphant, 1660-1830*. Paris : Fayard.
Clément, J. (2007). L'hypertexte, une technologie intellectuelle à l'ère de la complexité. *Humanités numériques,* Vol.1, 183-194.
Dacos, M. (2011). Manifeste des Digital humanities. *ThatCamp Paris 2010.* [En ligne] https://tcp.hypotheses.org/318.
Delengaigne, X., Deschamps, C., & Mongin, P. (2011). *Organisez vos données personnelles : l'essentiel du Personal Knowledge Management*. Editions Eyrolles.
Digitalhumanities (2008). A Digital Humanities Manifesto. [En ligne] http://manifesto.humanities.ucla.edu/2008/12/15/digital-humanities-manifesto/.
Dyens, O. (2008). *La condition inhumaine : essai sur l'effroi technologique*. Flammarion.
EPRIST. (2020). Le web sémantique. *Analyse I/IST* n°33 - novembre 2020. [En ligne] https://www.eprist.fr/le-web-semantique/
Flach, P. A. (2012). *Machine Learning: The Art and Science of Algorithms That Make Sense of Data*. Cambridge University Press.
Flon, E., & Jeanneret, Y. (2010). La notion de schème organisateur, outil d'analyse sémio-pragmatique des écrits d'écran. Revue des Interactions Humaines Médiatisées,

11(1). [En ligne] http://europia.org/RIHM/V11N1/RIHM11(1-1)-Flon-Jeanneret.pdf
Galleron, I., & Idmhand, F. (2020). De l'interopérabilité à la réutilisabilité des éditions électroniques », *Humanités numériques* 1 | 2020. [En ligne] http://journals.openedition.org/revuehn/350 ; DOI : https://doi.org/10.4000/revuehn.350
Gallot, S., & Verlaet, L. (2016). La transparence : l'utopie du numérique ?, Communication et organisation [En ligne] http://communicationorganisation.revues.org/5277.
Gardiès, C., & Fabre, I. (2010). Re-questionner les pratiques informationnelles. *Études de communication : langages, information, médiations, 35*, 121-132. https://doi.org/10.4000/edc.168
Geslot, J.-C., & Rebolledo-Dhuin, V. (2020). Du livre imprimé au Web sémantique : le projet du dictionnaire des éditeurs français du XIXe siècle. *Humanités numériques*, 2 | 2020 [En ligne]
Goffman, E. (1991). *Les cadres de l'expérience*, Paris : Les Éditions de Minuit.
Gonzalez Aguilar, A. (2019). Visualisation interactive de la jurisprudence de la Cour de cassation. *Les Cahiers de la Justice*, 2(2), 243-255. https://doi.org/10.3917/cdlj.1902.0243
Goody, J. (1979). *La raison graphique*. *La domestication de la pensée sauvage.* Paris: Les Éditions de Minuit.
Goody, J. (2007). L'oralité et l'écriture. *Communication et langages*, no 154, 3-10. [En ligne] http://doi.org/10.3406/colan.2007.4684
Goody, J., & Lejosne, J.-C. (2006). La technologie de l'intellect. *Pratiques*, *131*(1), 7-30.
Hanson, N. R. (1958). *Patterns of Discovery: An Inquiry into the Conceptual Foundations of Science*. Cambridge University Press.
Hottois, P. (1999). *Technoscience : science, technique et société*. Presses Universitaires de France.
Jeanneret, Y. (2011). *Y a-t-il (vraiment) des technologies de l'information ?*, Villeneuve d'Ascq, Presses universitaires du Septentrion.
Jeanneret, Y. & Souchier, E. (1999). Pour une poétique de l'écrit d'écran, *Xoana*, n° 6, p. 97- 107.
Juanals, B. (2003). *La culture de l'information : Du livre au numérique*. Paris : Hermès.
Kembellec, G. (2019). Semantic publishing, la sémantique dans la sémiotique des codes sources d'écrits d'écran scientifiques. *Les Enjeux de l'information et de la communication*, n°2/2019, 55-72.
Latour, B. (1987). *Science in action: How to follow scientists and engineers through society*. Cambridge, MA : Harvard University Press.
Legroux, J. (2008). *De l'information à la connaissance*. Paris : Éditions du CNRS.
Leleu-Merviel, S. (2002). De la navigation à la scénation. Un grand pas vers une dramaturgie du numérique. *Les Cahiers Du Numérique*, *3*, 97–120.
Leleu-Merviel, S. (2004). Effets de la numérisation et de la mise en réseau sur le concept de document. *Revue I3 - Information Interaction Intelligence*, Cépaduès, 4 (1). fsic_00001019
Leleu-Merviel, S., Useille, P. (2008). Quelques révisions du concept d'information. *Problématiques émergentes dans les sciences de l'information*, Hermès-Lavoisier, 25-56.

Le Moigne, J.-L. (2008). Edgar Morin, le génie de la Reliance. *Synergies Monde*, n°4, 177-184. [En ligne] https://www.gerflint.fr/Base/Monde4/lemoigne.pdf
Lévy, P. (1990). *Les technologies de l'intelligence. L'avenir de la pensée à l'ère informatique*. Paris : La découverte.
Mazza, R. (2009). *Introduction to information visualization*. Springer Science & Business Media.
Meyriat, J. (2006). Pour une compréhension plurisystématique du document (par intention). *Sciences de la société, 68*, 11-26.
Milette, M., Millerand, F., Myles D., Lattzko-Toth (2020). Méthodes de recherche en contexte numérique. Une orientation qualitative. Québec : PUM.
Morin, É. (1986). *La méthode. 3. La connaissance de la connaissance*. Paris : Éditions du Seuil.
Morin, E. (1994). « Sur l'interdisciplinarité » Bulletin Interactif du Centre International de Recherches et Études transdisciplinaires n° 2 - Juin 1994 [En ligne, consulté en septembre 2018] http://ciret-transdisciplinarity.org/bulletin/b2c2.php
O'Reilly, T, Steele, J., Loukides, M. & Hill, C. (2012). *How Data Science Is Transforming Health Care. Solving the Wanamaker Dilemna.* Sebastopol, O'Reilly Media.
Paquienséguy, F. (2019). De l'opendata à l'information, vers de nouvelles pratiques professionnelles? *Colloque international ISKO-France "Données et mégadonnées ouvertes en SHS : de nouveaux enjeux pour l'état et l'organisation des connaissances ?*", co-organisé par le Chapitre Français de l'ISKO et le LERASS-Céric, Université Paul-Valéry Montpellier 3, 10 octobre 2019.
Piaget, J. (1967). *Psychologie de l'intelligence*. Paris : Collection Agora, Edition Armand Colin.
Reigeluth, T. (2018). La prédiction algorithmique comme activité sociale. *Réseaux*, 211(5), 35-67. https://doi.org/10.3917/res.211.0035
Rieffel, R. (2014). *Révolution numérique, révolution culturelle ?* Gallimard, collection folio actuel.
Robert, P. (2000). Qu'est-ce qu'une technologie intellectuelle ? *Communication & Langages*, 123(1), 97–114.
Salvador X.-L. (2018). Transformations disciplinaires en Littérature et Sciences Humaines à l'heure numérique », *Quaderni* [En ligne], 98 | Hiver 2018-2019, mis en ligne le 05 février 2019, consulté le 06 octobre 2020. URL : http://journals.openedition.org/quaderni/1399 ; DOI : https://doi.org/10.4000/quaderni.1399.
Schütz, A. (1987). *Le chercheur et le quotidien*. Paris : Librairie des Méridiens Klincksieck.
Simondon, G. (1958 : 2024). *Du mode d'existence des objets techniques*. Flammarion.
Souchier, E., Candel, É., Jeanne-Perrier, V., & Gomez-Mejia, G. (2019). *Le numérique comme écriture : théories et méthodes d'analyse*. Armand Colin.
Verlaet, L. (2017). Introduction : Enjeux et apports des recherches en humanités numériques. *Revue Les Cahiers du numérique*, Édition Lavoisier, vol.13, n°3-4/2017, p.9-18.
Verlaet, L., & Chante, A. (2017). La notion d'autorité dans l'édition de la recherche : évolutions des formes, changements d'autorité ? *Revue Communication & Langage*,

n°192, 27-46.
Vygotski, L. (1934, éd. 1997). *Pensée et langage*, 3e éd. Paris : La Dispute.
Wenger, E. (1998). *Communities of practice: Learning, meaning, and identity*. New York : Cambridge University Press.

9. Notes

[1] Propos tenu lors de la journée d'étude sur l'intelligibilité du numérique organisée dans le cadre de la revue éponyme en juin 2021.
[2] Les smart data ou données intelligentes sont des données issues de sources préalablement identifiées et choisies, l'on parlera alors de données primaires (et non brutes) souvent liées à l'internet des objets, leur volume sera moins important ce qui permet un traitement en flux continu (en streaming).
[3] Notamment du fait de l'internet des objets. Source : https://www.lebigdata.fr/big-data-2025-idc#:~:text=A%20l'heure%20actuelle%2C%20en,soit%20175%20milliards%20de%20t%C3%A9raoctets . (consulté le 4 août 2020)
[4] Ces features déterminent le degré de dimensionnalité des données à traiter.
[5] « *Cette faculté de l'esprit qui est d'articuler ce qui est séparé et de relier ce qui est disjoint, de distinguer des formes ou des 'patterns' sans pour autant les découper, et d'identifier les 'tiers' ou les 'liants' qui les interfacent, n'est-ce pas aussi cela que nous entendons maintenant par la faculté de reliance ? Faculté de reliance qui appelle l'exercice de l'ingenium, lequel nous épargne les navigations cognitives qui vont du réductionnisme qui sépare, au holisme qui fusionne ; la reliance révèle et organise des 'patterns' d'interactions possibles par lesquelles les 'complexes' nous deviennent intelligibles assez pour que nous puissions agir intentionnellement en 'raison gardant' : 'En cela consiste notre dignité'.* » (Le Moigne, 2008, p.181).