N° 7 | 2025
Le 4ème paradigme en question

Le 4ème paradigme en question

Volet 1

Lise VERLAET *Professeur des universités*
Institut des Technosciences de l'Information et de la Communication
UR LHUMAIN
Université de Montpellier Paul-Valéry
Bruno BACHIMONT *Professeur des universités*
COSTECH (EA 2223)
University of Technology of Compiègne

Édition électronique :
URL :
https://intelligibilite-numerique.numerev.com/index.php/articles/revue-7/3741-le-4eme-paradigme-en-question
Date de publication : 04/05/2025

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

Pour **citer cette publication** : VERLAET, L., BACHIMONT, B. (2025) Le 4ème paradigme en question. *Intelligibilité du numérique*, (7). https://doi.org/10.34745/

Ce numéro s’inscrit dans une réflexion approfondie sur les conditions contemporaines de production et d’intelligibilité des données à l’ère du « quatrième paradigme » scientifique. À mesure que les environnements numériques se structurent autour de l’exploitation de données massives et de traitements automatisés, la question n’est plus seulement celle de l’accès à l’information, mais bien celle de sa compréhension, de sa mise en forme et de sa capacité à faire sens. Les contributions réunies proposent ainsi de déplacer le regard, en interrogeant la donnée non comme un donné, mais comme un construit, inscrit dans des dispositifs techniques, des cadres théoriques et des pratiques situées. Loin de toute illusion d’objectivité immédiate, elles mettent en évidence les processus par lesquels les données sont sélectionnées, transformées, instrumentées, représentées et interprétées, participant pleinement à la fabrique des savoirs. Ce faisant, ce numéro engage une réflexion sur les médiations multiples – techniques, sémiotiques, communicationnelles et épistémologiques – qui conditionnent l’intelligibilité du numérique. Qu’il s’agisse de visualisation, de narration ou encore de formes alternatives de restitution comme la sonification, les contributions montrent que la donnée ne devient intelligible qu’au prix d’un travail de configuration et de mise en relation. Elles invitent ainsi à penser les artefacts numériques comme des opérateurs de connaissance, au cœur d’une dynamique où se redéfinissent les rapports entre calcul, interprétation et compréhension.

The Fourth Paradigm in Question. Part 1.

This issue engages in an in-depth reflection on the contemporary conditions of data production and intelligibility in the era of the “fourth scientific paradigm.” As digital environments increasingly rely on the exploitation of massive datasets and automated processing, the challenge is no longer merely one of access to information, but rather of understanding, structuring, and making sense of it. The contributions gathered here seek to shift the perspective by considering data not as given, but as constructed—embedded within technical dispositifs, theoretical frameworks, and situated practices. Far from any illusion of immediate objectivity, they highlight the processes through which

data are selected, transformed, instrumented, represented, and interpreted, thus actively participating in the production of knowledge. In doing so, this issue foregrounds the multiple mediations—technical, semiotic, communicational, and epistemological—that shape the intelligibility of digital environments. Whether through visualization, narration, or alternative forms of representation such as sonification, the contributions demonstrate that data only become intelligible through processes of configuration and relational structuring. They ultimately invite us to conceive of digital artefacts as operators of knowledge, at the heart of a dynamic in which the relationships between computation, interpretation, and understanding are being redefined.

Mots-clés :
Humanités numériques, Épistémologie du numérique, Visualisation, Construction des connaissances, Data visualization, 4eme paradigme, Données massives, Intelligibilité des données, Big data, Data intelligibility, Fourth paradigm, Digital epistemology, Knowledge construction, Sociotechnical mediations, Narration and storytelling, Scientific instrumentation, Digital humanitie, Médiations sociotechniques, Narration et mise en récit, Instrumentation scientifique

---

Au cours de cette dernière décennie, les évolutions des technologies furent étonnamment rapides, transformant la plupart des activités humaines. En particulier, les pratiques scientifiques ne sont pas épargnées par la révolution numérique qualifiée de 4ème paradigme. Pour Tansley et Tolle (2009) le quatrième paradigme de la science est basé sur l'exploration des données massives comme processus de découverte scientifique. Il s'agit d'une nouvelle façon de faire de la science qui repose sur la capacité de collecter, de stocker et d'analyser de grandes quantités de données. Cet appel à articles de la revue interdisciplinaire Intelligibilité du numérique vise à interroger ce 4ème paradigme, qui repose des questions anciennes mais de manière radicale à ce qui constituent la connaissance scientifique et ses matériaux d'élaboration mais également à ce qui cimente les communications dans nos sociétés contemporaines**.** En effet, ce paradigme semble proposer d'adosser la science à un nominalisme radical et d'aborder la société à travers sa « datafication » totale.

Questions anciennes en effet. Il est évident que les évolutions techniques et technologiques dans le domaine scientifique n'ont pas commencé avec le numérique. 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. Car comme l'indique Jack Goody (2007), l'écrit est bien plus qu'un simple « mode d'enregistrement de la parole », il a non seulement permis à l'auteur d'user de la rhétorique adéquate pour formuler – et penser – son propos ; mais également aux lecteurs de l'assimiler ou de s'en accommoder (Piaget, 1967) en façonnant ainsi leur réflexion sur le monde qui les entoure. 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. Il est très clair que ces millénaires d'écriture et ces siècles d'imprimerie ont permis de diffuser et d'organiser les connaissances humaines, et forme un héritage fondamentale dans l'histoire de toutes sociétés avancées et en particulier des sociétés savantes. Toutefois, force est de constater que les évolutions techniques de la communication électronique et des écrits d'écran avec l'invention de technologies permettant le traitement de l'information programmable (Rieffel, 2014), soit les ordinateurs, l'informatique et le numérique, pourraient conduire sciences et sociétés vers une véritable rupture (Schwab, 2017). En effet le tsunami du 4ème paradigme serait le déclencheur de la 4ème révolution industrielle impactant tous les aspects de notre réalité. Pour Halevy (2016) ce n'est donc pas simplement une révolution, mais une renaissance, laquelle s'appuie sur 5 éléments communs pour vivre une bifurcation
: rupture écologique, rupture économique, rupture technologique, rupture organisationnelle et rupture philosophique.

1. Les données, un nouveau paradigme ?

Comme le suggère l'expression de 4e paradigme, l'idée sous-jacente est que les données et leur exploration constituent un nouveau cadre épistémique pour élaborer les connaissances scientifiques (Kuhn, 1983). Ce cadre remplacerait celui instauré par les sciences dont s'est doté la modernité depuis les révolutions galiléo-cartésiennes (Koyré, 1973 ; Kuhn, 1973). Ainsi, aux mesures élargissant le regard de l'observateur au-delà de ce que la nature montre, considérant les réponses qu'elle produit dans les expériences qu'on lui soumet, se succèdent les données qui formalisent et encodent notre représentation du réel. Or, il semble bien qu'une donnée ne soit pas une mesure dans son acception classique. La mesure est le produit d'une théorie scientifique de la nature qui spécifie comme l'interroger via des expérimentations : la mesure est le résultat d'une continuité épistémologique entre la question est la réponse. La donnée est la représentation formalisée d'un fait, c'est-à-dire une expression formatée selon un langage donnée, porteuse d'une valeur de vérité. Ce qui fait qu'une donnée est une donnée ne résulte donc pas nécessairement d'une théorisation particulière, mais peut simplement consister dans l'enregistrement d'un fait quelconque (Bachimont, 2016). Les données du 4e paradigme ne sont alors pas réduites aux données que l'on peut trouver dans les bases de données scientifiques, convertissant via leur formalisation les mesures en données, mais peuvent résulter de n'importe quelle représentation pourvu

qu'elle porte une affirmation. Les logs d'un serveur, les enregistrements d'une base de données administratives, les bases textuelles, etc., sont autant d'affirmations que l'on peut reprendre et explorer : affirmation de textes (suites de caractères), d'images (suites de pixels), d'enregistrements, etc.

Aussi assiste-t-on à un nominalisme radical où les données ne sont que l'affirmation d'elles-mêmes (Bachimont, 2014) : on explore les liens entre les données tels que leur formalisation et encodage permettent de les établir. Non pas qu'il y aurait un lien entre ce dont les données sont les données (ce qu'elles représentent), mais seulement entre leurs expressions. Comme l'illustre l'IA générative, la connaissance apprise à partir des données n'est pas une connaissance portant sur le monde mais une corrélation entre des expressions figurant ou non le monde, de manière vérace, fictive ou erronée.

Pourtant, l'exploration de données permet de découvrir des liens insoupçonnés entre les données et ce qu'elles représentent, d'optimiser les décisions qu'elles suggèrent, d'agencer des dispositifs divers, de la reconnaissance ou diagnostic, de la décision ou intervention, en passant par la création ou proposition de nouvelles alliances entre données. S'il y a donc bien des usages productifs liés aux données, la question se pose de savoir comment les caractériser, les situer et d'en déterminer la portée.

Si on s'accorde sur le principe d'un 4e paradigme en ce sens, on peut s'interroger sur la manière de construire des connaissances, de faire de la science avec des matériaux a-théoriques comme des données. Quelles questions faut-il se poser ? Quel agenda de recherche découle d'un tel paradigme ? Comment sont élaborées, discutées et validées les connaissances découlant des données ? En quoi sont-elles des connaissances et de quoi sont-elles la connaissance ? En quoi un entrepôt de données peut-il donner lieu à des explorations en lieu et place d'expériences ? Peut-on imaginer une complémentarité, une coopération entre le paradigme de la donnée et celui de la mesure ? Ou faut-il considérer, puisqu'un paradigme chasserait un autre, qu'on aurait des connaissances incommensurables entre elles ?

2. Des technologies de l'intelligence aux technologies intelligentes

Un deuxième axe de réflexion entend interroger les aspects historiques, socio-culturels et sociétaux, et philosophiques des évolutions technologiques afin de mieux saisir les incidences et enjeux inhérents au 4ème paradigme. En effet, d'un point de vue historique, il est admis que l'invention des systèmes de traitement de l'information programmable à la base de l'ordinateur tient ses origines dans la machine à calculer et donc que l'ordinateur n'est en ce sens qu'une évolution technologique de la Pascaline. L'objectif de cet instrumentation technique était de permettre à l'homme d'accomplir à grande échelle et plus rapidement, ce que nous maîtrisions déjà, voire de décupler le raisonnement humain et l'intelligence humaine. Rappelons que 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. Les technologies de l'intelligence peuvent ainsi être définies comme des solutions techniques et technologiques permettant à l'humain d'accroître ses propres facultés de raisonnement, afin de mieux comprendre la complexité de nos sociétés. Pour concevoir ces technologies de l'intelligence, il apparait nécessaire d'anticiper les comportements et les usages lors de la conception d'artefacts techniques, mais aussi de les inscrire dans les pratiques professionnelles des acteurs concernés, lesquelles évoluent corrélativement aux progressions de ces artefacts numérique (Simondon, 1969, 1989 ; Stiegler, 1998). Cette co-construction, co-constitution ou individuation selon les acceptions, sera particulièrement intéressante à questionner.

Mais revenons un instant sur le concept même de technologie de l'intelligence que nous retrouvons également dans la littérature scientifique sous la terminologie de « technologie intellectuelle ». Le concept qui a été popularisé par les recherches de Jack Goody (1979, 2007) lequel considère l'écriture comme une technologie de l'intellect, l'écrit donc comme une innovation des moyens de communication, une technologie fondée sur la « raison graphique », qui a pleinement participé à l'évolution culturelle de l'humain. L'on retrouve ce concept dans les travaux de Pierre Lévy (1990 ; 2011), avec ce que l'on peut qualifier de « raison hypertextuelle » qui est à l'honneur dans ce nouvel espace d'information et de communication qu'est le numérique et désigné par l'auteur de cyberespace. L'écriture hypertextuelle comme nouvelle métaphore de la communication et dont les principes, méthodes et outils sont particulièrement pertinents à l'ère de la complexité (Verlaet et al., 2013). Les technologies intellectuelles sont également présentes dans les travaux de Bruno Bachimont (2015 ; 2020), qui soutient notamment la théorie du support. Cette théorie part du postulat que tout objet technique est l'inscription matérielle d'une connaissance et que toute connaissance est d'origine technique, la connaissance étant, selon l'auteur, la capacité à l'agir téléologique (Bachimont, 2004). En ce sens, le support numérique introduit de nouvelles formes et de nouvelles dynamiques de représentation de l'information et des connaissances basées sur des manipulations calculatoires, ce que Bruno Bachimont (2007) qualifie de « raison computationnelle ». Cette raison computationnelle se constituent depuis les modèles de multiples techniques de manipulation des contenus (ingénierie des connaissances, les différentes IA, l'apprentissage machine, etc.) et procède également des différentes pratiques de la lettre et du nombre engendrées par le numérique (par exemple le copier-coller).

Point de vigilance, si l'en est, 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. Cependant, la frontière entre technologies de l'intelligence et technologies intelligentes est mince.

Les technologies de l'intelligence permettent de construire de nouveaux schèmes de pensée par et pour l'homme. Alors que l'intelligence artificielle se substitue à la raison humaine et crée une indépendance décisionnelle sur laquelle nous nous déchargerions

de toute activité mémorielle et de raisonnement. Toutefois, l'écart entre les deux concepts est subtil car permettre aux machines d'assister l'humain, consiste à leur octroyer les paramétrages nécessaires à cette fin. Ceci soulève la question de l'équilibre à trouver pour que ce soit les machines qui soient au service des hommes et non l'inverse.

3. Les promesses technoscientifiques : un enrôlement technologique ?

Il convient dès lors de s'interroger sur cet équilibre entre technologie de l'intelligence et technologies intelligente ? Comment configurer l'instrumentation scientifique ? Faut-il privilégier des systèmes experts, également désignée de IA « déterministe » que l'on peut rapprocher d'une instrumentation disciplinaire ? Ou bien les approches statistiques ou probabilistes, que l'on peut relier à une instrumentation transversale ? Dominique Boullier (2015) distingue l'instrumentation disciplinaire de l'instrumentation transversale, chacune d'elles ayant incontestablement avantages et inconvénients. Pour ce qui est de l'instrumentation scientifique disciplinaire, il précise que celle-ci n'est pas sans incidence sur les disciplines qui les utilisent. Certaines d'entre elles ont été affectées en profondeur dans leur dynamique scientifique interne. Dominique Boullier prend pour exemple le Traitement Automatique des Langues (TAL) pour les Sciences du Langage ou encore les Système d'Information Géographique (SIG) en Géographie et souligne que la puissance de calcul de cette instrumentation disciplinaire a occasionné un contournement théorique des disciplines en question, l'homme ne faisant pas le poids contre la force de calcul des machines. Cette instrumentation scientifique disciplinaire comporte donc des risques sur la nature et sur l'évolution intrinsèque de la discipline quand bien même ces outils sont construits pour les finalités de la discipline. Autre instrumentation citée par D. Boullier, les approches transversales portées par les politiques de recherche en France dont le Plan S et la Science ouverte, qui ont mis l'accent sur une dynamique technique mêlant informatique et cycle de la donnée et de l'information. S'il est entendu que ces initiatives techniques sont louables et nécessaires, l'auteur souligne qu'elles ont parfois été développées parfois hâtivement et au détriment d'une réflexion sur les concepts et paradigmes propres aux disciplines. Ces architextes
(Jeanneret& Souchier, 1999), détenteurs d'un monopole du sens, enferment ainsi les disciplines dans un mode de gouvernance méthodologique parfois très éloigné de celui des disciplines. Une partie des disciplines relevant des SHS se voient ainsi obligées par ces outils de faire cohabiter les méthodes quantitatives et qualitatives, alors que les approches positivistes ne faisaient jusqu'alors pas partie de leur pratique voire étaient écartées pour des raisons épistémologiques. Comment alors les disciplines peuvent-elles et doivent-elles s'approprier ces instrumentations et pour quels usages ? Quels avantages et inconvénients peut-on observer/expérimenter ? Les outils du 4ème paradigme sont-ils d'ores et déjà en train d'occulter les paradigmes traditionnels au sein des disciplines ?

Ceci pose aussi la question du développement et de la régulation de ces outils scientifiques, lesquels peuvent par surenchère de spécialisation supprimer des pans théoriques de recherche ou, par une volonté politique visant des consensus larges, de conduire à une rupture de la part de certaines disciplines qui ne se reconnaissent pas dans les fonctionnalités proposées voire considèrent ces outils comme une remise en question fondamentale de leur méthode. Pierre Mounier (2017) précise que si les promesses technoscientifiques laissent espérer des résultats inédits, ceci est largement grevé par la prédominance des méthodes et modèles en sciences techniques et médicales (STM) pour concevoir ces technologies. La difficulté que présentent ces instrumentations scientifiques est qu'elles obligent les chercheurs à s'asservir aux fonctionnalités d'outils qui ne répondent pas (totalement) à leurs paradigmes et aux méthodologies inhérentes, et donc à se soumettre au déterminisme technoscientifique qui les sous-tend. Mais sur quelles épistémologies les outils du 4ème paradigme scientifique s'appuie-t-il réellement ? Comment s'articule-t-il avec les paradigmes précédents ? Quelles approches méthodologiques gouvernent cette nouvelle façon de faire des « découvertes scientifiques » ? Quel avenir pour les paradigmes précédents ?

4. La question du sens du corpus : un donné ou une donnée ?

Enfin, reprenant le questionnement initial sur le statut des données, il ne faut pas oublier que, quel que soit le corpus donné à la machine, celle-ci - aussi « intelligente » soit-elle - n'est pas en mesure d'en apprécier la valeur et la pertinence, et fournira toujours un résultat froid - dit objectif - mais dont la véracité ou la pertinence devront être établies par ailleurs dans une reprise scientifique, épistémologique et finalement politique des résultats produits.

Or, la question de la donnée se pose dans le contexte désormais prégnant d'Internet et des modes de communication associés : une production non régulée indéfinie pour ne pas dire infinie d'inscriptions numériques en circulation dont tout un chacun peut se saisir comme données. En fonction des motivations présidant à ces saisies ou reprises, on peut assister tant à la construction de connaissances qu'à des manœuvres de désinformation et des *fakenews*. Ceci nous conduit *de facto* aux problématiques de corpus. En effet, une collecte d'information ou de données sur Internet ne constitue pas d'emblée un corpus qui répond à une question de recherche et qui motive et justifie la sélection des informations qu'il rassemble. Mais en face de données collectées par les outils ressortissant du 4e paradigme, sur quelles bases algorithmiques ces motivation et justification peuvent-elles être opérées ?

Une des difficultés souvent mise en avant est que les algorithmes utilisés ont un fonctionnement qui s'apparente généralement à une boîte noire… Or, comme le rappelle Sylvie Leleu-Merviel (2003), il y a toujours un maître du sens à tout dispositif

technoscientifique. Outre les difficultés épistémologiques déjà évoquées, ces éléments – parmi tant d'autres – viennent en outre renforcer les dystopies liées à ces technologies, de la crainte d'une surveillance absolue par la collecte des données et l'étude des traces, (cf. le « panopticon » de Foucault ou les réserves de Wiener sur les usages malintentionnés des machines).

Mais au-delà des questions de principes, il demeure le problème pratique et technique de traiter des corpus et collectes de données massives, conduisant à ces volumétries inédites. Dans ce contexte, les outils statistiques et métriques peuvent apparaître comme des méthodes d'analyse indiquées pour traiter (semi)-automatiquement ces grandes masses de données. Il est à cet égard intéressant de constater que le phénomène *big data* a parfois laissé la place à celui de *smart data*, car une fois le défi technologique éprouvé, c'est bel et bien la notion de corpus ou de jeu de données qui revient pour obtenir des résultats pertinents. Connaître et comprendre le corpus à sélectionner pour s'assurer de la fiabilité des (res)sources exploitées. Dans cette optique, on peut s'intéresser aux exploitations concrètes des entrepôts de données dont l'ambition outre le stockage consiste en leur réutilisation pour capitaliser sur l'existant.
La donnée est souvent présentée comme le nouvel or noir de la recherche scientifique, ces entrepôts de données devenant la clef de la connaissance. Mais, faisant écho au premier axe, la notion même de « donnée » est problématique : données brutes, données produites, données construites (Prost & Schöpfel, 2019). 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). Si l'on reprend la pyramide DIKW utilisée en gestion des connaissances, où s'arrête l'utilisation de cette instrumentation scientifique ? Quand et comment intervient l'intellection humaine dans cette nouvelle configuration ?

5. Sommaire du numéro

David Lamboley (2025) propose une exploration originale des modalités d'intelligibilité des données en mobilisant le registre sonore, encore largement marginal dans les pratiques scientifiques. À partir d'un cadre théorique articulant épistémologie du numérique, phénoménologie de la perception et sciences cognitives, l'auteur montre que la sonification – et plus particulièrement la musification – constitue une voie alternative pour appréhender la complexité des données massives. En s'appuyant sur les capacités perceptives de l'audition, cette approche permet de révéler des structures, des dynamiques et des invariants souvent imperceptibles dans les représentations visuelles. Loin d'une simple esthétisation, la musification est envisagée comme un dispositif de médiation épistémique, fondé sur une mise en correspondance rigoureuse entre structures de données et formes sonores. L'article met ainsi en évidence les conditions de validité scientifique de ces dispositifs et interroge leur capacité à produire des connaissances en articulant expérience sensible et

formalisation. En proposant un déplacement vers un réalisme perceptif puis structural, il ouvre la voie à une réconciliation entre sciences de la nature et sciences humaines, où le sonore devient un véritable instrument de connaissance.

Dans l'article "*Construire des connaissances en sciences du langage : données langagières, corpus numériques et instrumenta(lisa)tions*", Jérémi Sauvage (2025) propose une réflexion épistémologique sur les conditions de production des connaissances en sciences du langage à l'ère du « quatrième paradigme » fondé sur l'exploitation massive des données. À rebours d'une vision techniciste qui associerait mécaniquement volume de données et objectivité scientifique, l'auteur met en évidence les médiations multiples qui affectent la donnée langagière : processus de constitution des corpus, choix d'instrumentation, opérations d'étiquetage et de visualisation. En mobilisant notamment la pensée complexe d'Edgar Morin, il montre que la donnée ne constitue jamais un donné brut mais une construction située, traversée par des cadres théoriques, des choix méthodologiques et des biais interprétatifs. L'article alerte ainsi sur les risques d'instrumentalisation de la « data », y compris dans des contextes scientifiques, et appelle à une vigilance critique renouvelée quant au statut épistémologique des données. Il réaffirme, en creux, la nécessité de maintenir une articulation exigeante entre raisonnement théorique, construction des corpus et interprétation, dans une perspective réflexive et non réductrice des pratiques scientifiques.

Christophe Tuffery (2025) aborde le domaine de l'archéologie en rappelant que cette discipline est faite d'inscriptions, tant à travers les traces qu'elle reconnaît dans les chantiers qu'elle entreprend que dans celles qu'elle constitue à partir de ces traces. Science des vestiges, certes, mais technique d'inscription assurément. À ce titre, les sciences physiques, le numérique et l'IA transforment globalement le mode de production de ces inscriptions et de leur traitement : enregistrements physiques, codification numérique et traitement algorithmique, et dernièrement reconfiguration textuelle des contenus. L'auteur montre cependant qu'une rupture s'effectue avec l'IA générative dans la mesure où le travail s'effectue selon deux modalités inédites - la quasi-instantanéité du traitement, l'opacité de son fonctionnement - pour un résultat surprenant par sa qualité. En effet, le travail a porté sur la transcription numérique d'archives de fouille, notamment d'un carnet de fouille datant des années 1970 d'environ 160 pages, renseigné quotidiennement pendant plusieurs campagnes par plusieurs des fouilleurs présents sur le chantier. Deux outils ont été mobilisés pour des résultats insatisfaisants, les difficultés restant nombreuses (écritures manuscrites de scripteurs différents, schémas divers, insuffisance des données pour l'apprentissage). Mais le résultat fut incomparablement meilleur en utilisant des outils d'IA générative (Gemini de Google d'une part, et Le Chat de Mistral d'autre part). Le résultat proposait non seulement des transcriptions de qualité, mais une structuration et instrumentation des informations permettant une consultation optimisée. Cela montre à la fois l'importance, mais aussi le décalage, du passage de l'interprétation au calcul : la transformation textuelle interroge sur le type de culture lettrée d'une part et d'expertise scientifique à élaborer dans ce nouvel environnement d'inscription. Si la science a

toujours été une affaire d’écriture et d
’inscription, l’IA générative nous installe dans un nouveau rapport à la textualité, pas seulement littéraire, mais aussi scientifique comme le montre fort bien cet article.

Cassandre Burnier (2025) dans son article intitulé "*Comment fait-on parler les données par elles-mêmes ? Une approche communicationnelle de la production de connaissance*" propose une analyse des conditions de production du sens à partir des données, en déconstruisant le mythe d’une donnée « auto-explicative » au cœur du quatrième paradigme. À partir d’une enquête ethnographique menée dans un groupe de presse, l’autrice montre que les données ne parlent jamais d’elles-mêmes mais font l’objet d’un travail collectif d’interprétation, de qualification et de mise en récit. En mobilisant notamment la théorie de la ventriloquie, l’approche constitutive de la communication et la notion d’agencement narratif, elle met en évidence une agentivité distribuée entre humains et artefacts techniques, où la donnée devient un actant participant à la production organisationnelle du sens. L’étude du « Profil SSO » révèle ainsi que l’intelligibilité des données repose sur des processus situés de négociation entre logiques techniques et impératifs métiers, donnant lieu à la construction de véritables « biographies numériques » des usagers. Loin d’une objectivité immédiate, la donnée apparaît comme une production sociotechnique, dont la valeur cognitive dépend des dispositifs narratifs et organisationnels qui la rendent interprétable et actionnable.

La question de la transformation de la production scientifique, abordée dans les articles précédents, est posée directement par l’article de Cléo Collomb (2025) qui revient sur le programme AlphaFold qui valut le prix Nobel de chimie à ses concepteurs, alors qu’ils ne sont pas chimistes mais informaticiens. Ce programme permet en effet d’aborder et de traiter un problème majeur de la biologie structurale, à savoir la prédiction de la structure d’une protéine à partir de la séquence de ses acides aminés. Cléo Collomb montre que la rupture n’est pas tant le fait d’avoir obtenu un résultat de manière inédite, la technique d’écriture prenant le pas sur la consultation de la nature, mais dans la crise qu’entraînait l’usage de l’IA dans le fonctionnement scientifique et en particulier ce qui permet à un résultat scientifique d’être considéré comme tel. Cette crise se déploie à trois niveaux : la reproductibilité computationnelle, où un résultat proposé doit pouvoir être retrouvé, reproduit par la communauté dès lors qu’il est publié ; la reproductibilité économique qui renvoie aux moyens nécessaires pour produire le résultat, et enfin la reproductibilité technique qui consiste dans le savoir-faire inhérent à la mise en œuvre des outils, particulièrement sensible dans l’IA où l’effet des paramètres choisis est plus constaté qu’expliqué. L’IA générative ne permet pas en effet de reproduire à l’identique des résultats, demande des moyens techniques et des masses données inaccessibles pour les membres des communautés scientifiques, et finalement repose sur des tours de main sans lien avec la connaissance produite. Certes, certains résultats scientifiques (on peut penser à la physique des particules ou la cosmogonie) posent des problèmes semblables : coûts astronomiques (!), complexité des calculs, déluge de résultats. Mais ces expériences sont la production d’une communauté scientifique rassemblée dans des organisations publiques internationales, où ces reproductibilités sont immanentes à leur fonctionnement. Alors que les

programmes comme AlphaFold sont le fait de compagnies privées, dialoguant avec les communautés scientifiques portant ces valeurs, sans en faire partie. L'article conclut sur le dernier déplacement, sans doute le plus important, notamment pour une revue comme la nôtre, à savoir l'abandon de l'explication et de l'intelligibilité au profit de l'efficacité (et efficacité du profit). Ce déplacement induit que la science ne cherche plus à comprendre, mais se contente de faire pour optimiser : la production de résultat, des profits, l'accumulation de faits, etc. On est loin de la science telle qu'on a pensé la mener jusqu'à aujourd'hui, des Grecs jusqu'à Einstein : comprendre restait l'objectif. Comme le dit Cléo Collomb, ce quatrième paradigme pourrait bien signifier la disparition des trois premiers.

L'article de Simon Dumas-Primbault et Paul Guhennec (2025) s'intéresse également à la rupture apportée par l'IA et le quatrième paradigme en l'abordant depuis la notion de données et de la métaphore de la liquidité mobilisée pour décrire leur nature et leur évolution. En effet, l'expression « lac de données » est régulièrement utilisée dans le monde du Web des données (ou Web sémantique) pour évoquer comment les multiples sources de faits et « connaissances » alimentent un unique répertoire, synthèse universelle, qu'il suffirait d'interroger et explorer pour avoir à sa disposition tout le savoir du monde. Autre métaphore de liquidité, la notion de déluge, insistant sur le flux massif que le calcul et ce paradigme rendrait enfin maîtrisable. On pourrait également évoquer dans le même esprit la métaphore des données comme nouvel or noir (fuel ou pétrole) de la connaissance. L'article explique à travers la notion d'espace latent et les modes de sa construction comment la liquidité invoquée se traduit par la continuité des métriques utilisées dans cet espace de représentation des données. Le propre de cet espace est alors de permettre la navigation entre les données, de reformuler par des mots des notions voisines que l'on trouve via leur représentation dans cet espace. Mais les auteurs montrent bien que, même si cet espace remplace par le calcul entre positions géométrique l'herméneutique interprétative permettant la compréhension des termes, il n'en reste pas moins soumis à des hypothèses, des situations, des compromis techniques qui relèvent d'une herméneutique technique et numérique, clarifiant les préjugés et les présupposés pour les intégrer dans la compréhension à mener. C'est ainsi que, contrairement aux ambitions affichées de remplacer la théorie par l'objectivité des calculs effectués sur le lac ou l'océan des données, il y aurait plus que jamais le besoin de théorie pour la critique des données (au sens de la critique des textes, l'établissement d'une version accordée d'un texte – critique philologique, d'une donnée – critique numérique) et leur interprétation. Les auteurs concluent sur une métaphore non plus liquide mais géographique, bien qu'en lien avec l'univers marin, à savoir l'archipel du savoir, des communautés épistémiques se réappropriant les données comme connaissances par leur connaissance sur les données.

6. Références citées

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. (2014). Le nominalisme et la culture : questions posées par les enjeux du numérique. In B. Stiegler (Ed.), *Digital Studies, organologie des savoirs et technologies de la connaissance* (pp. 63-78). Paris: FYP Editions.
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). Traces, calcul et interprétation : de la mesure à la donnée. *Azimuth : Philosophical Coordinates in Modern and Contemporary Ages, IV*(7), 13-36.
Bachimont, B. (2020). L'IA, le brin d'herbe, la caresse et le regard. *Interfaces numériques.* Vol.9, n°1. DOI : 10.25965/interfaces-numeriques.4135
Boullier, D., (2015). Les trois humanités numériques et le positivisme algorithmique, Colloque *Humanités Numériques*, Toulouse, 26 Novembre 2015, [En ligne] https://www.canal-u.tv/video/fmsh/humanites_numeriques_dominique_boullier.20859
Burnier, C. (2025). Comment fait-on parler les données par elles-mêmes ? Une approche communicationnelle de la production de connaissance. *Revue Intelligibilité du numérique*, 7|2025. [En ligne]
Collomb, C. (2025). Le quatrième paradigme ou la remise en cause des normes de la science. Une enquête socio-économique et épistémologique sur AlphaFold. *Revue Intelligibilité du numérique*, 7|2025. [En ligne]
Dumas Primbault, S. & Guhennec, P. (2025). La fabrique de la liquidité des données. Plongements dans un milieu navigable et herméneutique(s) de l'espace latent. *Revue Intelligibilité du numérique*, 7|2025. [En ligne]
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
Halévy, M. (2016). Qu'est-ce qui nous arrive ? Laurence Massaro.
Jeanneret, Y. & Souchier, E. (1999). « Pour une poétique de l'écrit d'écran », *Xoana*, 6 : 97-107.
Leleu-Merviel, S. (2003). Les désarrois des « Maître du sens » à l'ère numérique, *H2PTM'03 Conférence international, Hypertextes et Hypermédias, Créer du sens à l'ère numérique*, Paris : Edition Hermès.
Koyré, A. (1973). *Du monde clos à l'univers infini* (Traduction R. Tarr). Paris: Gallimard (Idées).
Kuhn, T. (1983). *La structure des révolutions scientifiques*: Flammarion.
Kuhn, T. S. (1973) *La révolution copernicienne*. Paris: Arthème Fayard.
Lamboley, D. (2025). Les apports du son dans la restitution du sens : de la sonification à la musification des données. *Revue Intelligibilité du numérique*, 7|2025. [En ligne]
Lévy, P. (1990). *Les technologies de l'intelligence. L'avenir de la pensée à l'ère informatique.* Editions La Découverte.
Lévy, P. (2011). *La sphère sémantique 1 : computation, cognition, économie de*

l'information. Lavoisier.
Mazza, R. (2009). *Introduction to information visualization*. Springer Science & Business Media.
Mounier, P. (2017). Les Humanités numériques, gadget ou progrès : Enquête sur une guerre souterraine au sein de la recherche. *Revue du Crieur*, 7(2), 144-159. [En ligne] doi:10.3917/crieu.007.0144.
Piaget, J. (1967). *Psychologie de l'intelligence*. Paris : Collection Agora, Edition Armand Colin.
Prost, H., & Schöpfel, J. (2019). Les entrepôts de données en sciences de l'information et de la communication (SIC). Une étude empirique. *Études de communication*, *52*, 71-98.
Rieffel, R. (2014). *Révolution numérique, révolution culturelle ?* Gallimard, collection folio actuel.
Sauvage, J. (2025). Construire des connaissances en sciences du langage : données langagières, corpus numériques et instrumenta(lisa)tions. *Revue Intelligibilité du numérique*, 7|2025. [En ligne]
Schwab, K. (2017). *The fourth industrial revolution*. Crown Currency.
Simondon, G. (1969). *Du mode d'existence des objets techniques*. Paris : Aubier Montaigne.
Simondon, G. (1989). *L'individuation psychique et collective*. Paris : Aubier Montaigne.
Stiegler, B. (1998). Temps et individuations technique, psychique et collective dans l'oeuvre de Simondon. *Intellectica*, *26*(1), 241-256.
Tansley, S., & Tolle, K. M. (2009). *The fourth paradigm: data-intensive scientific discovery* (Vol. 1). T. Hey (Ed.). Redmond, WA: Microsoft research.
Tufféry, C. (2025). Les pratiques numériques en archéologie annoncent-elles un changement de ses paradigmes ? *Revue Intelligibilité du numérique*, 7|2025. [En ligne]
Verlaet, L., Gallot, S., & Gonzales Aguilar, A., (2013). Le paradigme de la complexité. Apports pour les usages formels de l'hypertexte In Jeanneret, Y et al. *Pratiques et usages numériques : H2PTM'13*, Hermès-Lavoisier, 75-90.