

# Cognition artificielle et production de savoirs : enjeux épistémologiques d'un paradigme computationnel dans les humanités

Denise Scandarolli<sup>1, 2</sup> 

<sup>1</sup> Université du Québec à Montréal (UQAM), Canada

<sup>2</sup> Chaire de recherche du Canada en jeu, technologies et société

## Abstract

This paper explores the increasing integration of artificial intelligence in social sciences and humanities research, framing these devices not as neutral tools but as socially situated semiotic and epistemological artifacts. By reflecting on the emergence of a computational paradigm in the humanities, this article draws upon debates within cognitive computing to discuss the pragmatic and ethical implications of using these systems in knowledge production. The analysis thus aims to contribute to a critical reflection on the challenges facing research in light of these technologies.

**Mots-clés:** informatique cognitive, intelligence artificielle, sciences humaines et sociales, éthique, biais algorithmique.

**Keywords:** cognitive computing, artificial intelligence, human and social sciences, ethics, algorithmic bias.

## 1 Introduction

Dans l'édition inaugurale de la revue *Computers and the Humanities* [18], Louis Milic affirme la consolidation de l'utilisation de méthodes computationnelles dans les études en humanités. Intitulé *The next step*, son texte revient sur le franchissement d'une première étape « when to admit that you were working on a literary problem with the help of a computer was equivalent to saying that you were an eccentric (...) and possibly an underminer of the liberal tradition » [18]. Il confirme ainsi l'émergence d'un nouveau paradigme [16], qui assume les technologies computationnelles comme moyen de traiter l'information provenant d'événements et de phénomènes de l'humain en tant qu'individu social : un paradigme computationnel des sciences humaines et sociales (SHS).

L'enthousiasme de Milic [18] pour la « realization of the advantages of the new technology », ne reflète pas uniquement les progrès techniques de l'époque, il s'inscrit principalement dans l'intégration, au sein des SHS, de discours qui normalisaient la capacité des outils informatiques à compenser les limites humaines. Ces discours accompagnent l'émergence de nouvelles logiques issues des relations politiques et scientifiques établies au cours du XXe siècle, marquées par la rationalisation autoritaire, le contrôle des subjectivités et l'instrumentalisation politique de l'universel [1 ; 14], où la cybernétique s'est imposée comme moyen de régulation, d'adaptation et de pilotage des systèmes vivants, cognitifs et sociaux [4 ; 9 ; 25].

Des événements tels que la conférence de Dartmouth (Dartmouth Workshop, 1956) et les conférences Macy (1942-1953) ont contribué à légitimer l'idée selon laquelle ces technologies seraient capables d'agir dans la société. En revendiquant une prétendue neutralité, ces dispositifs ont opéré

---

Denise Scandarolli. "Cognition artificielle et production de savoirs : enjeux épistémologiques d'un paradigme computationnel dans les humanités." *Actes de la Conférence Humanistica*, éd. par Serena Crespi, Simon Gabay, Martin Grandjean, Ariane Pinche, Marie Puren et Léa Saint-Raymond. Vol. 4. Anthology of Computers and the Humanities. 2026, 79–84. <https://doi.org/10.63744/mUwpuEm3qlj9>.

un glissement sémantique, consacrant les technologies comme des artefacts impartiaux, aptes à produire des connaissances objectives. C'est dans cette perspective que les SHS pourraient en tirer profit, comme le souligne Milic [18], « computer-aided study in the humanities is beginning to provide us with all the good things we have been lacking for so long ».

Cette inflexion se prolonge dans le contexte contemporain et trouve un écho singulier dans la recherche scientifique, notamment dans le contexte des avancées de l'intelligence artificielle (IA). Les algorithmes, structures codées qui composent ces systèmes informatiques, demeurent pourtant peu problématisés en tant que produits d'un savoir articulé et institutionnalisé, ou comme vecteurs de signification. Leur statut d'artefact tend à les réduire à de simples instruments de calcul inférentiel, perçus comme dépourvus de toute subjectivité.

Les questions relatives à la nature sémiotique des algorithmes et aux processus de production de sens par les systèmes informatiques restent encore marginales. Elles s'avèrent pourtant d'autant plus cruciales au regard des développements récents dans le domaine de l'apprentissage automatique, de l'apprentissage profond, dont les grands modèles de langue (LLM), qui ont ouvert des horizons inédits à la recherche en SHS. Combinées à la numérisation croissante des pratiques sociales [4; 15], source de données massives (*Big Data*), ces technologies offrent des perspectives d'analyse renouvelées, mais non sans implications symboliques majeures.

L'usage de l'IA comme outil pour l'analyse de données de recherche suscite d'importants questionnements quant aux processus de formalisation et à la valeur des savoirs produits. Cette contribution explore ainsi les discussions théoriques de l'informatique cognitive afin de problématiser la double nature des algorithmes : celle de producteurs de sens dans le processus de traitement de l'information et celle de produits eux-mêmes de réflexions et de débats épistémologiques préalables. L'enjeu consiste à examiner, dans cette perspective, leurs fondements et leurs implications pour contribuer au débat sur les répercussions éthiques et politiques du déploiement de ces technologies dans la recherche en SHS.

## 2 Les algorithmes comme conception scientifique

Depuis le test de Turing (1950), conçu pour offrir une définition opérationnelle de l'intelligence par l'imitation, le fonctionnement des algorithmes a considérablement évolué, tant en termes de performance que de complexité des tâches qu'ils accomplissent. Dans la pratique scientifique contemporaine, ces algorithmes interviennent de deux manières : indirectement, dans un usage « instrumental » (par exemple dans la recherche bibliographique via les moteurs de recherche); et directement, lorsqu'ils sont intégrés à la méthodologie de recherche comme instruments d'analyse. Dans ce dernier cas, une valeur factuelle est souvent accordée aux résultats produits par la machine.

L'une des définitions les plus largement diffusées considère l'algorithme comme « any well-defined computational procedure that takes some value, or set of values, as input and produces some value, or set of values, as output » [6]. En tant que procédure computationnelle, il est souvent perçu comme physiquement inscrit dans le matériel informatique ; cette réification tend à le naturaliser en simple objet technique, occultant sa dimension symbolique. Toutefois, dans les débats en informatique cognitive, l'algorithme assume une fonction représentationnelle. Il devient une forme de modélisation computationnelle de la cognition, visant à comprendre, par la reproduction, les mécanismes fondamentaux du fonctionnement de l'esprit (*mind*) [11; 18].

Le processus rhétorique et épistémologique de construction de la signification conceptuelle des algorithmes n'est pas anodin. En tant que séquences de procédures codées, leur structure n'est accessible qu'à un cercle restreint d'experts. Pour la plupart des individus, y compris de nombreux spécialistes, notamment dans le cas des modèles d'apprentissage profond (*deep learning*) [2; 8; 13], ces systèmes restent structurellement opaques. Cette opacité institue une asymétrie analytique qui entrave toute capacité effective de remise en question, de vérification ou de jugement critique sur la validité des résultats qu'ils produisent [3; 4]. C'est pourquoi il est essentiel de comprendre

les logiques conceptuelles qui président à la construction de ces systèmes, des « lois » elles-mêmes par nature négociables, et qui affectent les résultats informatiques générés.

L'intelligence artificielle est décrite comme un domaine qui « attempts not just to understand but also to build intelligent entities » [21] capables d'opérer dans leur environnement. Ces systèmes sont conçus à partir d'une distinction conceptuelle entre deux paradigmes principaux : le courant symbolique (systèmes experts, ontologies et systèmes à base de connaissances) et le courant connexionniste naturaliste (réseaux de neurones artificiels, algorithmes génétiques et LLMs).

En tant que systèmes de traitement de données, les résultats (*outputs*) générés par ces algorithmes sont signifiés par des processus de raisonnement forgés à partir de conceptions distinctes. Les courants symbolique et connexionniste conçoivent différemment la nature de l'intelligence et de la représentation, tout comme celles de la mémoire et de la catégorisation. Nous nous limiterons ici aux deux premières notions.

Pour simplifier une discussion vaste et beaucoup plus complexe, la distinction entre symbolistes et connexionnistes quant à la nature de l'intelligence, issue de débats interdisciplinaires (sciences cognitives, linguistiques, psychologie), oppose deux visions principales. Le courant symbolique soutient que l'intelligence artificielle peut reproduire l'intelligence « naturelle », postulant que toutes deux sont structurées par des mécanismes formels de traitement de l'information, constitués par des règles explicites [11 ; 19 ; 23]. Pour les connexionnistes naturalistes, l'intelligence est un phénomène émergent de l'activité neuronale et d'un processus évolutif biologique et qui se manifeste à travers les cognitions incarnées, situées et sociales [10 ; 20 ; 24]. Ces deux courants opposent ainsi une intelligence « désincarnée », fondée sur des règles logiques, à une intelligence « biologiquement située ».

La divergence entre ces deux paradigmes réside aussi dans leur conception respective des états mentaux représentationnels. Pour les symbolistes, la cognition se compose de systèmes de manipulation de symboles régis par l'application de règles formelles. S'inscrivant dans une démarche essentiellement déductive, ils postulent l'existence d'un niveau de représentation structuré par un « langage de la pensée » [11]. Dans ce contexte, les représentations mentales sont envisagées comme des constituants atomiques dont la structure syntaxique prime sur le contenu sémantique. Ainsi, selon le principe de compositionnalité, le sens d'une représentation complexe est déterminé par la syntaxique des parties qui la composent.

Par ailleurs, pour les connexionnistes naturalistes, qui adoptent une perception holistique de la cognition incluant le corps vivant en interaction avec son environnement, les symboles ne sont pas constitués par les données d'entrée (*inputs*), mais par des propriétés qui émergent de la perception. À travers un raisonnement inductif, cette perception s'affine et se corrige dans le processus itératif de l'apprentissage automatisé. La nature de la représentation ne détermine pas le contenu sémantique par sa relation à des fonctions syntaxiques dans une structure combinatoire. L'attribution du sens se déplace vers les « nœuds » de l'architecture connexionniste, envisagés comme des représentations distribuées dans le réseau. Toutefois, ce modèle peine à clarifier les relations causales entre ces unités, suscitant des critiques [11 ; 21] quant à l'opacité de ses processus décisionnels.

Pour les symbolistes, la représentation revêt un sens explicite, dans la mesure où elle constitue un signe discret et manipulable. Pour les connexionnistes, en revanche, la représentation est distribuée (sous forme de vecteur numérique), le sens y demeure implicite, n'émergeant que de procédures statistiques. Ce passage du signe discret à l'agrégat statistique redéfinit radicalement les conditions de lisibilité et d'interprétation des connaissances générées par le dispositif informatique.

### **3 L'IA comme méthode de recherche : enjeux éthiques**

Lors de la conférence *Computers for the Humanities?* (1965), Jacques Barzun reconnaît l'utilité du traitement informatique dans les sciences humaines pour « indexing, collating, verifying, and

similar drudgery » [17]. Il formule cependant une mise en garde fondamentale : les humanistes utilisant l'ordinateur à d'autres fins s'exposent à un réductionnisme méthodologique. Pour Barzun, l'erreur résidait dans la simplification et la réduction d'ensembles complexes en parties distinctes, déconnectées de la valeur ou de la nature globale de l'objet d'étude.

Soixante ans après les réserves de Barzun, le saut qualitatif des capacités inférentielles de l'IA a radicalement déplacé ce seuil. Le développement récent des LLMs a relancé le débat sur les frontières du raisonnement cognitif. Ces modèles ont révolutionné le domaine par leur polyvalence inédite, témoignant d'une grande capacité d'adaptation à une vaste gamme de tâches. Toutefois, une divergence importante subsiste parmi les spécialistes quant à leur réelle faculté de compréhension. Pour Sejnowski [22], cette controverse reflète avant tout un désaccord conceptuel plus profond sur la définition même de l'intelligence. Il décrit ces systèmes comme un « miroir inversé » qui renvoie à la personne utilisatrice ses propres projections, expliquant ainsi la grande variation des interprétations et résultats obtenus.

Le débat sur les LLMs met aussi en évidence les divergences conceptuelles fondamentales entre les deux paradigmes, ainsi que leur impact sur la nature des savoirs produits. Pour expliciter le fonctionnement de ces modèles, Sejnowski [22] les distingue des systèmes symboliques en affirmant que « in a symbolic representation, all pairs of words are equally similar, which strips words of their semantics. In LLMs, words are represented by pre-trained embeddings in large vectors that are already rich in semantic information ». Les critiques adressées à ces modèles mobilisent également des arguments issus des débats historiques entre symbolistes et connexionnistes, soulignant la persistance de tensions sur la nature de l'intelligence et de la représentation.

Bender et Gebru [2] soutiennent qu'un « LM (*Language Models*) is a system for haphazardly stitching together sequences of linguistic forms it has observed in its vast training data, according to probabilistic information about how they combine, but without any reference to meaning : a stochastic parrot ». Cette critique fait écho à celle de Fodor et Pylyshyn [11], qui qualifiait déjà les architectures connexionnistes de « stochastiques » pour dénoncer l'absence de règles déterminant les relations entre les unités. Pour ces auteurs [11], seule la théorie symbolique et sa structure combinatoire permettent de satisfaire les conditions de validité inférentielle.

Cette absence de « référence au sens », pointée tant par Bender que par Fodor, constitue l'angle mort du recours à l'IA connexionniste comme instrument d'analyse et de preuve. En l'absence d'ancrage sémantique et de règles inférentielles explicites, l'algorithme génère des corrélations susceptibles de simuler la connaissance, sans pour autant constituer un savoir véritable. Dans cette optique, les résultats qualifiés d'« hallucinations » ne doivent pas être interprétés comme une rupture de la chaîne logique, comme ce serait le cas pour les systèmes symbolistes, mais comme une propriété intrinsèque des systèmes connexionnistes. Conceptuellement, ces derniers privilégient des résultats probabilistes qui ne sont soumis à aucune règle logique préétablie, de sorte que la machine peut être « cohérente » tout en étant « fausse ».

Au-delà des débats théoriques et techniques sur la structure matérielle, ces systèmes n'acquiescent de signification que lorsqu'ils sont articulés à un contexte d'utilisation, dans le cadre du traitement des données. Dans une perspective sémiotique, certains travaux [4; 5] interrogent la nature et la valeur des connaissances issues des processus de production et de traitement du *Big Data*. L'utilisation de ces systèmes rend encore plus ténue la ligne frontière entre la production procédurale et processuelle du sens, qui relève de processus de signification humaine et de procédures mécaniques, puisque l'intervention humaine est constante dans la machine, du développement des algorithmes au calibrage des données.

## 4 Conclusion

L'intégration du domaine computationnel dans les paradigmes des sciences humaines nécessite une discussion interdisciplinaire fondamentale. La compréhension des divergences entre les deux

principales approches de l'informatique cognitive permet d'analyser les algorithmes comme des constructions épistémologiques dont le fonctionnement doit être appréhendé à partir des principes conceptuels sur lesquels repose leur architecture.

En tant qu'artefact socialement situé, l'algorithme doit faire l'objet d'une analyse critique visant à déconstruire le « fétichisme technologique » [7] et l'idée d'IA comme un miroir objectif de la réalité, une vision qui occulte les asymétries épistémologiques, éthiques et politiques de ces dispositifs. Dans la recherche en SHS, ce fétichisme tend à assimiler la procédure informatique à une vérité immédiate. Or, la valeur indiciaire des résultats ne réside pas dans l'implémentation de l'algorithme, mais dans l'interprétation humaine qui l'investit. Sans conscience critique des limites de cette production procédurale, le chercheur risque de confondre la trace technique avec une connaissance validée, compromettant ainsi l'éthique et la rigueur de la production des savoirs.

## Références

- [1] ARENDT, Hannah. *Condition de l'homme moderne*. Original work published 1958. Pocket, 2012.
- [2] BENDER, Emily M., GEBRU, Timnit, McMILLAN-MAJOR, Angelina et SHMITCHELL, Shmargaret. « On the Dangers of Stochastic Parrots : Can Language Models Be Too Big ? » In : *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*. 2021, p. 610-623. DOI : 10.1145/3442188.3445922.
- [3] BENDER, Emily M. et GRISSOM, Alvin. « Power Shift : Toward Inclusive Natural Language Processing ». In : *Inclusion in Linguistics*. Oxford University Press, 2024, p. 199-222. DOI : 10.1093/oso/9780197755303.003.0010.
- [4] BONENFANT, Maude, LAFRANCE SAINT-MARTIN, Laura I. et CRÉMIER, Lucile. « Affected Data : Understanding Knowledge Production in Algorithmic Events ». In : *Global Media Journal – Canadian Edition 11*, no. 2 (2019), p. 66-78.
- [5] BOYD, Danah et CRAWFORD, Kate. « Critical questions for big data : Provocations for a cultural, technological, and scholarly phenomenon ». In : *Information, communication & society 15*, no. 5 (2012), p. 662-679. DOI : 10.1080/1369118X.2012.678878.
- [6] CORMEN, Thomas H., LEISERSON, Charles E., RIVEST, Ronald L. et STEIN, Clifford. *Introduction to Algorithms*. 4<sup>e</sup> éd. The MIT Press, 2022.
- [7] DEAN, Jodi. « Communicative Capitalism : Circulation and Foreclosure of Politics ». In : *Cultural Politics 1*, no. 1 (2005), p. 51-74. DOI : 10.2752/174321905778054845.
- [8] DUCOFFE, Guillaume, LÉCUYER, Mathias, CHAINTREAU, Augustin et GEAMBASU, Roxana. « Web Transparency for Complex Targeting : Algorithms, Limits, and Tradeoffs ». In : *ACM SIGMETRICS Performance Evaluation Review 43*, no. 1 (2015). DOI : 10.1145/2745844.2745896.
- [9] DUPUY, Jean-Pierre. *Aux origines des sciences cognitives*. Paris : La Découverte, 1999.
- [10] EVANS, Jonathan St. B. T., NEWSTEAD, Stephen E. et BYRNE, Ruth M. J. *Human Reasoning : The Psychology of Deduction*. Lawrence Erlbaum Associates, 1993.
- [11] FODOR, Jerry A. et PLYSHYN, Zenon W. « Connectionism and Cognitive Architecture : A Critical Analysis ». In : *Cognition 28*, no. 1-2 (1988), p. 3-71. DOI : 10.1016/0010-0277(88)90031-5.
- [12] HARTOG, François. *Régimes d'historicité : présentisme et expériences du temps*. Paris : Éditions du Seuil, 2003.

- [13] HOLTZMAN, Ari, WEST, Peter et ZETTLEMOYER, Luke. « Generative Models as a Complex Systems Science : How Can We Make Sense of Large Language Model Behavior? » 2023. DOI : 10.48550/arXiv.2308.00189.
- [14] HORKHEIMER, Max et ADORNO, Theodor W. *La Dialectique de la raison : fragments philosophiques*. Gallimard : Gallimard, 2013 [1944].
- [15] KRAYNAK, Janet. *Contemporary Art and the Digitization of Everyday Life*. Oakland : University of California Press, 2020.
- [16] KUHN, Thomas S. *La Structure des révolutions scientifiques*. Paris : Flammarion, 1972 [1962].
- [17] LEED, John. « Reviewed Work : Computers for the Humanities? A Record of the Conference Sponsored by Yale University on a Grant from IBM, January 22–23, 1965 ». In : *Computers and the Humanities 1*, no. 1 (1966), p. 12-14.
- [18] MILIC, Louis T. « The Next Step ». In : *Computers and the Humanities 1*, no. 1 (1966), p. 3-6. DOI : 10.1007/BF00188010.
- [19] MINSKY, Marvin et PAPERT, Seymour. *Perceptrons : An Introduction to Computational Geometry*. Cambridge : The MIT Press, 1972.
- [20] RUMELHART, David E. et MCCLELLAND, James L. *Parallel Distributed Processing : Explorations in the Microstructure of Cognition*. The MIT Press, 1986.
- [21] RUSSELL, Stuart J. et NORVIG, Peter. *Artificial Intelligence : A Modern Approach*. 3<sup>e</sup> éd. Prentice Hall, 2010.
- [22] SEJNOWSKI, Terrence. « Large Language Models and the Reverse Turing Test ». In : *Neural Computation 35*, no. 3 (2023), p. 309-342. DOI : 10.1162/neco\_a\_01563.
- [23] TURING, Alan. « Computing Machinery and Intelligence ». In : *Mind 59*, no. 236 (1950), p. 433-460. DOI : 10.1093/mind/LIX.236.433.
- [24] VARELA, Francisco J., ROSCH, Eleanor et THOMPSON, Evan. *The Embodied Mind : Cognitive Science and Human Experience*. The MIT Press, 1991. DOI : 10.7551/mitpress/6730.001.0001.
- [25] WIENER, Norbert. *La cybernétique : information et régulation dans le vivant et la machine*. Sources du savoir. Paris : Éditions du Seuil, 2014 [1948].