Signata
Annales des sémiotiques / Annals of Semiotics

17 | 2026
La traduction multimodale dans l'IA générative

L'énigme de l'IA générative : des contenus qui nous parlent et qu'on croit comprendre

Bruno Bachimont

Édition électronique
URL : https://journals.openedition.org/signata/6412
ISSN : 2565-7097

Éditeur
Presses universitaires de Liège (PULg)

Édition imprimée
ISBN : 978-2-87562-524-3
ISSN : 2032-9806

L'énigme de l'IA générative : des contenus qui nous parlent et qu'on croit comprendre

Bruno Bachimont

1 L'intelligence artificielle, depuis que le terme fut forgé dès 1955 en vue d'une école d'été au Dartmouth College (McCarthy, Minsky, Rochester, & Shannon, 1955), a pu être associée à différentes approches technologiques. La métaphore cérébrale fut historiquement la plus ancienne : fondée sur les propositions de McCulloch et Pitts dans leur article fondateur de 1943 qui énonçaient le modèle du neurone formel, elle donna le coup d'envoi à la tradition dite connexionniste, dont les moments forts furent le perceptron de Rosenblatt (Rosenblatt, 1958) et l'algorithme de rétropropagation du gradient (Rumelhart, Hinton, & Williams, 1986). Ce dernier permettait l'apprentissage dans des réseaux connexionnistes profonds, palliant ainsi les limitations intrinsèques du Perceptron (dénoncées par Minsky et Papert (1969)). Pourtant l'IA connut une première vague d'intérêt et de succès à travers une autre approche, dite symbolique, où l'on modélisait des connaissances sous une forme logique pour qu'elles puissent être utilisées par un outil de raisonnement informatique (notamment, les fameux systèmes experts fondés sur un moteur d'inférences, tirant les conséquences des connaissances formalisées qu'on lui avait fournies). Cette approche se vit délaissée par les outils fondés sur les modèles neuronaux dans les années 1980, à la suite du renouveau que l'algorithme de rétropropagation permit. Ces modèles ne reposent pas sur des connaissances qu'on formalise pour ensuite en inférer des conclusions, mais sur des exemples à partir desquels sont automatiquement extraites des connaissances formalisées (apprentissage automatique) pouvant ensuite être mises en œuvre sur de nouveaux exemples. Les approches s'inscrivant dans ce paradigme d'apprentissage sont multiples, alliant le *machine learning* (mobilisant des approches mathématiques et statistiques), le *deep learning* (mobilisant les modèles neuronaux) et enfin l'IA dite générative mobilisant des variantes de ce dernier.

2 Mais la diversité des approches techniques ne doit pas masquer le fait que l'objectif reste identique, à savoir que l'on cherche à faire faire à la machine des tâches pour lesquelles les êtres humains mobilisent des facultés cognitives et perceptives comme le raisonnement, la planification, la vision, la reconnaissance de forme ou d'objets, etc. On en déduit parfois un peu vite que les outils conçus en ce sens mobilisent ou présentent les mêmes facultés, ce qui n'est pas obligatoirement le cas ni ce qui est visé. Ce que l'être humain parvient à effectuer selon ses facultés, l'outil technique tente de le réaliser par d'autres moyens. Mais si l'outil parvient à des niveaux de performances identiques voire supérieures aux êtres humains, on peut légitimement se demander si ce qu'on appelait jusqu'à présent l'intelligence, à savoir le fait de présenter et d'utiliser ces facultés, ne se retrouve pas finalement dans ces outils. Ainsi, malgré quelques différences évidentes entre l'humain et la machine (entre un substrat biologique ou en silicium par exemple), elles ne seraient que superficielles et il n'y aurait pas lieu d'y voir un argument pour refuser à la machine une capacité de penser, interpréter, comprendre et réfléchir, bref de faire preuve d'intelligence.

3 Or, l'IA contemporaine, et en particulier les outils d'IA générative, présente de telles performances. Si bien que les questions qui furent posées dès l'origine de l'IA sur la nature de son projet, se retrouvent d'une actualité renouvelée. On se souvient par exemple que John Haugeland (Haugeland, 1989) précisait que l'IA visait à produire une véritable intelligence dans la machine :

> Peut-être devrait-on appeler l'intelligence artificielle « intelligence synthétique » pour être plus près du langage industriel. En effet, les diamants artificiels ne sont que des pâles imitations, alors que les diamants de synthèse sont de véritables diamants, mais fabriqués et non extraits du sol (ou bien prenez un parfum artificiel à la vanille par opposition, par exemple, à de l'insuline synthétique). En dépit de son nom, on veut atteindre la véritable intelligence et non une pâle imitation de celle-ci.

4 Ou encore, l'IA vise à proposer une approche générale de l'intelligence, dans le but non de reproduire la manière dont les êtres humains y parviennent, mais avec l'objectif d'en dégager les principes généraux pour les mettre en œuvre dans différents dispositifs. Ainsi, l'intelligence ne serait pas une propriété idiosyncrasique de l'humain, un propre de l'humain comme le dirait la logique antique (une propriété co-extensionnelle à l'humain), mais une catégorie générale pouvant convenir à différents systèmes, naturels ou artificiels. On retrouve l'inspiration originale de la cybernétique (Wiener, 1961) qui recherchait une théorie générale des systèmes, que ceux-ci soient naturels ou artificiels, biologiques ou mécaniques.

5 Si, jusqu'à il y a peu, ces affirmations relevaient de prise de position métaphysique, tant elles étaient au-delà des capacités effectives des machines à montrer de l'intelligence dans ce qu'elles réalisaient, les performances accumulées ces dernières années ne manquent pas de donner quelque réalité et consistance à ces considérations. En particulier, les outils d'IA génératives permettent de créer des contenus comme s'ils étaient le fait d'êtres humains et de les produire dans le fil d'une interaction, voire d'une conversation (quand ces outils sont des agents conversationnels e.g. ChatGPT) montrant un « esprit » d'à-propos et de suite dans les idées. Autrement dit, les agents conversationnels nous parlent et nous leur répondons ; nous sommes persuadés, au moins de prime abord, que la suite de sons entendus est une parole qui s'adresse à nous, que la suite de caractères lus sur l'écran constitue un texte méritant interprétation. Et, à l'instar de Socrate qui, dans le dialogue platonicien intitulé *Phèdre*

(Platon, 2006), fait remarquer qu'« il en est de même aussi pour les discours écrits : on croirait que ce qu'ils disent, ils y pensent » (276c), nous sommes persuadés que non seulement il y a un vouloir dire dans ce qui est produit, mais également quelque chose qui est dit, affirmé, et qu'il existe donc une réalité visée intentionnellement dans le sens vulgaire et phénoménologique. À savoir au sens vulgaire, avec l'intention de dire quelque chose, mais aussi, dans le sens phénoménologique, dans la mesure où les lettres ou sons produits renvoient à autre chose qu'eux-mêmes, étant à propos d'une réalité ainsi désignée.

6 Ce sont ces deux aspects qui vont nous intéresser ici : interroger le vouloir dire et ce qui est dit dans les productions de l'IA générative, problèmes que nous paraphrasons comme étant respectivement celui de l'énonciation d'une part, et de la modélisation d'autre part. Auparavant, nous reviendrons brièvement sur le principe de fonctionnement de la technique mise en œuvre, en insistant particulièrement sur la notion d'espace latent qui, même si ce concept est désormais ancien, est à la base des performances des IA génératives et permet de comprendre à la fois leur puissance mais également leur limitation intrinsèque. Limitation qui n'est pas à prendre comme une imperfection, mais comme une fin de non-recevoir aux différentes vaticinations que l'on peut proférer à propos de ces techniques, notamment quant à leur intelligence et leur concurrence avec l'humain. La réalité est sans doute à la fois moins gigantomachique et bien plus intéressante.

Le principe de l'IA générative

Connaissance et écriture

7 On peut reformuler le but et l'enjeu de l'intelligence artificielle en s'intéressant non pas à la notion d'intelligence, mais celle de connaissance. Non que cette notion ne soit pas également pourvue d'ambiguïté, mais il est plus facile de la caractériser du fait de sa relation privilégiée avec les outils de son expression, en particulier le langage et l'écriture. La connaissance correspond à ce qu'il faut savoir, ce dont il faut être doté pour être en mesure de réaliser une tâche, assurer une fonction, atteindre un objectif. Or, pour évoquer une connaissance, il faut être capable de la cerner et d'en parler : une connaissance non formulable est inassignable. Que ce soit l'observateur qui constate que si telle entité est capable de faire ceci ou cela, alors elle possède la connaissance que cet observateur explicite pour rendre compte de cette capacité, ou une entité qui fait elle-même état de la connaissance qu'elle possède et met en œuvre, une connaissance n'existe comme telle que si elle est explicitée. Cela ne renseigne pas sur la nature de la connaissance ni ne permet de la définir, mais que dès lors que la notion en est mobilisée et qu'on évoque des connaissances, il y a toujours quelqu'un qui se soucie d'en expliciter le contenu de ces dernières, que ce soit dans un rapport réflexif ou d'observation externe.

8 Ce rapport à l'explicitation explique que, dès lors qu'on veut doter une entité de connaissances, on tente de lui transmettre les explicitations associées. Le langage est évidemment un médium privilégié d'explication et de transmission des connaissances même si ce n'est pas sa seule fonction. L'écriture permet de conférer à cette explicitation une permanence matérielle et objective nécessaire à la capitalisation culturelle d'une part et à sa disponibilité dans les sujets d'autre part. On connaît le

thème de la pensée comme langage mental, voire comme écriture, qui sourd d'une longue tradition (Fodor, 1975 ; Panaccio, 1999). Ce thème rentre en une résonance particulière dès lors qu'on dispose de machines qui reposent quant à elles sur le calcul, c'est-à-dire une forme d'écriture[1] (Lassègue, 2013, 2023 ; Lassègue & Longo, 2025).

Des connaissances implicites apprises ou des connaissances explicites modélisées

9 L'enjeu est alors de fournir une représentation scripturaire des connaissances pour que la manipulation calculatoire puisse en tirer parti. Autrement dit, on passe de l'écriture des connaissances exprimées en langue à l'écriture du calcul pour qu'elles soient exploitées par l'inférence calculatoire. La difficulté est alors de disposer des expressions pour les formaliser et les exploiter. Deux manières de faire se dégagent rapidement : soit, à l'instar du *Ménon* (Platon, 1999a), on dispose non seulement d'exemples mais aussi de leur abstraction en connaissances générales et universelles, et alors il faut les modéliser et les formaliser et définir comment en tirer des inférences. C'est l'approche historique de l'IA dite symbolique, dont les systèmes experts furent un temps l'image de marque (*Mycin* en étant le parangon (Shortliffe, 1976)) pour être ensuite anonymisée et sécularisée dans les techniques du web des données appelé également web sémantique (Berners-Lee, Hendler, & Lassila, 2001). Mais quand, comme le déplore Socrate, on ne dispose que d'exemples, il faut alors se mettre en quête des connaissances qui peuvent les généraliser et dont ils sont la manifestation ou la mise en œuvre. Les techniques d'apprentissage permettent, à partir d'un jeu de données dit d'apprentissage, de dériver un modèle appris que l'on peut ensuite appliquer sur de nouvelles données. Passer des données d'apprentissage au modèle appris nécessite de recourir à un modèle d'apprentissage. Les réseaux de neurones (*deep learning*), le *machine learning* sont des techniques proposant des modèles d'apprentissage. La difficulté de ces approches est que les connaissances apprises sont présentes dans le modèle appris, mais souvent de manière non explicite : on ne sait pas les lire ou les retrouver dans le modèle. Le passage du calcul comme écriture à l'écriture des connaissances comme savoir explicite ne se fait pas, fixant par conséquent une limite à l'intelligibilité de ce qui est calculé et à ce qu'on peut valider.

10 Ces outils revêtent donc une opacité épistémique (on comprend le principe de fonctionnement mais on ne peut comprendre comment les résultats sont produits (Humphreys, 2004)). Cette opacité est donc plus délicate à surmonter que la simple ignorance des principes de fonctionnement : c'est l'idée que la compréhension de ces derniers ne suffit pas à comprendre la manière dont est obtenu le résultat. Cette opacité épistémique se redouble également d'une opacité représentationnelle (Humphreys & Alvarado, 2017) ou interprétative (Bachimont 2018) dans la mesure où, même si on peut comprendre comment mécaniquement ou algorithmiquement le résultat est produit, on n'est pas pour autant capable d'interpréter le résultat et savoir le rapprocher du monde dont les données manipulées sont les données. La dé-sémantisation nécessaire pour le calcul impose une étape de réinterprétation des résultats qui n'est pas aisée dans le cas des approches connexionnistes. C'est la raison pour laquelle le courant de l'IA « explicable » (XAI pour eXplainable AI) s'est constituée et constitue un domaine actif de recherche (Mattioli, Cinà, & Pelillo, 2024). Outre sa dimension technique qui relève d'une complexité certaine, les enjeux éthiques pour

une utilisation assumée et responsable de ces outils donnent encore une importance accrue à ses approches (Pegny, 2024).

11 Néanmoins, ces approches ont l'immense avantage de justement proposer une manière d'extraire des connaissances, alors qu'on ne les a pas explicitées, d'exemples où ces connaissances sont mobilisées de manière implicite. C'est notamment le cas de certaines tâches d'expertise qui permettent difficilement la généralisation, étant la rencontre entre une situation singulière et une expérience accumulée.

12 L'IA générative est une forme particulièrement aboutie de cette deuxième approche. Elle correspond à l'une des trois manières classiques de mobiliser les approches par apprentissage :

- Reconnaissance d'une nouvelle donnée :
  à partir des données existantes réputées appartenir à des catégories préexistantes et explicites, comment définir la catégorie d'une nouvelle donnée ? C'est une tâche souvent présentée comme une prédiction, car il s'agit de prédire, reconnaître la catégorie d'une donnée connaissant celles des données d'apprentissage. Il s'agit alors d'apprentissage supervisé (par des classes déjà connues).
- Structuration des données :
  à partir des données existantes, on apprend des catégories permettant de les classer et de les différencier. Une nouvelle donnée sera alors classée selon ces catégories apprises. Cet apprentissage des catégories est non supervisé (on ne connaît pas a priori les classes) mais il peut également faire de la prédiction car une nouvelle donnée peut alors être rapprochée des catégories dégagées par un tel apprentissage non supervisé.
- Génération de données :
  à partir des données existantes, on apprend à produire de nouvelles données qui leur sont le plus proche possible, le plus ressemblant. Cette dernière tâche est différente des deux premières dans la mesure où on invente des données alors que les deux premières travaillent avec des données fournies et réelles.

L'IA générative

13 L'IA générative, celle donc qui engendre de nouvelles données se réalise à travers plusieurs approches que l'on peut regrouper en 3 modèles principaux.

14 Le premier modèle correspond aux *Generative Adversarial Networks* (GAN) (Goodfellow *et al.*, 2014). Le principe est de faire travailler deux réseaux de neurones en compétition : le premier (c'est le « faussaire ») apprend à générer des contenus propres à tromper le second (c'est l'« expert ») qui est entraîné quant à lui à discriminer entre de faux contenus et de véritables. Il s'agit dans les deux cas d'un processus d'apprentissage supervisé habituel, mais l'expert apprend à partir de véritables contenus dont on connaît la qualité (vrais ou faux) alors que le faussaire apprend à partir des réponses fournies par l'expert. On comprend aisément pourquoi on produit alors de nouvelles données.

15 Le second modèle correspond aux modèles de diffusion qui possèdent une grande efficacité, et donc popularité, dans le domaine des contenus non textuels, notamment le son et l'image. Le principe est d'ajouter du bruit, c'est-à-dire de l'information aléatoire aux données pour ensuite apprendre à le retirer pour retrouver des contenus proches des données d'origine. Autrement dit, on se déplace dans l'espace des données

initiales, en s'éloignant arbitrairement des données d'apprentissage grâce au bruit, pour ensuite revenir au sous-espace propre à ces données. L'apprentissage porte donc sur la capacité à débruiter les images. En partant d'une image donnée par exemple, le débruitage aboutit non à l'une des données initiales d'apprentissage mais à une donnée voisine, inventée. Ce modèle peut être complexifié en prenant en compte des catégories d'images, et donc des types de débruitage (une classe permettant de débruiter pour obtenir des images de chats, une autre de chiens, etc.), voire des catégories plus complexes mélangeant plusieurs caractéristiques. Pour cela, on ne raisonne plus sur l'espace des données, mais un espace différent, un espace latent, où les données sont projetées de manière à introduire des proximités entre elles reflétant leur appartenance commune à des catégories. Le débruitage est alors conditionné à partir de cet espace et de ces catégories. C'est ainsi que on peut obtenir dans cet espace des images de chat sur un tapis plutôt que dans le jardin.

16 Enfin, le dernier modèle est celui correspondant aux modèles dits « *transformers* » et aux algorithmes gérant l'attention, c'est-à-dire les dépendances entre les données de la situation traitée. Le principe de ces modèles s'inspire des réseaux de neurones capables de poursuivre ou engendrer de nouvelles séquences de données à partir de séquences fournies en entrée. L'exemple le plus connu à présent est celui des outils tels que ChatGPT qui renvoient des chaînes de caractères répondant à, complétant, des suites de caractères fournies en entrée (les requêtes ou *prompts*).

17 Ces outils reposent également sur le principe des espaces latents, c'est-à-dire une autre représentation des données initiales dans un espace de dimension plus réduite d'une part et permettant d'autre part de coder les données afin que leur distance mutuelle reflète leur proximité sémantique (au sens que l'on voudra, mais la plupart du temps il s'agit de l'appartenance à de mêmes classes). Les étapes sont les suivantes :

- Un formatage des données initiales, qu'on appelle une tokenisation puisqu'il s'agit d'une mise en *tokens* ou unités élémentaires. Dans le cas des textes, il s'agit de coder un contenu textuel (au sens large, avec des graphiques, des caractères, du code informatique, etc.) au moyen de ces *tokens* : il correspondra à une succession de tels *tokens*. L'intérêt est de considérer seulement quelques dizaines de milliers de *tokens* possibles (100 000 dans ChatGPT4 (Alammar & Grootendorst, 2024)) plutôt que l'ensemble des formes textuelles possibles d'une langue et des contenus associés (de l'ordre de 500 000 mots, voire à quelques millions en ajoutant ceux venant des différents langages informatiques, graphiques, etc.). On réduit alors la dimensionnalité du problème (une dimension par mot à une dimension par *token*). Mais pas de manière suffisante.
- C'est la raison pour laquelle on effectue un plongement dans un espace latent de dimension généralement fixée à 512. Cet espace latent est construit également par apprentissage de manière à coder les *tokens* en fonction de leur contexte d'usage : deux *tokens* seront d'autant plus proches qu'ils sont utilisés dans de mêmes contextes. Différents algorithmes permettent de le faire, le plus connu étant *Word2Vec* (Mikolov, Chen, Corrado et Dean, 2013). Cette approche repose sur une théorie sémantique classique de la langue naturelle, à savoir la sémantique distributionnelle de Harris, où l'enjeu est de modéliser la sémantique d'un terme par sa distribution dans les corpus d'usage[2] (Gastaldi, 2021 ; Harris, 1954). Les axes sémantiques appris dans cet espace latent ne sont pas toujours faciles à interpréter, voire jamais. Mais le principe est que les 512 dimensions de cet espace sont autant de dimensions sémantiques permettant de décrire le sens d'un *token*, qui est donc un vecteur de cet espace.
- Un dispositif d'encodage permet ensuite d'apprendre à encoder dans l'espace latent.

- Un dispositif de décodage apprend à produire une séquence en fonction de la séquence encodée. Les deux dispositifs, d'encodage et de décodage, sont entraînés à partir des séquences fournies en entrée et des réponses pertinentes attendues (une traduction, une suite de la séquence d'entrée, etc.).
- L'originalité de l'encodage et du décodage est de pouvoir faire dépendre les *tokens* des séquences d'entrée et de sortie les uns des autres par un mécanisme dit d'attention (Vaswania *et al.*, 2017) permettant de traiter des dépendances longues d'une part et en parallèle d'autre part. Ces dépendances longues sont tant spatiales (par exemple, repérer des motifs ou formes dans une image) que temporelles (repérer des dépendances entre éléments d'une séquence temporelle, musicale ou langagière par exemple). Ces types de dépendances étaient jusque-là respectivement traités par les réseaux convolutifs (CNN ou convolutionnal neural networks) et récurrents (RNN ou recurrent neural networks). Le parallélisme dans l'analyse des séquences temporelles est important dans la mesure où tous les items de la séquence d'entrée sont traités en même temps, ce qui permet d'avoir un temps de réponse quasi constant quelle que soit la longueur de la séquence d'entrée. Mais une difficulté devient alors d'avoir, pour chaque item traité en parallèle, une information sur les voisins qui le précèdent pour le prendre en compte dans le traitement du mot et garder ainsi l'information que les items sont produits en séquence. Cette information est traitée en IA générative en intégrant dans l'encodage du token sa position.
- Une traduction des vecteurs obtenus (produits par le décodage) dans l'espace du langage cible (qui peut être aussi varié que l'on veut, dès l'instant qu'on sait passer de l'espace latent à celui-ci). En particulier, l'espace latent permet de représenter des vecteurs correspondant à l'encodage de textes, et d'autres correspondant à l'encodage d'images. L'essentiel est qu'on ait des vecteurs dans ces espaces, et qu'on soit capable de calculer leur proximité (ou distance) et de la manipuler. De même, à partir d'un vecteur, on peut choisir de le décoder en image, texte ou autre chose.

18 Ce qu'il faut retenir de ces dispositifs pour comprendre le statut de ce qui est produit par de tels outils peut être résumé par les éléments suivants :

- La codification des données d'entrées : la mise en token ;
- Le plongement (*embedding*) des tokens dans un espace latent ;
- Le calcul des vecteurs composant la réponse dans cet espace du fait des proximités vectorielles constatées ;
- La traduction des vecteurs obtenus dans l'espace du langage cible.

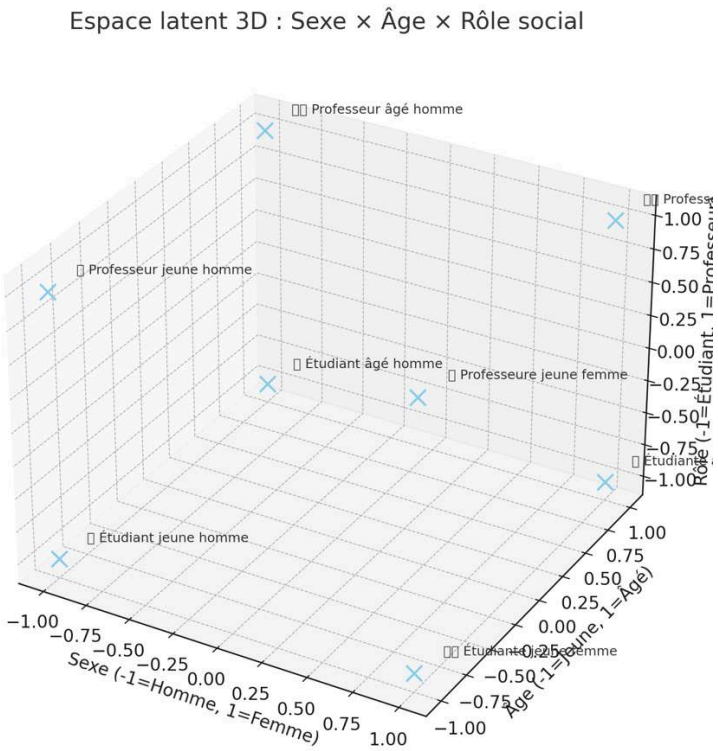
19 Nulle part le processus ne dépend d'autre chose que des données initiales et de l'encodage dans l'espace latent, ce dernier possédant une structure restant opaque pour l'interprétation car ses dimensions sont trop nombreuses pour être labellisées et combinées d'une part, et parce que le lien avec les données initiales n'est pas conservé d'autre part. Le résultat produit s'évalue uniquement par sa pertinence et non par son mode de production.

L'espace latent

20 La notion d'espace latent est un concept très puissant pour deux raisons essentielles : d'une part, il réduit la dimensionnalité des données et permet de rendre les calculs plus efficaces et de mieux les différencier. En effet, si on prend l'exemple d'un texte, il est écrit à partir d'un dictionnaire comprenant par exemple 500 000 mots. À chaque mot utilisé correspond un vecteur de dimension 500 000 (une par mot du dictionnaire), où

toutes les coordonnées sont nulles sauf celle correspondant à la dimension du mot utilisé. Un texte sera donc codé par autant de vecteurs que de mots le composant. On pourra donc repérer des textes qui utilisent certains mots plutôt que d'autres, représenter un texte par les mots majoritaires, mais on ne pourra pas rendre compte du fait que certains mots ont des proximités sémantiques, et que leur co-présence renforce cette proximité et ce sens partagé. L'idée est donc de réduire la dimension du codage vectoriel pour forcer à ce que des mots de sens voisins aient des vecteurs dont la distance vectorielle soit d'autant plus réduite. Si chaque mot, un parmi 500 000, doit être codé sur 512 dimensions (le nombre de dimensions le plus fréquent dans ces approches), on devra le projeter sur ces 512 dimensions par des coefficients traduisant la pondération / contribution de chacune de ces dimensions dans le sens du mot. Un mot sémantiquement voisin aurait alors un codage qui serait également projeté sur ces 512 dimensions avec des coefficients voisins du précédent. La question est alors de savoir comment effectuer un tel codage d'un mot parmi 500 000 sur les 512 dimensions envisagées. L'approche est de ne pas fixer *a priori* ces dimensions mais de les apprendre en exploitant la proximité d'usage des termes : comme évoqué ci-dessus, l'intuition est que des mots sont voisins s'ils partagent des contextes identiques. Le codage déterminera donc les coefficients respectifs en fonction de ces voisinages communs rencontrés. Les dimensions de l'espace latent, fixées arbitrairement à 512, n'ont pas d'explicitation sémantique immédiate, car elles résultent d'une optimisation des coefficients en fonction des voisinages rencontrés dans les textes considérés. Une fois encodés, on ne sait pas dire en quoi les termes sont voisins mais on peut exploiter le fait qu'ils le sont.

Fig. 1



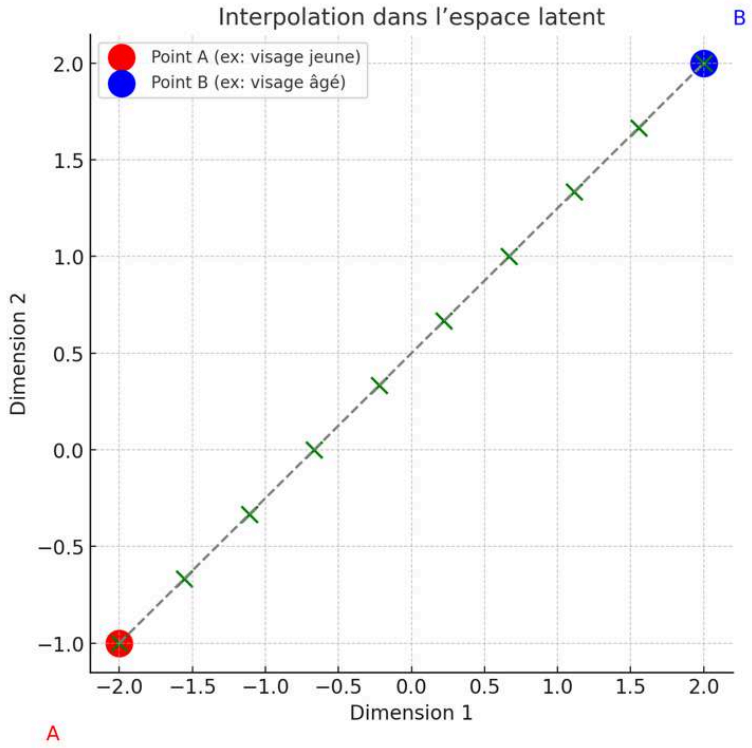
Espace latent 3D.

21 Par conséquent, un tel espace latent constitue un espace codant les données où leur distance mutuelle traduit une proximité sémantique. Ainsi, si deux données (par exemple deux mots), correspondent à deux points de cet espace, les points intermédiaires, sur la droite reliant ces points correspondent à des données ayant un sens voisin partagé par ces deux points.

22 On comprend les conséquences de cette approche :

- Les points (mots, pixels, sons, etc.) interpolés correspondent à des données irréelles mais réalistes ;
- On peut suivre une trajectoire dans l'espace latent, en poursuivant un axe « sémantique » défini sur les différentes dimensions de cet espace.
- L'interpolation fait que les données ressemblent aux points d'entrée correspondant aux données initiales, si bien que le point interpolé est interprétable.
- Néanmoins, puisque ce point est interpolé, il peut ne correspondre à rien de réel ni même de possible, l'espace latent n'étant qu'une approximation et optimisation de la représentation des données initiales.

Fig. 2



Interpolation dans l'espace latent.

23 Cette approche est très féconde comme le démontre les multiples déclinaisons et applications que l'on peut rencontrer. Si le réalisme des résultats est aisément compréhensible, il est plus délicat en revanche de déterminer le statut qu'il convient de donner à ces résultats. Ce sera l'objet de notre dernière partie.

Énonciation et modélisation

24 Les IA génératives sont souvent utilisées dans un fil de discussion ou d'interaction où l'outil répond à des requêtes (des prompts) que l'on peut progressivement affiner, donnant lieu à quelque chose ressemblant à une conversation. La puissance de l'espace latent est qu'il est possible de demander à l'interpolation d'emprunter un style, une approche « à la manière de », où l'outil feint de parler à la place de quelqu'un qu'il n'est pas. Nous y reviendrons, mais c'est une dimension de la réflexivité que de pouvoir se mettre à la place d'un autre pour parler à sa place, comme si on était lui (Paolucci, 2025), de renvoyer ainsi à une niveau supérieur, un « méta niveau[3] » qui conditionne et positionne le niveau de ce qui est dit.

25 Mais pour définir plus avant le statut de ces productions par l'IA générative, nous allons les aborder sous deux aspects complémentaires : d'une part, en tant qu'il s'agit d'une action correspondant à une production de contenu, une sorte d'énonciation, d'autre part en tant qu'il s'agit de représentations produites consistant en un discours sur quelque chose, sur une réalité (quelle que soit sa nature). Pour l'aborder, nous allons mobiliser une première analyse en termes de dire, dit et désigné pour caractériser les contenus produits par ces outils.

Le dire, le dit et le désigné

26 Dans toute production de contenu, on peut distinguer trois niveaux d'analyse de ce contenu :

- Le niveau du dire comme action ou événement selon lequel le contenu est énoncé comme tel. Le dire est une irruption, une instauration qui confère au contenu un statut de réalité communicationnelle au-delà de la simple virtualité qu'il possède sinon. Il y a donc un avant et un après du dire.
- Le niveau du dit qui correspond à la matérialité du contenu en tant qu'il correspond à un substrat matériel façonné selon certaines formes conventionnelles à travers lesquelles il est envisagé et interprété. Comme nous l'avons dit ailleurs (Bachimont, 2017), le contenu est la rencontre entre un substrat matériel de manifestation et une forme sémiotique d'expression. Le principe est que le substrat matériel confère au contenu sa perceptibilité, sa manipulabilité technique, tandis que la forme sémiotique confère au contenu sa capacité à faire signe et sens, c'est-à-dire à se rapporter à ce qui n'est pas lui, à propos de quoi il est un contenu.
- Enfin, le niveau du désigné qui découle directement du dit comme contenu. Il s'agit de ce que le contenu prétend affirmer à propos de ce qu'il n'est pas mais auquel il renvoie. Le désigné est ainsi la référence, ou le représenté si le dit est le représentant ou la représentation.

27 Ces trois niveaux sont complémentaires mais distincts. Le dire est avant tout un faire, il s'inscrit dans un comportement et une action prenant place dans un monde matériel, qui crée un objet matériel, le dit. Le dit est une chose, une entité matérielle, un contenu. Résultat d'un faire, il n'en est pas un. Enfin, le désigné est une entité du monde visé et peut être à la fois une chose, un faire, un agencement de faire et de choses. Néanmoins, on peut s'interroger sur l'articulation de ces derniers : comment un contenu, de l'ordre du dit, peut-il être une chose en rapport avec d'autres qui ne sont

pas des contenus ou des dits et en « parler » ? De même, en quoi est-ce qu'un faire de l'ordre du dire peut avoir un rapport avec des faire qui appartiennent au monde sans pour autant être des productions de contenus, c'est-à-dire des dire ?

28 Concernant la dualité entre les choses dites et non dites, il est établi que les choses dites entretiennent les unes avec les autres des relations linguistiques déjà bien étudiées et élucidées, notamment leur capacité de juxtaposition (l'axe syntagmatique) et de substitution (l'axe paradigmatique). Mais l'essentiel est que les choses dites s'impliquent les unes les autres, une phrase en appelant un autre, un texte une interprétation, un discours, une réponse (en s'intéressant ici non au faire qui a produit ces interprétations, discours, réponses, mais à son résultat, le dit produit). Si on voit bien que l'on peut sortir de l'ordre du discours pour entrer dans celui du faire (l'instruction, l'ordre, etc.), le rapport du mot au faire et à la chose reste une énigme. Si l'on suit la sémiotique peircienne par exemple, le dit ne peut entraîner qu'un autre dit dans une sémiose illimitée qui ne peut se résoudre *in fine* que dans une action, exténuation ultime du dit et de son interprétation, mais à la condition d'en faire elle-même un signe. De même, le faire comme dire semble d'un ordre bien distinct de l'ordre du faire comme geste : le geste et la parole sont deux ordres d'intervention dans le monde qui n'interfèrent pas de manière évidente ; si parler est agir, c'est agir vis-à-vis d'interlocuteurs qui parlent également. De même, agir par ses gestes, c'est modifier l'organisation matérielle des choses autour de nous.

29 Mais on peut aussi considérer que, si énigme il y a, il n'y a pourtant aucun obstacle de principe entre les choses dites et non dites, les dire et les faire. C'est précisément là que l'IA générative est intéressante car elle montre leur proximité. En effet, l'IA générative propose de nouveaux contenus sous la forme de nouveaux pixels (pour les images) ou de nouveaux agencements de caractères (pour les textes). Or, ces contenus sont des choses matérielles agencées par le processus technique : nulle part les pixels, les caractères, sont considérés comme des signes, simplement comme des occurrences dans un espace idéal et virtuel de types eux-mêmes idéaux et virtuels (Bachimont 2025b). Ces contenus, compris comme signes, sont assemblés et créés par ajustements mécaniques de symboles (i.e. ici, des occurrences matérielles de types idéaux, par exemple la lettre « a »). C'est que les symboles constituent un système complet et autonome d'explicitation de contenus permettant de les formuler indépendamment de leur signification et de leur appartenance à l'ordre du dit. Nous rejoignons en cela les analyses de Lassègue et Longo (2025) qui montrent comment l'alphabet, en permettant de lire une langue qu'on ne connaît pas et d'en reproduire les sons, a permis de dissocier la matérialité symbolique de sa signification. Il n'est pas nécessaire de connaître un mot au préalable pour être capable de le lire. Cette dissociation crée en retour l'événement de la sémantisation où l'agencement des symboles peut produire des expressions renvoyant à ce qu'on n'a encore jamais dit ni imaginé pouvoir dire, tout comme à des expressions ne voulant rien dire. La sémantisation s'empare de ces agencements pour y retrouver un sens qui ne leur est pas préalable[4]. Cela renvoie à la remarque de Platon (1999b) dans le *Théétète* (203b et seq.) où il constate que des lettres sans signification composent néanmoins des syllabes pourvues de signification, le problème étant de comprendre comment survient cette signification et comment elle s'articule à ces composants insignifiants[5].

30 L'ambiguïté vient du fait que si le contenu est bien un objet comme un autre, pourquoi le recevons-nous comme l'énonciation de quelque chose qui est dit ? Autrement dit,

pourquoi considérons-nous que les contenus produits par l'IA disent quelque chose, à savoir sont des dits se rapportant à un désigné et émanent d'un dire ?

31 On a là le double problème de l'énonciation et de la modélisation : en quoi le contenu produit comme agencement de symboles ressortit-il d'un dire, d'une énonciation (un faire produisant un énoncé) ? Par ailleurs, en quoi le dit produit comme agencement de symboles renvoie-t-il à un désigné et en affirme quelque chose, comme s'il en était un modèle (représentation pouvant tenir lieu du désigné) ? Bref, quand nous sommes confrontés à de tels contenus issus de l'IA générative, pourquoi ça nous parle d'une part, et pourquoi pensons-nous que ce contenu dit quelque chose à propos d'un désigné ?

Le problème de la modélisation

32 Prenons le risque de caractériser ce qu'est un modèle, tant les déclinaisons disciplinaires peuvent apporter des nuances à cette notion fondamentale. Reprenant l'approche que nous défendons dans (Bachimont 2023), mais proche également de (Varenne, 2012), un modèle est une représentation d'un objet ou d'une situation concrets, empiriquement situés, permettant de les remplacer dans l'interaction que nous avons avec cet objet ou cette situation. L'idée principale est de pouvoir interroger le modèle en lieu et place de la réalité (objet ou situation) qu'il représente. L'interrogation peut être par exemple une expérimentation : au lieu d'expérimenter dans des conditions réelles, on procédera à une simulation grâce aux modèles, comme le permettent par exemple les modèles numériques de différents dispositifs. Mais l'interrogation peut être épistémique : le modèle permet de raisonner sur le réel représenté et d'en tirer une compréhension globale ; le modèle est alors une source d'intelligibilité. Le modèle est à la fois une idéalisation, car il ne reprend pas toutes les caractéristiques de la réalité considérée mais l'abstrait selon différents points de vue théoriques, et une concrétisation car le modèle se veut synthétiser à l'échelle de l'objet ou situation représentés des points de vue hétérogènes pour lesquels on établit des approximations, compromis, extrapolations. Ainsi mêlera-t-on des considérations à la fois économiques et psychologiques sur le comportement d'usager pour mieux les comprendre (finalité épistémique) ou les prédire (finalité expérimentale), ou des considérations de thermodynamique et résistance des matériaux dans un moteur ou une centrale. Le modèle pallie l'absence d'une théorie globale du réel en proposant localement, à l'échelle d'un objet ou d'une situation, une manière de synthétiser les différents points de vue et de les considérer dans leur totalité. Le modèle est ainsi établi selon un point vue pratique adopté, la finalité pour laquelle il est élaboré : ce point de vue permet d'établir les compromis ou les approximations nécessaires pour faire cohabiter des considérations que les théories établissent indépendamment les unes des autres (la théorie de l'électricité et celle de la gravitation, ou encore un modèle de l'utilisateur intégrant ses connaissances et son environnement).

33 La plupart des usages que nous pouvons avoir des IA génératives reposent sur la production mécanique de symboles que la lecture sémantise pour les rapporter à une réalité visée. La question est de savoir si cette production symbolique est un modèle de la réalité visée, que l'on peut donc interroger pour en savoir plus sur elle. Sur une question posée concernant des événements, une théorie, une personne, etc., la réponse sera interprétée comme une suite d'affirmations se rapportant à ces derniers. De plus,

l'interaction se faisant dans une sorte de dialogue où les questions (prompts) de l'utilisateur se succèdent, nous nous comportons vis-à-vis de l'outil d'IA générative comme vis-à-vis d'un modèle qu'on interroge pour appréhender la réalité dont non seulement il parle, mais qu'on considère qu'il représente.

34 Pourtant, à l'aune du processus que nous avons rappelé, on peut constater que nulle part n'est considérée ou prise en charge la relation articulant les représentations symboliques produites avec le sens qu'on leur prête. En revanche, l'espace latent qui est au fondement de ces approches se comporte bien comme un modèle, mais un modèle non pas du réel mais des documents et données qui ont été utilisés lors de l'apprentissage. Ainsi l'expression de modèle appris est-elle bien adaptée : l'espace latent permet de représenter les documents sources et on peut l'interroger et avoir une réponse, comme si l'on consultait les documents mais sans avoir à le faire. C'est la raison pour laquelle les réponses fournies ne sont pas toujours sourcées ou référencées d'une part, et que l'outil génératif n'est pas un outil de navigation dans une bibliothèque mais plutôt une simulation d'un lecteur idéal ayant lu tous les ouvrages de la bibliothèque et qu'on peut donc interroger sur ce qu'ils disent et ce dont ils parlent. Mais les documents ne sont pas présents dans l'espace latent : on ne les retrouve pas. C'est en cela que l'IA générative, à travers les usages qu'on en dégage, s'inscrit à la fois dans la tradition de la documentation mais pour s'en échapper. En effet, la documentation a pour but de fournir des informations qualifiées en réponse à des questions posées par des usagers (Bachimont, 2017 ; Briet, 1951). Les réponses fournies sont les documents apportant ou comportant ces informations. Mais, peu importe la nature du document, sa forme, son support, dès lors que l'information est pertinente et qualifiée. L'IA générative de ce point de vue s'émancipe du format documentaire en reformulant le contenu et en délivrant directement l'information, indépendamment de son origine documentaire. La classification et l'indexation sont remplacées par le plongement dans l'espace latent qui compile l'information contenue dans les documents. De la même manière que la documentation pouvait renvoyer des documents au contenu douteux, l'IA générative peut renvoyer des réponses fautives. Mais, dans la première, l'origine documentaire étant présente, on peut aller l'interroger directement, la documentation produit des documents ! En revanche, l'IA générative ne permet pas de retourner à l'origine, on perd le lien et la capacité de vérifier les sources et de mettre en perspective les informations reformulées.

35 Avec Antonio Somaini[6] (Somaini, 2024), on peut considérer l'espace latent comme une méta-archive : une archive car elle renvoie aux documents issus du passé, mais une méta-archive dans la mesure où les documents initiaux sont rendus inaccessibles, noyés dans l'espace latent.

36 Mais si l'IA des données ne modélise que les données, et si le modèle obtenu correspond aux connaissances permettant de manipuler ces données pour répondre à leur place, ce modèle n'est pas un modèle du monde ni des entités dont les données sont les données. Or, nous ne pouvons nous empêcher de les lire de cette manière : toute production sémiotique, si elle est perçue ou reçue comme telle, est comprise comme une expression sur un monde. Nous sommes soumis à une illusion transcendantale (Kant, 1997/1781) où, bien que nous sachions que ce soit une illusion, nous interprétons le résultat produit non comme une extrapolation de données, mais comme une donnée sur le monde. À l'instar du *Phèdre* de Platon (Platon, 2006), les textes et images produites semblent être des énoncés, ils semblent penser à ce qu'ils disent ou montrent

alors qu'ils ne font que répéter la même chose dès qu'on les interroge. Cependant, le trouble contemporain s'accroit puisque, détournant la critique de Platon, les outils comme les IA génératives fondées sur les LLM donnent l'illusion d'un dialogue, répètent le contenu compilé dans la méta-archive de manière toujours renouvelée mais pour toujours dire des choses semblables.

Le problème de l'énonciation

37 Nous avons déjà relevé qu'une IA n'a de monde que réduit à ses données, et que ces dernières sont sans arrière-plan ni contexte. Autrement dit, les données représentant le monde n'ont pas de monde dans lequel elles sont inscrites et s'interprètent. Quand la machine répond ou produit un énoncé, cela ne dépend pas de la personne avec laquelle elle interagit, ni de la situation, ni de lieu ou du temps. L'interaction machinique n'est pas située et n'a pas de milieu. Or, il n'y a d'énonciation que si on s'engage et assume, pour le meilleur et pour le pire, l'énoncé produit : on peut en être tenu pour responsable. La machine ne manipulant que des informations codées dans un espace latent ne s'engage ni n'est responsable de ce qu'elle produit.

38 Cependant on pourrait considérer que cette argumentation repose sur le mythe du signifié et de l'intentionnalité. (Paolucci, 2025) fait ainsi remarquer que la capacité cognitive ne se reconnaît pas à la capacité à dire « je » ou à être un ego totalisant les expériences vécues, mais bien plutôt à la capacité de renvoyer à d'autres énonciations que la sienne, quand on parle à la manière de quelqu'un d'autre. Or, les outils d'IA générative ont montré qu'ils savaient produire des énoncés appartenant à un répertoire stylistique et argumentatif propre à un auteur et à s'exprimer selon des canons particuliers. Ne faut-il pas y voir la preuve que la machine énonce alors aussi bien que nous et que l'engagement énonciatif que nous avions évoqué ne serait qu'une modalité existentielle, un vécu qui ne constituerait pas en tant que tel une compétence cognitive ni ne définirait la pensée ? Cela reste discutable car on remplacerait ici une totalisation par l'ego (l'ego étant la totalisation de l'expérience du sujet) par une totalisation symbolique par la machine (l'espace latent totalisant le savoir documentaire). En effet, la critique du « je » vient de fait que ce serait une structure transcendantale totalisant l'expérience, étant celui dont c'est l'expérience et à qui cette expérience survient. Mais si la raison n'est pas le fait de dire « je » mais de s'exprimer « à la manière de », la totalisation de l'expérience n'est plus assumée par le sujet mais par l'encyclopédie (au sens de Eco, 1985, 1988), l'espace latent étant le support et la matérialité de cette totalisation. Mais le « je » est disqualifié précisément parce qu'on fait appel à la totalisation, qui est aussi aporétique quand on parle de l'encyclopédie qui prend un statut transcendantal dès qu'on veut déborder le simple principe de la base documentaire compilée. Et si on se rapporte à cette dernière, considérant le modèle appris à partir des données et documents d'apprentissage, on tombe sur le problème selon lequel les données sont décontextualisées, isolées et séparées pour être réunies dans un corpus d'apprentissage. L'apparence de la discussion n'en est donc pas vraiment une car l'outil ne peut répondre à la question qu'on lui pose depuis le partage d'un monde commun ou d'une situation vécue ensemble, mais selon une base documentaire qui reste la même quelle que soit les utilisateurs et les questions posées. Si par conséquent, il répond aussi bien que les livres, l'outil n'en reste qu'une émanation, et ne produit pas une énonciation pour prendre part à un monde partagé.

39 L'énonciation ne relève pas tant d'un ego qui serait le principe transcendantal de l'expérience, ou d'une encyclopédie compilée, mais d'un engagement dans une situation ou un milieu qui nous englobe et nous contient, vis-à-vis duquel on se détermine et se positionne. Le milieu, ou contexte, c'est ce qui fait que nos données changent en permanence, qu'elles sont formatées et sélectionnées selon ces critères évoluant en permanence, et que finalement les règles de raisonnement se déterminent en fonction de l'horizon de notre engagement. Or, la machine a des données fixes, un modèle d'apprentissage intangible et pas d'horizon d'engagement.

40 En revanche, il y a bien un engagement de l'utilisateur qui adhère, rejette, reprend, réutilise les productions symboliques. Par l'interprétation qu'il en fait, par le statut qu'il leur confère, l'utilisateur sélectionne et instaure, il énonce et s'engage. Par conséquent, l'interaction avec de tels outils relève bien de l'énonciation, mais elle procède de l'utilisateur et de sa situation, dans laquelle il convoque un outil lui proposant des productions symboliques dont l'interprétabilité le plonge dans l'illusion transcendantale d'un monde dont ces productions seraient une représentation et dont il serait l'interprétation[7].

Conclusion

41 L'IA générative n'en finit pas de nous surprendre mais elle s'inscrit dans une tradition symbolique ancienne car elle relève de la grammatisation, au sens où le système technique de l'expression permet de la manipuler et de la transformer. Or, le fait de pouvoir travailler sur des caractères, des symboles, qui représentent de manière autonome et complète une expression (la complétude devant se comprendre au niveau du système phonologique des phonèmes - comme le fait remarquer (Lassègue & Longo, 2025), où ce qui est écrit permet de reproduire ce qui est dit sans pour autant toujours le comprendre, comme lorsque l'on lit machinalement un texte sans faire attention à ce que l'on lit), permet d'avoir des productions mécaniques et machinales d'écritures qui ne disent rien en propre mais qu'on lit et interprète pour leur faire dire quelque chose. La réception de ces productions produit une énonciation où l'on fait dire à ces productions ce qu'elles ne disent pas, car elles ne parlent pas du monde sur lequel on les interroge mais répliquent seulement l'ordre mécanique des symboles disposés dans l'espace latent documentaire. La puissance de ces outils repose donc bien sur un principe de modélisation et d'énonciation : mais au lieu de modéliser le monde, les IA se comportent comme un modèle des organisations documentaires soumises à l'apprentissage ; de même, les IA n'énoncent rien en propre parce qu'elles n'ont pas de monde où énoncer, mais en revanche leur usage est une énonciation établie par l'utilisateur qui interprète et sémantise les productions symboliques comme des contenus porteurs de sens, ce qu'ils n'ont pas besoin d'être pour être collectés, assimilés et transformés dans l'espace latent, leur mise en caractères ou symboles suffit.

42 Comme souvent, les débats sur les IA ne nous en apprennent pas beaucoup plus sur ce que sont ces outils, dont le principe de fonctionnement permet de trancher de manière claire à propos des questions posées (modélisation des données et non du monde produisant ces données, énonciation par l'utilisateur et non par la machine), mais plutôt sur nous qui nous posons ces questions. En tant qu'animaux symboliques, on ne peut s'empêcher de reconnaître dans les objets qui nous entourent des signes qui

s'adressent à nous, voyant en eux des énonciations à propos du monde dans lequel nous sommes plongés et qui par conséquent ne peut que nous échapper.

BIBLIOGRAPHIE

Alammar, J., & Grootendorst, M. (2024), *Hands-On Large Language Models: Language Understanding and Generation*, Sebastopol, CA : O'Reilly.

Bachimont, B. (2017), *Patrimoine et numérique : Technique et politique de la mémoire*, Bry sur marne : Ina-Éditions.

Bachimont, B. (2018), « Formats : transparence manipulatoire et opacité interprétative ; la question du sens dans les dispositifs techniques », in E. Mitropolou & N. Pignier (éds.), *Le Sens au cœur des dispositifs et des environnements* (pp. 189-221), Saint-Denis : Connaissances et savoirs.

Bachimont, B. (2025a), « Digital Ethics: Empowering Agents and Taking Care of Systems », in M.-H. Abel, N. Matta, H. Karray, & I. Saad (eds.), *Ethics and Digital Transition* (pp. 1-27), London : ISTE Wiley.

Bachimont, B. (2025b), « L'espace-temps du calcul : position virtuelle et distance réelle », *Signata, Annales des sémiotiques / Annals of Semiotics*, 16 (Sémiotique de l'espace à l'ère de la logique computationnelle et des pratiques numériques : une perspective contemporaine). Doi : https ://doi.org/10.4000/14rhm.

Berners-Lee, T., Hendler, J., & Lassila, O. (2001), « The Semantic Web », *Scientific American* (284), pp. 34-43.

Briet, S. (1951), *Qu'est-ce que la documentation ?*, Paris : Éditions Documentaires, Industrielles et Techniques.

Eco, U. (1985), *Lector in Fabula ; Le rôle du lecteur ou la coopération interprétative dans les textes narratifs*, Paris : Grasset & Fasquelle.

Eco, U. (1988), *Sémiotique et philosophie du langage*, Paris : Presses Universitaires de France.

Fodor, J. (1975), *The Language of Thought*, Harvard University Press.

Gastaldi, J.L. (2021), « Why Can Computers Understand Natural Language? », *Philosophy & Technology*, 34 (1), pp. 149-214. Doi : https://doi.org/10.1007/s13347-020-00393-9.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Bengio, Y. (2014), *Generative Adversarial Networks*, Paper presented at the Proceedings of the International Conference on Neural Information Processing Systems (NIPS 2014).

Harris, Z.S. (1954), « Distributional Structure », *WORD*, 10 (2-3), pp. 146-162. Doi : 10.1080/00437956.1954.11659520.

Haugeland, J. (1989), *L'Esprit dans la machine ; Fondements de l'intelligence artificielle*, Paris : Odile Jacob.

Heidegger, M. (1986 [1927]), *Être et Temps* (*Sein un Zeit*, Traduction F. Vézin), Paris : Gallimard (Bibliothèque de de Philosophie Série : œuvres de Martin Heidegger).

Herrenschmidt, C. (2007), *Les Trois Écritures : langue, nombre, code*, Paris : Gallimard (Bibliothèque des Sciences Humaines).

Humphreys, P. (2004), *Extending ourselves: Computational science, empiricism, and scientifique method*, New-York : Oxford University Press.

Humphreys, P., & Alvarado, R. (2017), « Big Data, Thick Mediation, and Representational Opacity », *New Literary History*, 48, pp. 729-749.

Kant, E. (1997 [1781]), *Critique de la raison pure* (*Kritik der reinen Vernunft*, Traduction A. Renaut), Paris : Aubier (Bibliothèque philosophique).

Lassègue, J. (2013), « Quelques remarques historiques et anthropologiques sur l'écriture informatique », in F. Nicolas & A. Tonneau (éds.), *Les Mutations de l'écriture* (pp. 83-103), Paris : Éditions de la Sorbonne.

Lassègue, J. (2023), « Machines "à mouvement" et machines "à signes" ; Jalons pour une histoire de l'alphabet », in V. Charolles & É. Lamy-Rested (éds.), *Les Philosophies des techniques, un levier pour l'action* (Vol. 7, pp. 63-72), Londres : ISTE Editions Ltd.

Lassègue, J., & Longo, G. (2025), *L'Empire numérique : de l'alphabet à l'IA*, Paris : Presses universitaires de France.

Mattioli, M., Cinà, A.E., & Pelillo, M. (2024), « Understanding XAI Through the Philosopher's Lens: A Historical Perspective », *arXiv preprint :* https://arxiv.org/abs/2407.18782.

Mc Culloch, W., & Pitts, W. (1943), « A logical calculus of the ideas immanent in nervous activity », *Bulletin of Mathematical Biophysics*, 5, pp. 115-133.

McCarthy, J., Minsky, M.L., Rochester, N., & Shannon, C.E. (1955), *A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence*, Retrieved from http://jmc.stanford.edu/articles/dartmouth/dartmouth.pdf.

Mikolov, T., Chen, K., Corrado, G.S., & Dean, J. (2013), « Efficient Estimation of Word Representations in Vector Space », *Proceedings of Workshop at ICLR*, 2013.

Minsky, M., & Papert, S. (1969), *Perceptrons: An Introduction to Computational Geometry*, Cambridge, Ma : MIT Press.

Panaccio, C. (1999), *Le Discours intérieur : de Platon à Guillaume d'Ockham*, Paris : Seuil (Des Travaux).

Paolucci, C. (2025), « The myth of meaning: generative AI as language-endowed machines and the machinic essence of the human being », *Semiotica*, 2025 (262), pp. 5-23. Doi : https://doi.org/10.1515/sem-2024-0204.

Pegny, M. (2024), *Éthique des algorithmes et de l'intelligence artificielle*, Paris : Vrin (Pour demain).

Pitrat, J. (1990), *Métaconnaissance. Futur de l'intelligence artificielle*, Paris : Hermés.

Platon (1999a), *Ménon* (Traduction M. Canto-Sperber), Paris Garnier-Flammarion.

Platon (1999b), *Théétète* (traduction et notes de Michel Narcy éd.), Paris : Garnier-Flammarion.

Platon (2006), *Phèdre, suivi de la Pharmacie de Platon par Jacques Derrida* (Traduction, introduction et notes de Luc Brisson éd.), Paris : Garnier-Flammarion.

Pylyshyn, Z.W. (1984), *Computation and Cognition, Toward a Foundation for Cognitive Science*, Cambridge MA : The MIT Press.

Rastier, F. (1991), *Sémantique et Recherches Cognitives*, Paris : Presses Universitaires de France.

Rosenblatt, F. (1958), « The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain », *Psychological Review*, 65 (6), pp. 386-408. Doi : https://doi.org/10.1037/h0042519.

Rumelhart, D.E., Hinton, G.E., & Williams, R.J. (1986), « Learning representations by back-propagating errors », *Nature*, 323 (6088), pp. 533-536. DOI : https://doi.org/10.1038/323533a0.

Shortliffe, E.H. (1976), *MYCIN: Computer-based Consultations in Medical Therapeutics,* American Elsevier.

Somaini, A. (2024), « Le visible et l'énonçable. L'IA et les nouveaux liens algorithmiques entre images et mots », *Nouvelle revue d'esthétique,* nº 33(1), pp. 47-58. Doi : 10.3917/nre.033.0047.

Stich, S.P. (1983), *From Folk Psychology to Cognitive Science, The Case Against Belief*, The MIT Press.

Varenne, F. (2012), *Théorie, réalité, modèle : épistémologie des théories et des modèles face au réalisme dans les sciences*, Paris : Éditions Matériologiques (Sciences & Philosophie).

Vaswania, A., Shazeer, N., Parmar, N., Uszkoreit, J., Llion, J., Gomez, A. N., Polosukhin, I. (2017), *Attention is all you need*, Paper presented at the NIPS'17: Proceedings of the 31st International Conference on Neural Information Processing Systems, Long Beach, CA.

Wiener, N. (1961), *Cybernetics, or Control and Communications in the Animal and the Machine* (2nd ed.), Cambridge, MA : The MIT Press.

NOTES

1. Le calcul comme écriture n'implique pas qu'une culture ayant dégagé des méthodes de calcul possèdent d'emblée une écriture renvoyant à la parole et au langage. Cependant, la thèse ici, que nous partageons et tirons des travaux cités de Jean Lassègue, mais aussi de ceux de Clarisse Herrenschmidt (Herrenschmidt, 2007), est que le calcul et l'écriture renvoient à une matrice commune de symbolisation et manipulation, qu'on peut aussi appeler grammatisation (Bachimont 2019).

2. Cette conception de la sémantique et du sens peut être discutée dans la mesure où la clôture sur le corpus coupe l'interprétation du contexte dans lequel elle s'effectue et celui dans lequel les contenus analysés se sont constitués. Si le corpus est un instrument d'observation de la sémantique, il n'est pas la langue et il n'est ni nécessaire ni suffisant pour dégager le sens des énoncés (voir par exemple (Rastier, 1991)).

3. Cette réflexivité ainsi comprise a souvent été caractérisée comme ce qui manquait à l'IA pour posséder ou simuler les facultés cognitives dont on voulait la doter. L'IA symbolique a notamment donné lieu à un programme de recherche sur les métaconnaissances (Pitrat, 1990).

4. Ce pourrait être la preuve que, puisque l'écriture comme agencement de symboles peut se passer du sens, c'est que ce dernier n'existe pas et constitue un vestige mythique. Seul compterait la matérialité symbolique comme explication et explicitation des échanges communicationnels et productions de connus. Ces idées furent notamment soutenues par la psychologie et les sciences cognitives computationnelles (Pylyshyn, 1984 ; Stich, 1983). Des raisonnements analogues sont soutenus parfois pour les modèles d'apprentissage profond mais la discussion est plus compliquée car il faut comprendre quel statut conférer aux modèles mathématiques utilisés par ces modèles. Les sciences computationnelles en restent à une fondation symbolique, calculatoire et inférentielle à l'instar des systèmes formels.

5. Cité également dans (Lassègue & Longo, 2025).

6. Voir également son intervention, très stimulante, au séminaire international de sémiotique de Paris, « IA et cultures visuelles : une théorie des espaces latents », le 25 janvier 2025, https://www.youtube.com/watch?v=im0auFKwx74.
7. Mais si l'énonciation relève de l'engagement et de l'être-jeté (au sens de cette notion dans (Heidegger, 1986/1927)) dans une situation ou un contexte, dans un monde, elle renvoie également à des considérations éthiques qu'on ne peut aborder ici mais qui appellent à l'être (Bachimont 2025a).

RÉSUMÉS

Dans la recherche d'outils permettant de réaliser des tâches des ressources cognitives quand elles sont effectuées par les êtres humains, l'IA, et en particulier l'IA générative, semble atteindre un niveau de performance inédit. Ces performances relancent les questions sur la nature de ces outils et des productions qui en découlent : ont-elles du sens, sont-elles des énonciations d'un discours, proposent-elles des vérités sur le monde auxquelles elles semblent faire référence ? Cet article revient sur le principe de fonctionnement des IA génératives, en particulier celles fondées sur les larges modèles de langage (LLM) pour aborder ces questions et proposer des réponses. Les productions de l'IA comme séries de symboles agencés sont analysées selon trois niveaux : le dire, le dit et le désigné. Si l'IA nous trouble, c'est que nous analysons ses productions comme des dits, donc émanant d'un dit et renvoyant à un monde désigné. Il y aurait alors une énonciation (un dire) et une modélisation (représentation d'un désigné). L'article argumente qu'en fait de modélisation du monde, c'est plutôt une modélisation du déjà-dit qui est effectuée, notamment grâce à l'espace latent de représentation, qui fonctionne comme une méta-archive, et qu'en fait d'énonciation, il n'y a qu'un agencement symbolique. La machine n'ayant ni contexte, ni situation effective dans lesquels ces productions sont faites, il n'y a ni en effet monde à représenter ni langage à proférer si on se rappelle que ce dernier est d'abord un dialogue situé dans un monde. Ce qui fonde et explique l'efficacité constatée maintes fois de ces approches, mais en indique d'emblée la portée et les limites.

In the search for tools capable of performing tasks that require cognitive resources when carried out by humans, AI—and generative AI in particular—appears to be reaching unprecedented levels of performance. This performance raises questions about the nature of these tools and the outputs they produce: do they have meaning, are they utterances of a discourse, do they propose truths about the world to which they seem to refer? This article revisits the operating principles of generative AI, particularly those based on large language models (LLMs), to address these questions and propose answers. AI outputs, as series of arranged symbols, are analyzed at three levels: the saying, the said and the signified. If AI unsettles us, it is because we analyze its outputs as 'said' statements, thus emanating from a 'saying' and referring to a 'signified' world. There would then be an enunciation (a saying) and a modelling (representation of a content). The article argues that, in terms of modelling the world, what is actually taking place is a modelling of what has already been said, notably thanks to the latent space of representation which functions as a meta-archive, and that, in terms of enunciation, there is merely a symbolic arrangement. Since the machine has neither context nor an actual situation in which these productions are made, there is in fact neither a world to represent nor a language to utter, if we recall that the latter is first and foremost a dialogue situated within a world. This underpins and

explains the frequently observed effectiveness of these approaches, but also immediately indicates their scope and limitations.

INDEX

Mots-clés : IA générative, espace latent, modélisation, énonciation, situation
Keywords : generative AI, latent space, modelling, enunciation, situation

AUTEUR

BRUNO BACHIMONT

Bruno Bachimont est professeur d'informatique et d'épistémologie à l'université de technologie de Compiègne. Il a été directeur de la recherche et conseiller scientifique de l'Institut national de l'audiovisuel (INA). Il s'intéresse aux questions de la connaissance, de la mémoire, et des archives dans le contexte de la numérisation. Parmi ses publications récentes sur les questions abordées dans l'article, voir : « AI and algorithms: what digital technology can teach us about our content », in Pietro Conte, Anna Caterina Dalmasso, Maria Giulia Dondero, & A. Pinotti (eds.), *Algomedia. The Image at the Time of Artificial Intelligence* (pp. 17-36), Cham, Switzerland : Springer, 2026 ; « Le récit comme connaissance résiliente : aborder la complexité en situation », *Cahiers Risques et Résilience, 7* (Résilience de l'ingénieur), pp. 37-56, 2025 ; « Digital Ethics: Empowering Agents and Taking Care of Systems », in M.-H. Abel, N. Matta, H. Karray, & I. Saad (eds.), *Ethics and Digital Transition* (pp. 1-27), London : ISTE Wiley, 2025 ; « L'espace-temps du calcul : position virtuelle et distance réelle, *Signata 16* (2025), https ://doi.org/10.4000/14rhm. Parmi ses ouvrages personnels, voir : *Patrimoine et numérique : technique et politique de la mémoire*, Bry sur marne : Ina-Éditions, 2017 ; *Le Sens de la technique : le numérique et le calcul*, Paris : Encres Marines/ Les Belles Lettres, 2010.