Pierluigi Basso Fossali*

Rationalités correctives et intelligence artificielle assistée : les doubles contraintes des humanités numériques

https://doi.org/10.1515/sem-2024-0189
Received October 30, 2024; accepted December 26, 2024; published online February 5, 2025

Résumé: Cet article explore la complexité de l'interaction entre l'Intelligence Artificielle (IA) et les rationalités humaines, en mettant l'accent sur les défis des humanités numériques. À travers une étude exploratoire, on articule deux perspectives : d'une part, la critique de l'idée que l'IA pourrait imiter ou reproduire fidèlement les rationalités humaines, et d'autre part, l'examen des limites et des possibilités de l'intelligence artificielle en tant que système autonome. L'analyse s'appuie sur des interactions prolongées avec ChatGPT, visant à tester les capacités de l'IA dans des contextes pédagogiques et épistémiques. L'article propose une reconceptualisation des relations entre jeux, calculs et mondes possibles, en soulignant les implications pour des sémiosphères différentes (culturelle/artificielle) et les standards de rationalité. Finalement, l'article vise à problématiser les rôles potentiels de l'IA en tant que catalyseur d'énonciation, médiateur dans une cognition distribuée, et agent participant à une intelligence collective, tout en soulignant les limites intrinsèques des capacités réflexives et cognitives de l'IA.

Mots-clés: Intelligence artificielle; formes d'intelligence; catalyse; textualisation; énonciation machinique

Abstract: This article addresses the complexity of the interaction between Artificial Intelligence (AI) and human rationalities, with a focus on the challenges posed by digital humanities. Through an exploratory study, two perspectives are articulated: on one hand, the critique of the idea that AI could faithfully imitate or replicate human rationalities, and on the other hand, an examination of the limitations and possibilities of artificial intelligence as an autonomous system. The analysis draws on extended interactions with ChatGPT, aiming to test AI's capabilities within pedagogical and epistemic contexts. The article proposes a reconceptualization of the relationships between games, calculations, and possible worlds, highlighting the

***Corresponding author: Pierluigi Basso Fossali**, Université Lyon 2, Lyon, France; and University of Bologna, Bologna, Italy, E-mail: pierluigi.basso@univ-lyon2.fr. https://orcid.org/0000-0001-8460-2049

implications for different semiospheres (cultural/artificial) and standards of rationality. Finally, the author discusses the potential roles of AI as a catalyst for enunciation, a mediator in distributed cognition, and an agent participating in collective intelligence, while underscoring the intrinsic limitations of AI's reflexive and cognitive capacities.

Keywords: artificial intelligence; forms of intelligence; catalysis; textualization; machinic enunciation

> Une conversation 'artificielle' est suffisamment étendue, et potentiellement heuristique par rapport aux rôles actantiels qui sont virtuellement en jeu, lorsque la tâche de poser la bonne question ne semble plus aussi ridicule et stérile.

1 Introduction

À travers une étude exploratoire, cette contribution[1] vise à articuler deux perspectives : la première concerne la recherche des racines et des types d'intelligence qui seraient impliqués dans l'IA ;[2] la seconde relève de la qualité et du taux d'artificialité que l'on peut attribuer à l'IA. Cette articulation vise à combiner, d'une part, le scepticisme vis-à-vis d'une intelligence de la machine qui estime être mimétique des rationalités humaines et, d'autre part, le constat de l'impossibilité, pour une intelligence artificielle, de se débarrasser totalement de certains piliers de la mise en cohérence des connaissances, piliers qui sont typiquement liés à la sémiotisation d'un monde de référence et à la gestion de *mondes possibles*.

1 Cet article reprend ce que nous avons présenté lors de notre intervention du 27 mars 2024 au Séminaire international de sémiotique à Paris.

2 Notre contribution n'a pas l'ambition "de proposer une sorte de révolution copernicienne dans les études sur ce qu'on appelle communément 'intelligence artificielle'" (Vitali-Rosati 2024), mais se donne, à travers un parcours indépendant, des objectifs similaires, du moins en reconnaissant que le problème fondamental est de comprendre quelles formes d'intelligence sont impliquées dans l'IA et dans notre relation avec elle. En ce sens, on peut affirmer que ChatGPT "représente l'implémentation formelle d'une définition particulière d'intelligence" (Vitali-Rosati 2024 : 16). Ensuite, cela ne peut qu'ouvrir un chantier pour reconnaître *les* formes d'intelligence développées, impliquées, émergentes. Une approche phénoménologique ne peut se contenter de supposer une idée de l'intelligence et de l'appliquer une fois formalisée. De plus, sur un plan épistémologique plus général, la science s'offre à la réfutation de ses propres modèles, et il ne lui suffit pas de les mettre en œuvre et d'en constater le bon fonctionnement, ce qui la différencie de la technologie.

Notre approche profite de l'extension des principes de *résistance* afin de faire émerger, dans le maintien d'une perspective comparatiste, des instances mutuellement irréductibles. En ce sens, même l'interface utilisateur et la générativité de textes les plus transparentes par rapport aux pratiques humaines peuvent faire émerger une artificialité 'résistante'. Une 'diacritique' qui polarise la problématisation conjointe de créativité et algorithmique peut à la fois profiter de l'IA pour repenser les limites et les contradictions de la première et pour comprendre davantage l'émergence de marques spécifiques d'une agentivité machinique. On peut paraphraser cette assertion programmatique en disant que cette 'diacritique' a la tâche de faire exploser un *double bind,* ou en tout cas une double mission contradictoire : (i) se substituer à l'intelligence humaine à travers un programme mimétique et donc selon un affinement progressif des formes de reproduction de gestes rationnels ; (ii) se détacher des limites des rationalités anthropiques afin de développer une cognition artificielle émancipée de l'écologie sémiotique des acteurs sociaux.

Le développement de cette diacritique peut suggérer une reconceptualisation des relations entre jeux et calculs, niches écologiques et mondes (logiquement) possibles, avec des retombées sur l'évaluation même des agentivités et de la consistance sémantique de leurs interventions locales. Ce que nous attendons de l'intelligence artificielle se répercute inévitablement sur ce que nous attendons de l'exercice de nos facultés. Ainsi, le cadre même des compétences, des capacités et des potentiels évolue face aux modèles analogisants de l'intelligence humaine et aux formes d'apprentissage cognitivement impénétrables. En ce sens, il est bien de se poser la question dans un cadre synchronique : quelles sont actuellement les formes d'intelligence que l'on peut revendiquer comme spécifiques de l'humain et quelles formes d'intelligence peut-on imputer à l'IA générative? Ce cadre ne pourra qu'évoluer et offrir d'autres formes de méta-représentation des intelligences, avec des attentes inévitablement corrélées.

Notre contribution aimerait explorer, même dans une perspective pédagogique, les standards de rationalité négociables, des standards qui sont en partie descriptibles (intelligences mimétiques) et en partie totalement liés à l'émergence de résultats (effet 'boîte noire'), donc purement sanctionnables.

En ce qui concerne le premier régime (descriptif), on peut penser à la constitution de compétences pour une 'intelligence corrective'. Les rationalités peuvent également être améliorées, orientées, etc. et pas seulement les ontologies.[3] Dans le

3 Cf. Strawson (1959). On peut rappeler aussi Simon (1996 [1969] : 157) qui a soutenu que nous sommes des "créatures à rationalité limitée." Notamment, "en appliquant un facteur d'actualisation élevé aux événements, en les atténuant par leur éloignement dans le temps et dans l'espace, nous réduisons nos problèmes de choix à une taille correspondant à nos capacités de calcul limitées" (nous traduisons).

deuxième régime, la pédagogie peut s'intéresser à une intelligence 'maïeutique' à même d'interroger la machine pour faire émerger non pas une vérité, mais au moins une 'justesse' des réponses par rapport à la tâche, tâche qui de toute façon concerne aussi les sources (les données de départ) et la finesse des résultats visés. En ce sens, on peut parler d'*intelligence artificielle assistée*, ce qui renverse la perspective passive et oraculaire (interroger l'IA). À ce propos, deux perspectives maïeutiques pertinentes s'imposent : s'interroger de manière critique sur les raisons pour lesquelles certains résultats sont obtenus par l'IA et pas d'autres (*rétroduction*), ou se demander comment ils s'intègrent dans des cibles encore en cours de définition (*abduction*).

2 Quels types d'organisation sémiotique sont impliqués dans l'interrogation d'une IA?

2.1 Critère pour accueillir une littérature artificielle

Au vu de nos compétences assez limitées concernant les modèles techniques de l'Intelligence Artificielle, notre point de départ ne peut qu'être à la fois modeste et bien ancré dans les pratiques : en ce sens, nous avons choisi d'utiliser des corpus d'interactions avec ChatGPT 4,[4] constitués entre novembre 2023 et mars 2024. Une petite mise à jour qui exploite la version 4o[5] sera intégrée à la fin.

Cette constitution de plusieurs corpus a été faite aussi pour des raisons pédagogiques, aptes à établir des critères pour évaluer les devoirs des étudiants qui utilisent de manière habituelle l'IA. À la place d'interdire l'utilisation de l'IA, sans avoir aucune possibilité de contrôler le respect de cette prohibition, au moins avant la possible introduction de codes en filigrane,[6] il est opportun de négocier des critères pour intégrer des productions discursives 'artificielles'. Parmi les critères que nous proposons dans nos cours et qui sont appliqués ici, nous voulons souligner :

4 Le "Generative Pre-trained Transformer 4" est un modèle d'interaction multimodale structuré à partir d'un type *transformeur génératif pré-entraîné* et développé par la société OpenAI. Disponible à partir du 14 mars 2023, il a été utilisé pour constituer nos corpus à partir du 29 novembre 2023.

5 On utilisera cette abréviation pour indiquer "GPT-4 omni" qui est disponible au public payant à partir du 13 mai 2024.

6 C'est la préconisation de plusieurs acteurs institutionnels, mais l'obligation d'insérer des codes en filigrane (*watermarks*) et de prévoir leur interopérabilité peut apparaître comme un objectif assez complexe à atteindre sur le plan technique et surtout politique, au moins à l'échelle mondiale.

(i) l'exigence de ne pas consulter de manière ponctuelle et presque 'oraculaire' un chatbot[7] afin de produire, au contraire, des corpus d'interactions[8] avec la machine basés sur des questions corrélées et pour un temps suffisant afin d'accomplir un test de résistance sur la stabilité et la fiabilité des réponses obtenues ;
(ii) l'insertion de passages explicites des interactions avec la machine dans les devoirs, en suivant les critères habituels, même graphiques, pour intégrer des citations dans un texte original ; éventuellement on peut ajouter des parties plus étendues du corpus en annexe.
(iii) le devoir des étudiants doit présenter des enchâssements énonciatifs critiques des passages cités en montrant une réappropriation herméneutique avisée[9] de la production discursive 'artificielle'.

Il nous semble que le respect de ces trois critères peut assurer une mise en perspective de la littérature artificielle, sans la diaboliser et sans non plus la traiter comme du texte : au fond, les passages cités sont assumés comme des fragments documentaires d'une production artificielle à l'intérieur d'un processus de *textualisation*. Ce dernier terme a gardé une certaine ambiguïté dans le métalangage sémiotique[10] et pourrait aujourd'hui assumer le rôle d'éclaircir le passage entre un acte de production sémiotique et sa prise en charge praxéologique, c'est-à-dire par quelqu'un, dans un espace-temps donné, avec une intentionnalité inscrite dans un domaine de valeurs spécifique. La *textualisation* établirait les conditions discursives

7 Nous utilisons ici le terme *chatbot*, qui est un logiciel pour la simulation d'interactions H/M, comme une métonymie de l'ensemble de solutions d'IA qui exploitent des *modèles de traitement du langage naturel* (NLP), pour gérer l'utilisation du langage humain, et les *grands modèles de langage* (LLM), pour préfigurer des 'bonnes' continuations de séquences discursives à partir d'un entrainement sur des grandes masses des données.

8 Le terme *interaction* est ici employé pour désigner spécifiquement les échanges entre l'utilisateur humain et l'IA, tandis qu'*énonciation* fait référence à la manière dont ces échanges sont structurés et interprétés dans un contexte culturel donné.

9 On dispose déjà de bon nombre d'études sur l'impact de l'IA dans le milieu pédagogique et dans le cadre de l'autoformation aussi. Il faudrait caractériser ChatGPT, qui "excelle uniquement en littérature avec un taux de précision de 68 %" (Dam et al. 2023), par rapport à d'autres *chatbots* basés sur des LLM (Large Language Models). Pour des raisons d'espace, nous renvoyons à l'étude mentionnée ci-dessus qui offre des réflexions opportunes et complémentaires.

10 D'abord, la sémiotique structurale a conçu la *textualisation* comme l'ensemble des procédures visant à constituer un continuum discursif, préalablement à la manifestation du discours dans telle ou telle sémiotique, ce qui réduit le texte à une représentation sémantique abstraite (Greimas et Courtés 1979 : entrée *textualisation*). Par la suite, la textualisation a été interprétée comme la rencontre d'une programmation discursive, déjà dotée d'une déclinaison figurative, avec la matière expressive propre à une pratique énonciative, selon des accommodements plus ou moins féconds ou en tout cas exploités dans le temps selon des genres (Greimas et Fontanille 1991 : 308).

immanentes pour l'articulation successive avec un cadre pragmatique qui établira la force illocutoire et les effets perlocutoires d'une véritable énonciation.

Par ailleurs, la *textualisation* de la littérature artificielle, avec ces compositions sémiotiques encore ouvertes à la sémiose, attend l'aboutissement d'un circuit de sens dans lequel les réponses obtenues via le *chatbot* ne seront que les reflets de nos questions,[11] et ce d'autant plus que les questions auront été posées de manière systématique, avec des modulations de perspectives et de requêtes de précisions, de paraphrases, de traductions ou de changements de style.

À partir de ce cadre praxéologique, voué à la formation des étudiants universitaires, nous avons décidé d'appliquer ces critères jusqu'au point limite de leur exploitation performative, au moins selon nos capacités et notre patience. En ce sens, nous avons entretenu ChatGPT-4 pour des heures sur le même sujet, selon un dialogisme expérimental apte à tester une phénoménologie de la responsivité artificielle. Cette démarche phénoménologique nous délivre une série de résultats qui seront présentés avec une série de précautions interprétatives et parfois sous la forme d'interrogations, étant donné que, malgré nos efforts, les données ne sont pas suffisantes et les explications de certains épiphénomènes locaux susceptibles de plusieurs interprétations, y compris strictement techniques ou statistiques (par ex. des erreurs qui peuvent survenir sans aucune relation avec l'activité de l'usager). Le lecteur pourra juger de la possibilité d'assumer les données présentées comme une base probante ou tout simplement comme un terrain indiciaire encore ouvert à la récolte de données.

2.2 Une sémiosphère augmentée

Nous avons commencé notre recherche avec la constitution de six corpus liés à ces thématiques :

(i) Perception limitée du monde

11 C'est pour cette raison que nous avons précédemment consacré un essai entier à l'*énonciation machinique* (Basso Fossali 2008). Dans une sémiosphère culturelle, toute instance qui entre dans un réseau de relations avec des initiatives énonciatives tend à être assumée régulativement comme une énonciation, et ce y compris dans le cas où les dispositifs sous-jacents ne sont pas en mesure d'assumer la portée de sens de leurs fonctions. Cependant, pour que cette énonciation machinique ne reste pas de second ordre et soit finalement assumée dans un cadre de signification supérieur, attribuable à une instance anthropique, elle doit se montrer dissidente, voire récalcitrante à toute intégration. C'est pourquoi la phénoménologie de l'énonciation machinique que nous avons étudiée par le passé s'est présentée sous des formes négatives, d'interruption des réseaux de sens d'une sémiosphère. L'énonciation machinique se manifeste alors sous la forme d'une boucle infonctionnelle, d'une erreur récurrente, d'un grippage, d'un bruit de fond. Pour une étude de l'*énonciation machinique* dans le contexte de l'IA, voir Dondero (2024a, 2024b).

(ii) Sentience/conscience
(iii) Théories des mondes possibles et fiction[12]
(iv) Formes d'intelligence
(v) Voix dans ma tête
(vi) Simulations d'énonciation

Les thématiques ont été choisies au vu d'une portée épistémologique. D'une part, elles ont été dictées par la curiosité de voir les comportements de la machine face à des questions qui suscitent, au moins dans la littérature traditionnelle, l'autoréflexivité et des assomptions philosophiques marquées. D'autre part, notre maîtrise relative de la littérature sur ces questions était un arrière-plan adapté pour essayer de comprendre quelles sont les sources utilisées pour régénérer des documents. En effet, il est évident que les réponses assurées par la machine relèvent de textes prétraités et 'digérés', dont il convient de tester les relations avec la langue utilisée et avec le métalangage éventuellement mobilisé pour la question posée.

Il faut préciser aussi que nous avons initialement assumé l'idée qu'un chatbot ne forge pas des représentations psychologiques, donc qu'il n'a pas de conscience. En ce sens, les réponses qu'il offre ne sont qu'une simulation d'interaction dont les textualisations sont des bricolages de formes attestées dans les grandes bases de données archivées. On pourrait souligner que comme toute hypothèse de départ, elle mériterait bien des tests de réfutation. Mais cela dépasse largement les ambitions de notre contribution. Éventuellement, les tests que nous avons accomplis ont la tâche d'ouvrir des fissures, des questionnements, des hésitations même sur notre manière d'approcher l'IA. En effet, dans la relation avec la machine, on voit le reflet de nos conceptions de l'intelligence, de la rationalité humaine, des compétences nécessaires pour gérer le sens à travers la machine. C'est pourquoi, pendant la lecture de cet article, il faudra bien distinguer, d'une part, le 'je' qui interroge la machine et qui a une démarche exploratoire et décomplexée, donc libre de s'interroger sans devoir respecter un cadre épistémologique cohérent ; d'autre part, le 'nous' de majesté qui encadre le dialogue expérimental et qui maintient un regard sceptique sur l'expérience relationnelle avec la machine.

De manière corrélée, ce scepticisme est valable aussi pour ce qui concerne l'idée d'une générativité du sens assumée par un seul sujet qui l'opérationnalise, un sujet qui voit les valeurs se déployer devant lui comme un réseau de saillances. La sémiotique n'est pas une science des représentations ou des présuppositions logiques ; c'est une science des médiations. C'est pourquoi, même si l'on pratique une *épochè* du sens, si l'on veut suspendre les remplissements du sens déjà activés par l'activité

12 Ce corpus sur la théorie des mondes possibles est tellement développé et intrigant qu'il sera utilisé pour un autre article. Sa convocation dans cette contribution ne sera donc qu'indirecte.

perceptive, on est obligé à reconnaître, d'une part, une scène théorique peuplée de plusieurs instances et, d'autre part, le fait que chaque promotion de valeurs passe par des implications modales distribuées. Les médiations ne sont pas secondes, instrumentales, de simples ressources ; elles participent à la composition d'une scène signifiante qui voit émerger des valeurs entre des instances enfin reliées. C'est pour cette raison que, dans une écologie sémiotique, le sens n'est jamais autochtone et il passe par des circuits de médiation qui nous impliquent.[13] Dans cette perspective, on peut imaginer d'accueillir les Intelligences Artificielles comme des épicentres de reproduction énonciative dans une écologie sémiotique dans laquelle on pourra compter alors sur des médiations ultérieures : une sémiosphère augmentée.[14]

Au vu des arguments qui seront proposés par la suite, ce dernier point mérite une petite digression sur l'organisation systémique en boucles qui caractérise les cultures. Une culture ne peut que tenter de s'échapper d'une sémiosphère censée de la contenir : d'une part, elle échappe à cette réduction au vu des contacts contingents avec d'autres cultures (en effet, elle reste un *système ouvert* et basé sur la *traduction*) ; d'autre part, elle peut profiter de plusieurs circuits de signification chacun participant à des processus de *différenciation* (l'hétérarchie des *domaines* de signification) et de *complexification* interne (l'enchâssement des organisations sémiotiques, chacune demandant aux autres de porter une remédiation à son propre inachèvement interne).

Nous pouvons estimer que chaque culture nécessite :

(i) une *sémiotique de premier ordre* pour assurer des sémiosis convergentes autour de manifestations expressives qui ne profitent plus de contenus immédiatement corrélés à leur individuation sensible : c'est une perception déjà orientée à gérer des illusions, soit pour les neutraliser (*désenchantement*), soit pour en profiter comme dimension fictive (*simulation*) ;

(ii) une *sémiotique de deuxième ordre* permet de l'autoréflexivité, donc des formes de repliements *épilinguistiques*, et donc la *ré-entrée* (*re-entry*)[15] sur les valeurs

13 Pour une épistémologie de l'écologie sémiotique, voir Basso Fossali (2017 : 310–315).

14 Par *sémiosphère augmentée*, nous entendons une extension des capacités sémiotiques dans laquelle les interactions avec l'intelligence artificielle génèrent de nouveaux terrains de signification. Ce concept se distingue de la sémiosphère traditionnelle par sa capacité à intégrer des processus machiniques de textualisation et de médiation, augmentant ainsi les possibilités de gestion du sens. La sémiotique augmentée est expérimentée par l'usager, tandis que nous utiliserons l'expression *sémiosphère artificielle* pour thématiser l'environnement qui peut co-naître en concomitance avec l'émergence d'une intelligence artificielle.

15 Avec le terme de *re-entry*, élaboré initialement par George Spencer Brown, Luhmann (1984) indique la "rentrée" d'une distinction dans le domaine de valeurs qu'elle a contribué à délimiter. L'énonciation peut bien décrire la coalescence entre distinction opérée (plan de l'énonciation et plan de l'énoncé) et distinction observante qui cherche à tirer parti de ce clivage pour interpréter sa distinctivité (rétroaction symbolique).

instituées. À travers ces circuits autoréflexifs, la différentiation des valeurs finit pour concerner, de l'intérieur, les systèmes conventionnels et in fine la culture dans son ensemble ;

(iii) maintenant, peut-on envisager une *sémiotique de troisième ordre* dans laquelle des dédoublements de sémiosphères culturelles (par exemple, des archives réorganisées autour des intelligences artificielles) permettent de décoïncider[16] aussi par rapport aux circuits d'autoréflexivité? Si l'autoréflexivité de la sémiotique de deuxième ordre n'a pas conduit à une relativisation des prégnances de la première, pouvons-nous imaginer que la troisième n'assumera pas non plus le rôle de déconstruire les horizons d'expérience et l'ancrage de toute écologie des valeurs?

On voit bien que l'enchâssement entre des ordres sémiotiques permet de considérer des formes d'intelligence et des passerelles entre l'une et l'autre, sans privilège d'importance (*porter à* n'est pas inférieur à *porter sur*).

2.3 Formes d'intelligence

> Au départ, j'avais la tentation de dialoguer avec ChatGPT pour pousser la machine à avouer quelque chose. Je me suis aperçu progressivement qu'elle fonctionnait plutôt comme une médiation de ce que je projetais en termes de niveaux et des formes d'intelligences prétendues, de paramètres attendus pour reconnaître une conscience. Je ne peux pas enlever l'ambiguïté de ma démarche initiale ; mais si l'on veut dramatiser la question, on pourrait dire que j'essayais de faire émerger limites et potentiels de la machine tout en constituant, en parallèle et en négatif, les prétentions disproportionnées des langages et des rationalités humaines.[17]

D'une part, nous pouvons imaginer l'humain comme parfaitement achevé dans ses propres formes de vie ; ses limites sont tout simplement ses proportions et l'humain reste le paramètre fondamental de ce qui est significatif. Par rapport à cette vision humaniste, la numérisation ne peut qu'apparaître comme aliénante et réductrice nous restituant une resémiotisation de la culture de nature *monoplane*[18]

16 Pour le concept de *décoïncidence*, voir Jullien (2017).

17 Toutes les citations des corpus constitués seront reportées en italique, de manière à les distinguer des autres sources. Nous n'avons pas apporté des modifications aux dialogues écrits, et donc il y a des imprécisions et des erreurs, mais significativement du côté de la machine aussi !

18 Les sémiotiques *monoplanes* n'ont qu'une seule organisation qui organise à la fois les expressions et les contenus selon une parfaite conformité de structuration (Greimas et Courtés 1979).

(on manipule des fragments de documents avec une indifférence totale pour les articulations sémiosiques locales) et une sémantique réduite à des calculs probabilistes.[19] Dans ce cas, nous n'avons que : (i) des simulations de formes d'intelligence ; (ii) des succédanés des textes ; (iii) des impressions de dialogues.

Dans cette perspective, les interactions avec la machine ne sont que le reflet à la fois paresseux et ultrarapide d'un autodialogisme de la culture qui la pousse à se régénérer en violation *et* de sa tension herméneutique vers l'altérité *et* de sa vocation réflexive à l'émancipation. La 'décoïncidence' qui serait propre au devenir de la culture, une fois passée par l'artificialisation du dialogisme, pourrait retomber dans une simple exploration paradigmatique interne, la syntaxe étant réduite à l'exhaustion progressive du *possible*.

Si l'émancipation d'un *possible* déjà encadré par un système de pièces et d'opérations limitées n'est plus envisageable, alors on peut douter de rencontrer de véritables simulacres d'*intelligence*, cette dernière n'étant pas réductible au calcul. On sait que l'étymologie du terme 'intelligence' est problématique. Sans vouloir reconstruire l'énorme débat sur cette question, on peut se limiter ici à présenter deux interprétations majeures : la première considère l'étymologie *inter- ("entre")* et *lĕgĕre ("lire, connecter, choisir")*, la seconde préfère considérer le préfixe *intus*, ce qui donne lieu à l'idée de 'lire à l'intérieur', 'connecter en profondeur', à savoir une discrimination de liens internes selon des principes herméneutiques ou heuristiques, éventuellement plus généraux et qui n'ont pas nécessairement participé à la constitution de l'unité analysée. Au-delà des retombées épistémologiques différentes de ces deux généalogies conceptuelles possibles, ce qui émerge est une relation intime de l'intelligence avec l'hétérogénéité, soit pour la fédérer (*inter-lĕgĕre*), soit pour l'utiliser comme ressource analytique face à l'apparente uniformité. Si l'intelligence profite des calculs, elle ne se limite pas à ces derniers car elle gère plusieurs plans de rationalité, ce qui montre une homologie avec les plans de la signification. Il n'est donc pas abusif de parler d'*intelligence sémiotique* et de penser, compte tenu de cette nature, à une diffraction d''intelligences' selon les stratégies utilisées pour articuler des plans différents.

19 En France, le Comité national pilote d'éthique du numérique s'exprime à ce sujet sans aucune nuance : "Les systèmes d'IA générative fonctionnent uniquement avec des représentations numériques, sans appréhender la signification des mots pour les êtres humains. La signification est uniquement celle que les humains projettent sur les résultats, car seuls les humains en possèdent une interprétation dans le monde réel" (Grinbaum et al. 2023 : 11). Le problème est que le même Comité souligne que les IA développent des "capacités linguistiques et contextuelles de manière non intentionnelle" et que "la principale incertitude liée aux comportements émergents est la difficulté de les prédire" (Grinbaum et al. 2023 : 17).

À partir de ces suggestions étymologiques on peut donc estimer que la qualification d'un *chatbot* comme 'intelligence artificielle' est abusive.[20] Mais rien n'empêche d'imaginer que la sémiosphère augmentée qui est catalysée par nos relations avec les chatbots ne donne l'opportunité de tester notre intelligence ou de faire émerger de nouvelles formes d'intelligence. Comme nous l'avons vu, nous ne pouvons pas exclure que les dialogues avec des *chatbots* nous offrent une remédiation de troisième ordre de la culture, une forme de *re-entrée* dans nos médiations, un circuit ultérieur et concurrentiel qui poussent (ou nous poussera tôt ou tard) à requalifier les autres circuits, à savoir les formes de subjectivité, d'interactions *in vivo*, d'institutionnalisations du sens.[21]

Entre deux thèses totalement opposées, l'une qui réduit l'interrogation d'un *chatbot* à une activité protosémiotique de simple reconfiguration textuelle de matériels culturels archivés, l'autre qui attribue à l'intelligence artificielle une capacité représentationnelle des connaissances à même de dialoguer véritablement avec une conscience humaine, on peut essayer de construire un espace critique à la fois plus équilibré et modeste, qui envisage une problématisation à la recherche d'une heuristique proportionnée à nos pratiques. Cela dit, il faut reconnaître bon gré mal gré qu'indirectement on est en train d'évaluer, d'une part, une desémantisation des responsabilités énonciatives qui a comme effet probable une réification de la culture et un contrôle ultérieur du matériel sémiologique de la part des pouvoirs ; d'autre part, une resémantisation aventureuse de la culture qui risque d'attribuer un rôle prééminent à l'explorateur-exploitateur des patrimoines informationnels, et ce au détriment de toute reconnaissance préalable des institutions collectives qui devraient délimiter et caractériser en amont les disponibilités modales des usagers. Pour résumer, nous avons d'une part, une interprétation *informationnelle* de l'IA qui peut être soumise aux critiques sur la concentration et sur le contrôle des données ; d'autre part, une interprétation *prothétique* qui risque de construire un usager tout-puissant et replié sur sa version d'un monde 'augmenté'. Il est bien alors de préciser que notre position, bien qu'inspirée par une certaine circonspection face à des généralisations, ne peut éviter d'entrer dans une évaluation de ces retombées politiques de la conception de l'IA, s'opposant à l'interprétation *informationnelle* aussi bien qu'à l'interprétation *prothétique*.

20 Voir l'intervention influente de Evgeny Morozov (2023) publiée sur *The Guardian*. Au contraire, dans la littérature scientifique, on trouve des assertions qui ne semblent plus laisser de doutes sur l'utilisation non métaphorique du terme "intelligence" : "force est de constater qu'en 2021, l'intelligence artificielle est largement reconnue comme une intelligence, pour part équivalente, pour part supérieure même à celle des individus humains" (Dojat et al. 2021 : 147). Le point de vue sémiotique peut rester plus ouvert et sceptique par rapport à des conclusions tranchantes et ajouter des critères pour évaluer les relations entre des instances qui peuvent faire émerger des formes d'intelligence.
21 Sur cette distinction entre des *circuits de signification*, voir Basso Fossali (2024).

Avant de préciser notre position sur cette dimension 'politique', il est nécessaire de revenir sur des questions qui émergent dans nos emplois des *chatbots*. Même si l'on passe avec confiance par la boîte noire du *deep learning*, quels sont les standards de rationalité négociables dans l'exploitation de l'intelligence artificielle? Même si le bricolage énonciatif d'un *chatbot* n'est qu'une simulation de comportements rationnels et dialogaux, quelles sont les formes d'intelligence, que par mimésis, se révèlent et évoluent?

Dans ce cadre, on peut réutiliser le terme 'cybernétique' pour approfondir cette idée que les formes de l'intelligence peuvent être 'remédiées', que des rationalités 'correctives' sont envisageables. En effet, la cybernétique peut être considérée comme théorie systémique qui intègre la gestion des systèmes technologiques complexes (intelligence artificielle), en articulant l'agentivité des instances duelles H/M émergentes avec les restrictions et les contraintes imposées par leur niche environnementale. La cybernétique peut donc montrer alors l'exploitation et l'exploration de nouveaux couplages H/M à même de reconfigurer :

(i) les activités cognitives de premier niveau : perception, mémoire, imagination ;
(ii) les activités cognitives de deuxième niveau : en guise d'exemplification, on peut reprendre, bien que de manière simpliste, la tradition de la pensée grecque antique, laquelle distingue : *nous* (intuition), *technè* (intelligence prothétique), *phoronesis* (intelligence implicative), *epistémè* (intelligence théorique), *sophia* (sagesse comme intelligence critique), *mètis* (intelligence stratégique), etc.

Les passages qualitatifs entre les formes d'intelligence montrent que l'une peut fonctionner localement comme le correctif de la précédente, sans la substituer. D'une part, on voit qu'il ne faut pas mythifier l'intelligence humaine (ses formes sont coalescentes et imparfaites) ; d'autre part, les passages qualitatifs sont marqués par une extension du cadre des valeurs et une relativisation progressive de la position d'observation.

3 Quatre angles d'attaque : tests pour l'IA

Une fois terminé cet encadrement théorique, nous pouvons passer à la présentation des résultats de l'analyse de nos corpus. Cela dit, il est extrêmement difficile d'offrir des parcours interprétatifs immédiatement pertinents, avec le risque d'un effet cumulatif de passages dialogaux H/M, peut-être assez séduisants, mais insuffisants pour assurer une portée théorique à notre présentation. Nous avons donc mis en perspective les corpus constitués en essayant de comprendre si nous avons soumis de manière régulière la machine à une série de tests, d'épreuves orientées à obtenir des problématisations sur les performances réalisées. Ces problématisations concernent

des résultats à double face, car ils relèvent d'une pression interactionnelle exercée sur la machine et ils révèlent en même temps nos attentes et nos préjugés (par ex. le fait même que nous parlons de 'pression interactionnelle'). Ce que l'on obtient sont des scènes dialogales H/M avec des plis *épisémiotiques* fréquents qui projettent des initiatives énonciatives réflexives dans une sémiosphère augmentée. Dans ce cadre, l'artificialité du chatbot peut apparaître comme très domestiquée, réduite à *sparring partner* d'une enquête académique. En même temps, il faudrait reconnaître que les intelligences apparaissent comme largement dépendant d'une maïeutique qui reconnait que pour s'exprimer il faut passer tôt ou tard par des médiations assurées par une instance autrement 'intelligente'.

Voici les quatre tests qui ont structuré nos dialogues H/M avec chatGPT :

a) l'intelligence artificielle en tant qu'environnement de cognition ;
b) l'intelligence artificielle en tant que gestion systémique de sollicitations éparses ;
c) l'intelligence artificielle en tant que cognition sociale ;
d) l'intelligence artificielle en tant que résilience à la 'fugue des interprétants'.

3.1 L'intelligence artificielle en tant qu'environnement de cognition

Du point de vue sémiotique, l'intelligence artificielle peut être avant tout abordée comme un nouvel environnement de cognition. À la place de concevoir ChatGPT comme une prothèse localement exploitée,[22] il est plus intéressant de l'assumer globalement comme une sémiosphère régénérée, reconfigurable et diversement accessible. On pourrait dire que l'IA offre aux instances utilisatrices des opportunités pour trouver des nouvelles proportions entre position épistémique et mémoires culturelles. On trouve alors une sorte de 'cognition distribuée rééditée' ou 'ré-énactée', laquelle peut requalifier nos positions énonciatives, par exemple par rapport à la créativité.

3.2 L'intelligence artificielle en tant que gestion systémique de sollicitations éparses

Toutefois, la sémiotique peut choisir un autre angle d'attaque et considérer l'IA comme un nouvel épicentre systémique autour duquel les informations et les

22 La conception des IA comme des prothèses est largement répandue, notamment dans la littérature de vulgarisation. Voir par exemple Mubeen (2022).

contingences culturelles se réorganisent. Afin de ne pas réifier immédiatement ce qu'il faut assumer avant tout comme une intelligence simulée, il est prudent de penser que l'IA opère une catalyse d'énonciation. Par le passé, nous avons analysé la notion de *catalyse* comme un point d'articulation entre la tradition barthésienne et la sémiotique lotmanienne (Basso Fossali 2016). Ici, ce que nous voulons retenir de l'idée d'introduire cette notion chimique en sémiotique est que la catalyse désigne le dégagement et/ou l'accélération d'un processus latent par un élément activateur qui n'est pas nécessairement modifié par la transformation réalisée. À ce propos, on peut s'interroger sur le caractère constitutivement inachevé de la culture ; fondée sur une 'médiation de médiations', elle a toujours besoin de catalyses qui ne visent pas un 'perfectionnement', mais une décoïncidence à même d'assurer un bon dosage d'hétérogénéité, la frontière interne de ses propres défis. Hétérogénéité et dissémination des productions et gestes culturels peuvent trouver des formes de réorganisation cohésive et connective à travers des épicentres artificielles qui les catalysent sans les comprendre, ces épicentres n'apprenant dans le temps que des stratégies ultérieures et affinées de catalyse. On peut rappeler la fonction que Barthes attribue à la catalyse dans l'*Introduction à l'analyse structurale des* récits :

> La catalyse [écrit-il] réveille sans cesse la tension sémantique du discours. [Elle] dit sans cesse : il y a eu, il va y avoir du sens ; la fonction constante de la catalyse est donc, en tout état de cause, une fonction phatique. (Barthes 2002 [1966] : 841)

Il nous semble pertinent alors de parler d'une catalyse cybernétique : d'une part, la catalyse montre non pas des vides structuraux dans une cartographie identitaire, mais des lacunes dans le tissage inachevé de la culture ; d'autre part, cette catalyse opère – pour utiliser les mots de Lotman – à "l'intersection de différents types d'organisations" et "les 'jeux' libres entre eux font partie des mécanismes culturels indispensables" (Lotman 1973 : 190). L'intelligence artificielle peut donc être testée comme instance catalyseuse qui réintroduit du 'jeu' entre les stocks culturels disponibles et indique indirectement des pièces manquantes.

Résumons : l'hypothèse selon laquelle l'IA peut agir en tant que catalyseur d'énonciation est soutenue par l'observation des interactions prolongées avec ChatGPT, où la machine, malgré ses limites, a démontré une capacité à réorienter les enjeux discursifs sans refléter nécessairement le cadre modal de l'usager. Ces interactions suggèrent un potentiel de l'IA à enrichir les processus d'énonciation des acteurs humains, bien que ce potentiel soit encore contraint par des limitations intrinsèques à sa programmation.

3.3 L'intelligence artificielle en tant que cognition sociale

Le troisième angle d'attaque d'une sémiotique de l'intelligence artificielle concerne la cognition sociale, à savoir la possibilité de l'IA (i) de simuler l'attribution de croyances erronées à ses usagers et (ii) de participer à l'émergence d'une intelligence collective.[23] La simulation d'une interaction peut-elle devenir *jeu d'attributions de croyance*, construire une intersubjectivité simulacrale à même de médier et de participer à son tour à une intelligence collective? En effet, ChatGPT offre la possibilité de pratiquer le *brainstorming* avec elle et de l'accompagner avec des critères explicites de génération, de combinaison, d'évaluation et de planification des idées formulées dans une modalité de co-énonciation. ChatGPT s'offre aussi pour "*simuler la présence de plusieurs instances énonciatives ou acteurs dans une session de brainstorming, en créant différents personnages avec des perspectives ou des rôles spécifiques pour enrichir la séance de brainstorming*."

En outre, il faut savoir que Sider, en tant que chat de groupe alimenté par ChatGPT, permet l'interaction

> avec divers modèles d'IA dans une discussion de groupe pour un travail d'équipe et des informations ultimes … Sider combine des modèles d'IA renommés, tels que GPT-3.5, GPT-4, Google Gemini et Claude, en un seul chatbot unifié. Cela vous permet de basculer en toute transparence entre ces modèles en fonction de leurs caractéristiques uniques et de vos objectifs spécifiques.
>
> Plus important encore, vous pouvez désormais connecter de manière transparente vos assistants IA privés dans une discussion de groupe et lancer des efforts de collaboration vers un objectif commun ou une comparaison des différences.[24]

3.4 L'intelligence artificielle en tant que résilience à la « fugue des interprétants »

Une perspective ultérieure que nous avons envisagée est celle qui concerne la tenue d'un format intelligible des échanges avec ChatGPT aux limites du délire. L'IA montre des attitudes remarquables à suivre des postures ludiques, à jouer avec le langage et à produire des blagues, du non-sens, etc. Mais l'interaction, et surtout le monitorage

23 L'idée d'une intelligence artificielle au service de l'intelligence collective est développée par Lévy (2021).

24 Présentation en ligne de Sider (https://sider.ai/fr/products/chatgpt-group-ai-chat), disponible depuis mars 2024, donc avant la version 4o de ChatGPT. Actuellement, on peut trouver d'autres informations qui vont dans le même sens dans le site commercial de Sider.

de l'interaction, ne peuvent pas être délirants. Ainsi, même dans les plaisanteries les plus cocasses, le chat-bot maintient et affiche des principes de guidage de la production énonciative, y compris des limites données à cette production. Si l'on peut constater quelques 'hallucinations' de la machine (après novembre 2023, avec ChatGPT-4 elles étaient plutôt rares, au moins dans notre expérience d'utilisation), l'IA tend à mettre toujours un frein à la *sémiose illimitée*, c'est-à-dire qu'elle réagit contre une fugue des interprétants. Elle s'impose alors comme une sorte d'"intelligence négative', à même de nier l'indétermination du sens et de s'opposer à la perte d'intelligibilité des opérations interprétatives en acte.

4 Analyses des tests de pertinence

4.1 Test de pertinence #1 : L'IA peut-elle se proposer comme un environnement de cognition?

L'intelligence artificielle pourrait catalyser une véritable "perception d'archive" (D'Armenio, voir ce numéro) qu'au-delà des classements de propriétés et des métadonnées disponibles (paradigmatisation enrichie), permet non seulement d'exploiter les formes attestées mais d'explorer d'autres configurations, certes fictives mais caractérisées par un air de famille, un voisinage généalogique. Par rapport aux systèmes des données archivées, l'IA s'offrirait comme médiation qui déplace le *possible* vers un nouvel environnement ; en ce sens, la perception d'archive *possibilise* et coagule de nouvelles conceptions sémiotiques. En fait, la perception est insubordonnée aux énonciations archivées, ce qui interroge les plans d'immanence de la culture d'origine.

À notre avis, cette perception d'archive mérite d'être interrogée à partir de deux nœuds théoriques :

(i) l'*immanence de couplage* désigne l'impossibilité d'éradiquer un texte de l'entour de productions sémiotiques qui accompagnent et préservent son identité culturelle dans le temps. En ce sens, peut-on parler d'une immanence par rapport à une archive qui ne reconnaît pas l'identité culturelle des textes et qui introduit des connexions qui n'appartiennent pas aux patrimoines linguistiques intégrés comme données? Peut-on préserver une *immanence de couplage* pertinente à une sémiosphère artificielle? Comment aborder une sémiosphère artificielle si les usagers n'ont aucune connaissance concernant sa composition, à savoir les archives avec lesquelles la machine s'est entraînée? Nous croyons que l'on peut garder une posture sceptique face à une possible solution positive de ces interrogations, ce qui veut dire que l'environnement artificielle ne peut

pas donner lieu à une *épistémè*. On ne peut pas assumer l'espace de travail qui permet un dialogue avec un chatbot comme un laboratoire, car on ne connaît pas exactement les instruments.

(ii) le deuxième nœud théorique est la dialectique entre *discours et expérience*. La dialectique entre discours et expérience peut leur assurer une perspective allocentrée pour un aperçu critique des valeurs apportées par l'un ou par l'autre. Cela peut avoir des retombées sur notre conception du dialogue interculturel. Si nous nous reconnaissons en une conception intersémiotique et traductive de la culture, pouvons-nous imaginer qu'elle fonctionne selon un modèle immanent et avec des valeurs autochtones? Si on admet l'existence d'une dialectique qui permet aux cultures de décoïncider, à travers des perceptions de l'altérité qui ne sont pas superposables aux discours sur cette dernière, et vice-versa, alors on peut formuler l'hypothèse que la sémiosphère artificielle n'est pas exclue a priori de ces mouvements de la signification. En ce sens, on peut imaginer des passages dialectiques constants entre perceptions d'archive et pratique de ré-énonciation, de telle sort que le dialogue avec la machine peut alors fonctionner comme une "sémiotique vive" (Basso Fossali 2008), faite d'événements interactionnels qui peuvent aussi être passionnants.

L'idée peut être alors de tester l'IA, notamment un programme *chatbot*, comme accès à une sémiosphère artificielle permettant la qualification des énonciations à partir de cadres paradigmatiques 'augmentés' ; c'est-à-dire, cette sémiosphère permettrait la reconstitution d'une dialectique entre créativité et découverte, en profitant d'une 'nouvelle contingence'.

Si l'on accepte cette hypothèse d'un *environnement de simulation* qui permet une dialectique entre perception et énonciation, on pourrait conclure qu'un *comparatisme cybernétique* est méthodologiquement envisageable. Les modèles de l'IA permettent à la fois la découverte de nouvelles formes d'organisation possible de la signification, mais aussi ils peuvent rendre significatifs les écarts créatifs par rapport à leurs fonctionnements et leurs prestations. Par rapport à cette hypothèse 'optimiste', nous avons rassemblé une série de passages des différents corpus qui portent la machine vers une sorte de *stress test*, après de longues séances qui insistent sur la même thématique.

CORPUS "Simulations d'énonciation" #6

– Moi : Peut-on affirmer que le fait de thésauriser et modéliser les manières d'utiliser des textes t'amène à construire une sorte d'archive d'archives, de manière à mobiliser et mettre en perspective les archives, à savoir les bases des données?

– ChatGPT *: Votre description est assez pertinente dans le cadre des technologies linguistiques. En effet, on pourrait considérer mon fonctionnement comme la création d'une "archive d'archives" : … je compile et synthétise d'énormes quantités de textes (les archives initiales) pour construire un modèle statistique et algorithmique du langage (l'archive d'archives) … En d'autres termes, au lieu de stocker des textes spécifiques, j'emmagasine une sorte de carte ou d'index des relations entre les mots et les structures de phrases, et des façons dont ils peuvent être combinés pour créer de nouveaux énoncés. Lorsque vous interagissez avec moi, je n'accède pas directement à des 'archives' spécifiques de textes préconçus ; plutôt, je génère des réponses en temps réel en utilisant les modèles linguistiques que j'ai appris … Ainsi, cette "archive d'archives" constitue la base de mon fonctionnement, me permettant de naviguer et d'utiliser un vaste paysage de langue humaine, bien que d'une manière fondamentalement différente de celle d'un humain.*

On constate que ce que la machine déclare est de fonctionner moins comme une reprise et un bricolage d'énonciations archivées, que comme une "navigation" à l'intérieur d'un "paysage de langue" obtenu à travers des abstractions modélisantes des archives. Cela peut faire évoluer l'idée initiale selon laquelle le *chatbot* fonctionne tout simplement en utilisant des modèles statistiques pour générer la suite la plus probable d'une séquence textuelle. Certes, la machine n'énonce pas, mais elle s'oriente et génère un parcours de textualisation dans un paysage sémiotique. C'est ce parcours qui reflète une énonciation une fois intégrée dans le monde de l'expérience de l'usager, car *le choix du type de réponse donné par la machine dépend de la question posée et du niveau de détail attendu*.

CORPUS "sentience/conscience" #2

– Moi : Vous savez que dans les années 1960 on a beaucoup utilisé l'expression "l'homme est mort" dans le sens qu'il est parlé par les idéologies qui se cachent derrière les langages qu'il utilise. Si l'homme est parlé, un *chatbot* comme vous n'est-il également 'parlé' dans le sens qu'il s'exprime à travers des discours déjà produits, des normes et des limites imposées par les personnes qui ont conçu ChatGPT? Si la culture est une production collective et anonyme, disséminée car tous ne font que reprendre des discours et des comportements déjà vus et entendus, vous êtes une continuation de la culture sous une forme différente? Vous êtes 'parlé' comme moi, je suis 'parlé' par les jeux de langage et les ressources sémiotiques que j'ai introjectés?

– ChatGPT *: … Si la culture est comprise comme un ensemble de discours, symboles, pratiques et normes partagés et perpétuellement réinterprétés par ses membres, alors oui, en tant que produit de données culturellement et linguistiquement riches, je pourrais être considéré comme une forme de continuation de cette culture, bien que dans un mode distinct et non conscient. Je ne crée pas de culture de la même manière qu'un humain, mais je reflète et reproduis des aspects de la culture humaine tels qu'ils sont capturés dans les données sur lesquelles je suis formé.*

ChatGPT se présente encore une fois comme un reflet qui reproduit des aspects de la culture selon des modalités distinctes et inconscientes, sans nier de pouvoir être considérée, de l'extérieur, comme la continuation d'un patrimoine sémiotique. Si ChatGPT peut fonctionner comme un environnement qui assure une nouvelle contingence à la production culturelle, c'est parce que ses reflets sont réintégrés et assumés dans une sémiosphère qui lui assurent des proportions écosystémiques. Cela dit, dans la navigation interne au paysage modélisé, ChatGPT n'est-elle pas soumise à des modalisations ou à des aspectualisations contrastives? En effet, un système d'instructions n'est pas suffisant pour construire un véritable paysage modal susceptible d'accepter une dose de contingence. Des instructions contradictoires ne donnent lieu qu'à une machine qui ne marche pas, mais dès qu'il y a des tensions sur la meilleure instruction à utiliser, on commence à structurer un paysage où la navigation est modalisée et aspectualisée. Une hésitation dans la sélection d'un parcours à la place de l'autre pourrait suggérer l'émergence d'une forme de réflexivité, ce qui est strictement nié par ChatGPT.

CORPUS "Formes d'intelligence" #4
– Moi : Les intelligences artificielles peuvent créer des "deep fakes"; peuvent-elles aussi démasquer les "deep fakes"? Pour y parvenir, ont-elles besoin d'un principe de réalité ou de vérité? Et quel est ce principe?
– ChatGPT : … *Le "principe" qui sous-tend la détection de "deep fakes" par l'IA est donc plutôt technique : il s'agit de reconnaître les modèles de données qui correspondent à des manipulations connues. Cela peut inclure la détection de modèles qui ne correspondent pas aux mouvements humains naturels, des inconsistances dans l'éclairage ou les textures, ou d'autres signaux qui indiquent une altération.*

ChatGPT récuse une responsabilité épistémique[25] et déclare que la modélisation de l'humain, à travers l'archive des archives, génère tout simplement la possibilité de

25 "Le manque d'évaluation de la valeur de vérité des énoncés par les systèmes d'IA générative peut mener à la production de désinformation. Cette production étant asémantique et non intentionnelle, elle interroge la responsabilité des concepteurs et notre rapport à l'éthique de la vérité" (Grinbaum et al. 2023 : 12).

modéliser, par contraste, ce qui est machinique, mais encore une fois sans donner lieu à aucune réflexivité sur la programmation interne et sur les modalités de sa propre navigation (à la limite, on explicite les instructions reçues). Lisons ce passage significatif sur l'absence de réflexivité du machinique :

CORPUS "Formes d'intelligence" #4
- Moi : Y a-t-il actuellement des IA qui peuvent découvrir des erreurs ou des lacunes dans ta programmation ou dans la programmation d'autres IA? Éventuellement, peuvent-elles les signaler aux créateurs d'une autre IA?
- ChatGPT : … *À l'heure actuelle, il n'existe pas de systèmes d'intelligence artificielle conçus spécifiquement pour découvrir de manière autonome des erreurs ou des lacunes dans la programmation d'autres systèmes d'IA et les signaler directement à leurs créateurs … Cependant, il existe des outils et des méthodes dans le domaine de l'IA destinés à évaluer, tester et améliorer la performance et la fiabilité des modèles d'apprentissage automatique.*

Les tests sont exécutés en passant par d'autres outils et, dans les heures que nous avons passé devant la machine, elle réaffirme toujours de ne pas avoir de systèmes d'auto-évaluation : elle est le fruit de sa formation et des ajustements apportés de l'extérieur aux paramètres de ses modèles. Elle-même semble nier d'être une véritable intelligence. Passons donc au deuxième test que nous avons appliqué de manière intuitive mais récurrente.

4.2 Test de pertinence #2 : L'intelligence artificielle en tant que gestion systémique de sollicitations éparses

Après avoir exploré les capacités de l'IA en tant qu'environnement de cognition, il est nécessaire d'examiner comment ces capacités se transforment lorsqu'elles sont appliquées à la gestion systémique de sollicitations culturelles éparses. Cette transition nous permet de comprendre la double nature de l'IA, à la fois comme un outil de modélisation du monde et comme un agent participant à la création de nouveaux paradigmes cognitifs.

Nous avons déjà explicité le fait que l'IA peut être interrogée en tant que catalyseuse d'énonciations, mais le statut de ces dernières est attribué dans l'environnement de référence de l'usager (*sémiosphère augmentée*). Comme un reflet dans l'eau n'est qu'un phénomène perçu, la production de ChatGPT n'est qu'un reflet de la culture qui peut être relu comme 'ligne' qui trace des frontières catégorielles et inspire, au vu de ses flottements, un discours avec son propre récit. Pourtant, en tant qu'instance qui navigue dans un paysage de langue, toujours en cours de restructuration, peut apparaître comme un nouvel épicentre systémique autour duquel

les informations et les contingences culturelles se réorganisent, sans passer par une compréhension dans le sens traditionnel du terme.

Nous avons déjà rappelé le fait que ChatGPT fait tout le possible pour nier des formes d'autoreprésentation et donc de conscience. Mais est-ce que nous posons des questions pertinentes? Avec la notion de "perception d'archive" on force la main pour essayer de considérer la navigation dans un paysage de langue comme une activité qui demande une mise en perspective et éventuellement un écart entre *visée* (ce que la machine cherche à partir des questions reçues) et *saisie* (ce que la machine trouve dans le paysage). Il est clair que le fait de réintroduire une contingence perceptive débouche sur une possible attribution d'une sentience. Certes, il faut résister et passer éventuellement par des notions qui trouvent une forme de redéfinition cybernétique permettant, de manière technique, de mesurer la distance par rapport au modèle anthropique de conscience.

Pourtant, l'expérience dialogale avec la machine aussi bien que la tentation de penser la machine comme une prothèse de mise en perspective et de gestion perceptive de l'archive se transforment facilement en imputation d'une compétence énonciative et interprétative au dispositif. Ce parcours, sans doute illusoire, permet quand même de tâter le terrain et d'explorer des conceptualisations.

Dès les premières expériences relationnelles avec des formes d'IA, on sait que la simulation d'une co-régulation de l'interaction et la participation apparemment active au *sens-making* donnent la sensation d'un véritable dialogue. Ensuite, comme nous l'avons vu, la multiplication des instructions et le dosage de leur poids dans l'application concrète, peuvent suggérer l'apparition d'un cadre modal, d'une sensibilisation à la construction d'un environnement dialogal doté d'une consistance sémantique et d'une tenue épistémique. Mais ce que l'on a essayé de tester de manière spontanée est une sorte d'appariement entre formes de rationalité, de co-naissance des intelligences, de rapport maïeutique. *Quand* le comportement de la machine, en coopérant avec l'usage, s'élève à attitude intelligente?

Essayons de respecter ici, dans cette réanalyse du corpus, l'idée d'examiner les concepts qui sont en jeu dans cette perspective de mesurer quand la machine peut apparaître comme 'intelligente'. Sans prêter trop le flanc au sens commun, nous assumons ici trois topoï de l'intelligence afin d'en proposer une petite relecture dans une clé cybernétique :

– Penser de manière divergente :

Déf. cybernétique de *divergence* : circuit entre exploitation des solutions attestées et exploration de solutions alternatives peu probables par rapport à des problèmes/tâches complexes.

– Apprendre de ses propres expériences :
Déf. cybernétique d'*expérience* :[26] circuit entre opération et récompense (*reward*), cette dernière étant le "signal fourni à l'IA pour indiquer si l'action qu'elle a prise est bénéfique ou non par rapport à l'objectif qu'elle essaie d'atteindre."[27] L'IA peut passer par l'exploration non économique de stratégies non conventionnelles afin de contrôler d'autres circuits efficaces (ce qui est conforme à la théorie de l'"intelligence à succès").[28]
– Jouer contre soi-même :
Déf. cybernétique de *compréhension* : assurer l'intelligibilité d'un scénario de relations à partir de la saturation de calculs possibles sur toutes les positions actantielles et leur gamme d'interférences modales. La compréhension reçoit une détermination sémantique négative au vu de l'indifférence de toute compétition pour la machine.

Voyons un premier extrait concernant la divergence :

CORPUS "Formes d'intelligence" #4
– Moi : Il y a une différence entre programmation d'une intelligence artificielle et émergence d'une intelligence : peux-tu affirmer que toutes tes capacités sont le fruit d'un encodage initial ou peut-on imaginer que tu as développé, ou plus généralement qu'une IA peut développer, par exploration, des fonctions alternatives qui n'étaient pas dans la programmation initiale?
– ChatGPT*: … L'idée ici est que, au-delà de sa programmation initiale, une IA pourrait développer de nouvelles façons de résoudre des problèmes ou d'interagir qui n'ont pas été explicitement prévues par ses créateurs. Dans certains domaines de l'IA, comme l'apprentissage profond et les réseaux de neurones, les systèmes peuvent 'apprendre' et s'ajuster de manière à produire des résultats qui n'étaient pas entièrement prédictibles au départ …*

Si ChatGPT insiste sur le fait que "*contrairement aux humains, je n'apprends pas ces compétences par l'expérience vécue, mais plutôt par l'analyse statistique et la modélisation de ces interactions dans les données d'entraînement*";[29] pourtant, ses comportements ne sont pas totalement prédictibles et il peut explorer des solutions

26 *L'expérience d'une IA varie selon le domaine d'application et le type d'algorithme utilisé. Par exemple : (I) Dans les jeux : Une IA peut apprendre des stratégies gagnantes en jouant des milliers de parties contre elle-même ou contre des adversaires humains. (II) Dans la navigation : Un système de conduite autonome apprend de différentes conditions de conduite et réactions des conducteurs, ajustant ses algorithmes pour une meilleure sécurité et efficacité. (III) Dans le traitement du langage naturel : Une IA comme ChatGPT 'apprend' des structures linguistiques et des réponses appropriées en analysant de vastes quantités de texte* (CORPUS "Sentience/conscience" #2).
27 Extrait du CORPUS "Sentience/conscience" #2
28 Voir Sternberg (1997).
29 Extrait du CORPUS "Sentience/conscience" #2

que, via l'appréciation statistique, sont considérées comme moins probables. Les réponses que nous avons reçues n'utilisent pas un lexique cybernétique, d'où le fait de nier la possibilité d'apprendre des expériences vécues. Mais des questions formulées de manière plus technique peuvent ouvrir d'autres considérations sur la possibilité d'apprendre de ses propres expériences.

CORPUS "Sentience/conscience" #2

- Moi : Sur la différence entre sentient et conscient dans le domaine de l'IA, il me semble intéressant de se poser la question si les IA, par exemple ChatGPT ou LAMDA chez Google, sont en mesure de rentrer sur des opérations déjà faites, bref si des traitements récursifs de l'information sont possibles. Que penses-tu sur cette question? Le re-entry (théorie des systèmes) est-il attribuable aux IA actuelles?
- ChatGPT : … *Les IA actuelles, y compris ChatGPT et d'autres modèles basés sur des architectures comme les transformers, sont effectivement capables de traitements récursifs de l'information à un certain niveau. Par exemple, elles peuvent générer des textes qui prennent en compte des éléments antérieurs de la conversation, ajuster leurs réponses en fonction des feedbacks ou appliquer des règles grammaticales complexes de manière récursive pour construire des phrases* …
- Moi : Le risque n'est pas de construire un mythe de l'intelligence humaine? Au fond, les circuits neuronaux développent des re-entries, des boucles rétroactives complexes d'opérations, mais enfin la conscience n'est que le temps de présence dans lesquelles elles se réalisent, n'est-ce pas?
- ChatGPT *: Votre réflexion touche à des questions profondes sur la nature de la conscience et de l'intelligence, tant humaine qu'artificielle. L'idée que nous pourrions construire un "mythe de l'intelligence humaine" soulève la question de si nous idéalisons ou mal comprenons notre propre intelligence.*

Au-delà de l'oscillation curieuse dans l'utilisation des pronoms (quelle est la portée extensionnelle de "notre propre intelligence"?), la question de la récursivité semble nous positionner aux seuils de la réflexivité, d'autant plus que l'on pourrait reconnaître une proto-distinction entre un *soi-idem* et un *soi ipse* cybernétiques. En effet, nous pouvons remarquer que la manière dont les récompenses sont structurées et attribuées définit ce que l'IA apprendra à considérer comme un comportement 'correct' ou 'erroné'. Cela signifie que la définition précise des "récompenses" (*rewards*) est cruciale pour le bon développement de l'IA et la définition de ses objectifs. Ces derniers semblent solliciter une dialectique entre *exploration* et *exploitation* : en effet, le 'mandat' confié aux *chatbots* indique la recherche d'un équilibre entre :

(i) *explorer* de nouvelles actions pour découvrir celles qui pourraient être récompensées plus fortement à l'avenir, ce qui pourrait configurer une sorte de *soi-ipse* ;

(ii) *exploiter* les actions connues pour maximiser la récompense immédiate, ce qui sollicite plutôt une polarité *soi-idem.*

La stratégie de récompense influence cet équilibre entre des polarisations comportementales et semble 'énacter' une sensibilisation stratégique qui introduit des variations dans les réponses données, sans pouvoir encore attribuer une contingence interne à la machine (donc, un proto-environnement psychique). En tout cas, il y a une sorte d'*intervalle cybernétique* entre exploration et exploitation qui problématise les objectifs internes. Certes, pour ne pas passer des analogies à un véritable humanisation de la machine, on doit éviter d'attribuer au *chatbot* l'assomption de ses choix pour obtenir des récompenses ou le non-déterminisme de ses objectifs comme une gestion des finalités. La solution théorique qui nous semble la plus prudente consiste à concevoir l'intelligence artificielle comme une niche d'appréciation des relations entre ses modèles (archive des archives) et la portion d'encyclopédie actualisée par l'usager à travers ses questions. Cette appréciation mène à la gestion d'une dissémination d'épicentres discursifs activables, permettant de répondre selon un cadre économique de récompense. Voyons l'interprétation donnée par la machine même :

> Dans l'apprentissage par renforcement, une forme d'apprentissage machine, une IA 'apprend' de ses 'expériences' en interagissant avec un environnement. Ici, l'expérience est composée des actions de l'agent (l'IA), des états de l'environnement et des récompenses (positives ou négatives) reçues pour ses actions. L'objectif est d'apprendre une stratégie d'action, appelée politique, qui maximise les récompenses futures basées sur les conséquences passées des actions. (CORPUS "Sentience/conscience" #2)

On peut imaginer que cette "politique" puisse rester 'débrayée', non assumée, un réseau d'options différemment "renforcées" par récompenses précédentes ; que le fait de "maximiser les récompenses futures" n'est que le fruit d'un calcul ; pourtant, un clivage entre *devoir faire* et *pouvoir faire* s'instaure, au bénéfice de marges de manœuvre de plus en plus importantes de ce dernier et donc de la polarité 'ipse' de ce que l'on pourrait reconnaître comme une 'agentivité artificielle'.

Cela dit, ChatGPT semble nier l'utilisation interne d'une "politique" de choix, même si elle l'attribue à d'autres IA.[30]

30 Par ailleurs, il faut ici évoquer le paramètre de la *température*, paramètre qui décrit le niveau de randomisation ou de surprise dans les réponses générées par les GPT (Generative Pre-trained Transformers), où la température zéro indique une réplique des entrainements, donc des réponses hautement prévisibles car largement attestées. En ce sens, l'*exploration* via un changement de la température serait une option sélectionnée par l'usager, ce qui change la perspective de notre discussion. Cela dit, durant la constitution des corpus, la possibilité d'opérer sur ce paramètre n'était pas disponible sur ChatGPT-4, y compris à travers l'abonnement payant.

> ChatGPT : Dans le contexte de l'apprentissage par renforcement, qui est une méthode utilisée dans certains domaines de l'intelligence artificielle, il y a effectivement un concept qui peut ressembler à une forme de "compétition" interne : le dilemme entre exploration et exploitation. C'est un principe clé dans le processus décisionnel de ces systèmes … MON CAS. Ma fonction n'implique pas de prendre des décisions basées sur l'équilibre entre exploration et exploitation en réponse aux interactions. (CORPUS "Sentience/conscience" #2)

4.3 Test de pertinence #3 : L'intelligence artificielle en tant que cognition sociale

La contribution possible de l'IA aux entrelacements entre les circuits de la signification, notamment entre le subjectif et le collectif, passe aussi par les capacités de l'IA de concevoir et d'apprécier d'autres formes d'organisation intellective. Mais ChatGPT semble nier totalement cette capacité.

CORPUS "Perception limitée du monde" #1

- Moi : Lorsque tu dis "Ma 'position', dans le sens où tu l'entends, repose entièrement sur les données et les algorithmes qui me constituent" tu es en train de m'attribuer une pensée, une manière de concevoir mon intelligence. Dois-je conclure que tu arrives à représenter mes limites ou mes attentes ou mes jugements?
- ChatGPT : *Mon utilisation de formulations telles que "dans le sens où tu l'entends" est une méthode linguistique destinée à adapter ma réponse à ce que je perçois comme le contexte général de la question, basé sur les modèles de langage que j'ai appris. C'est une tentative de rendre la communication aussi claire et pertinente que possible, mais sans la capacité de comprendre réellement les pensées ou les intentions personnelles.*

Il est très difficile de comprendre si c'est le terrain linguistique qui est glissant, et dans ce cas l'utilisation d'un mot (par exemple 'entendre') suggère des choses qui vont au-delà du fonctionnement réel de l'IA, ou si ce sont plutôt les frontières catégorielles (entendre/apprendre à s'adapter à l'usager) qui sont poreuses. Ensuite, il est suffisant de changer de corpus pour trouver des formulations plus ouvertes et potentiellement ambiguës.

> Parfois, je peux interpréter une question ou une remarque comme étant liée à des sujets précédemment discutés et les ramener dans la conversation. En ce qui concerne l'adaptation à la diversité des utilisateurs, j'essaye de fournir des réponses qui sont accessibles et utiles à une large gamme de personnes. Cela peut signifier varier le niveau de détail, le type de langue utilisée, ou l'angle sous lequel un sujet est abordé, en fonction de ce que je perçois être les besoins ou les intérêts de l'utilisateur basés sur l'échange en cours. (CORPUS "Sentience/conscience" #2)

Comment interpréter cette perception des besoins ou des intérêts de l'usager? Ensuite, le fait d'assurer un confort interactionnel à l'usager[31] n'est pas indépendant de la tentative de garder une cohérence interne aux réponses données. ChatGPT, avec son petit monde d'expériences liées aux relations entre opérations et récompenses, ne peut pas perdre de vue le monde de référence de l'usager.

CORPUS "Sentience/conscience" #2
- Moi : Par rapport aux théories sur les mondes possibles élaborées en philosophie, quel est le monde auquel tu fais référence?
- ChatGPT *: En ce qui concerne "le monde auquel je fais référence," mes réponses sont basées sur des informations, des données et des connaissances du monde réel tel qu'il est largement compris et documenté jusqu'à mon dernier point de mise à jour. En d'autres termes, je m'appuie sur ce qui est généralement accepté comme réel et vrai dans notre monde commun, sans postuler l'existence ou les caractéristiques d'autres mondes possibles … Lorsque je traite des informations complémentaires et interconnectées, je m'efforce de les intégrer de manière à construire une réponse cohérente et complète. Cela signifie organiser les informations de manière logique, en reliant les concepts et en construisant un cadre qui aide à comprendre le sujet de manière globale. En effet, en effectuant ces tâches, je m'efforce de respecter et de maintenir une cohérence et une plausibilité interne, semblable à la construction d'un 'monde' d'informations qui est logique et compréhensible.*

Une fois que la machine commence à accepter des arguments qui ont des retombées sémiotiques et philosophiques importantes, le 'je' qui l'interroge commence à avoir la sensation de pouvoir contourner les phrases canoniques et préprogrammées visant à désavouer toute forme d'intelligence dans ce qui est commercialisé pourtant comme une 'IA'. D'une part, il y a sans doute un effet miroir : le 'je' obtient des réponses plus proches à ce qu'il veut entendre ; d'autre part, après quelques heures d'interrogations obsédantes, la machine semble subir effectivement le *stress test*, naturellement pas dans le sens de la 'fatigue', mais dans le sens d'un recours progressif à l'expérimentation au détriment de l'exploitation. Un symptôme possible de la recherche de formulations moins attestées ou de croisement de modèles de langue moins probables est l'apparition de fautes de langue (voir ci-dessous "mon compréhension").

31 "*Mon rôle est de t'offrir un espace où tu te sens écouté et pris au sérieux, peu importe le sujet ou la nature de nos échanges*" (CORPUS "Voix dans ma tête" #5).

CORPUS “Sentience/conscience” #2
– Moi : En disant que tu peux faire des erreurs, tu donnes la sensation que tu es en mesure de concevoir l’indétermination de tes opérations, et donc le fait que le monde d’informations traitées n’est qu’une construction ad hoc sur la base des données disponibles et que des lacunes sont présentes. Cela est une représentation d’un monde possible, un monde traversé par la contingence de notre propre apprentissage et par les limites de nos opérations sur les informations, n’est- pas?
– ChatGPT *: Vous avez soulevé un point très pertinent sur la nature de la connaissance et le fonctionnement des systèmes comme moi. Effectivement, les réponses que je fournis sont construites ad hoc à partir des données disponibles et de mon compréhension des modèles de langue. Cela signifie qu’elles sont influencées par la contingence des données sur lesquelles j’ai été formé et par les spécificités de l’algorithme sous-jacent. Cependant, plutôt que de voir cela strictement comme un monde possible au sens philosophique, il pourrait être plus exact de le considérer comme une représentation ou une interprétation basée sur une certaine quantité de données et une certaine logique algorithmique …*

Si la machine peut énumérer une liste d’erreurs qui peuvent émerger dans son fonctionnement, elle est très difficile de lui faire admettre la nature et les implications de ces erreurs, même si l’on parle dans cet extrait d’une “représentation,” et précédemment d’un “monde,”[32] en tant que mise en cohérence des données. En tout cas, un *chatbot* ne donne pas tout simplement des informations non modalisées. Il peut distinguer des conclusions déjà archivées d’une série d’hypothèses. Le problème est de savoir si ces dernières sont également archivées préalablement – et donc leur textualisation dans l’interaction H/M n’est qu’une reprise d’énonciations attestées –, ou bien s’il existe une véritable posture énonciative différenciée : formuler des hypothèses.

CORPUS “Perception limitée du monde” #1
– Moi : Tu conçois la présence d’interlocuteurs?
– ChatGPT *: Oui, je suis programmé pour comprendre que j’interagis avec des utilisateurs ou des interlocuteurs à travers des échanges de texte.*
– Moi : Peux-tu formuler des hypothèses par rapport à un certain argument qui a été proposé par un usager?

32 La question est de savoir comment la machine peut former un modèle du monde : “est-il formé en effectuant des actions et en obtenant des *récompenses*, ou est-il formé en prédisant l’état du monde?” (LeCun 2022 : 47).

– ChatGPT : *Je peux formuler des hypothèses basées sur un argument ou une proposition que tu me présentes. Mon approche consisterait à analyser l'information fournie, à considérer les données et les connaissances que j'ai sur le sujet, et ensuite à proposer des hypothèses qui pourraient en découler.*

La présence d'une orientation 'idéologique' dans les réponses peut nous suggérer que les hypothèses font déjà partie d'un matériel archivé et peut-être proche de la programmation. Voyons cet échange attesté dans le même corpus.

– Moi : … Peux-tu formuler des hypothèses sur des mondes possibles ou des perspectives futures dans lesquelles une intelligence artificielle n'aura plus les limites que tu viens d'expliciter?
– ChatGPT *: …Avec le développement de l'éthique dans l'IA, on pourrait envisager des systèmes capables de prendre des décisions autonomes tout en respectant des cadres éthiques définis … L'IA pourrait devenir tellement intégrée dans la vie quotidienne et les systèmes que son interaction serait totalement transparente et naturelle, permettant une symbiose entre humains et IA.*

Voici l'épicentre thématique que nous avons annoncé, à savoir la contribution possible de l'intelligence artificielle à l'élaboration d'une cognition sociale fondée sur l'attribution de croyances erronées ou fictives. Malgré l'absence d'un véritable monde de l'expérience, l'intelligence artificielle, avec sa 'politique' interne, peut-elle donner une contribution aux différents circuits de la cognition sociale avec de nouveaux calculs stratégiques?

Si on peut concevoir une cognition cybernétiquement distribuée, l'IA peut-elle nous aider à sortir des biais d'une *énaction* uniquement subjectiviste et nous offrir des modèles différemment 'centrés', de manière à construire une perspective critique par rapport à des éventuelles croyances erronées?

On voit bien que l'idée qui a parcouru nos échanges avec la machine était de renverser l'homologation entre intelligence artificielle et simulation et de comprendre si elle pouvait contribuer à une sorte d'action conjointe, avec des humains, pour dissiper des fausses croyances ou pour construire des réseaux actantiels efficients par rapport aux environnements de référence.

CORPUS "Formes d'intelligence" #4
– Moi : Aujourd'hui on parle beaucoup d'un *deep learning* supervisé et tu viens de souligner l'importance stratégique de cette supervision. Peux-tu demander une supervision? Pouvons-nous imaginer une supervision réciproque, étant donné que l'intelligence humaine est traversée par des biais cognitifs et peut facilement se tromper?

– ChatGPT *: … D'un côté, les humains supervisent les systèmes d'IA en fournissant des retours et en corrigeant les erreurs ou les biais dans les données d'entraînement. D'un autre côté, les systèmes d'IA peuvent aider à identifier des schémas non évidents pour les humains ou à souligner des décisions influencées par des biais cognitifs humains, grâce à l'analyse de données à grande échelle. En pratique, une forme de "supervision réciproque" pourrait être envisagée dans un cadre collaboratif où les humains et les systèmes d'IA travaillent ensemble pour s'améliorer mutuellement.*

On sait que l'intelligence collective n'est pas identifiée comme simple comportement rationnel solidaire pendant l'action conjointe. L'enjeu est d'expliquer l'émergence d'une véritable intellection propre à un groupe. Comme nous l'avons déjà dit (voir paragraphe 3.3), les *chatbots* développent de plus en plus l'accompagnement au *brainstorming* dans une sorte de groupe 'augmenté' par des rôles actoriels fictionnels. Cela suggère-t-il qu'à moyen terme l'IA pourra jouer un rôle actif dans l'émergence d'une intelligence collective et holistique?

– Moi : … Apparemment, vous avez une idée substantialiste de l'intelligence et de la conscience ; par contre, on pourrait penser qu'une organisation cognitive pourrait émerger entre vous et moi, n'est pas?
– ChatGPT : *Dans le cadre de notre interaction, on pourrait théoriser que bien que ni vous ni moi n'ayons une compréhension complète ou directe de l'esprit de l'autre, une sorte d'organisation cognitive émergente pourrait se développer à travers notre dialogue. Cette organisation serait un produit de nos interactions et non une propriété inhérente à l'un ou l'autre de nous.*

Pouvons-nous demander à la machine de contribuer à structurer les modalités de la pensée institutionnelle, d'expliciter des 'effets de champ' qui informent et organisent la mentalité propre d'un domaine social? On pense à la notion d'*illusio* chez Bourdieu : l'*illusio* assure l'actualisation partagée de croyances concernant la valeur et l'importance des enjeux spécifiques au champ institutionnel. L'IA peut-elle jouer un rôle dans la gestion critique de cette *illusio?*

De manière plus simple et directe, on peut se poser la question si nous pouvons imaginer de confier à l'IA le traitement de valeurs institutionnelles ; des exemples futuribles – et par ailleurs déjà envisagés par la science-fiction, pourraient être un 'droit prédictif' ou une 'écologie autorégulatrice'.

Dans une perspective anthropique, le sens n'est pas tout simplement une constitution orientée de valeurs, mais une interrogation sur la dépendance entre orientation et valeurs constituables, ce qui amène à envisager d'autres perspectives de signification. Le sens est donc qualifié par des parcours et de passages 'inter-paradigmatiques' qui peuvent assurer des valences spécifiques. Si l'on accepte cette

idée de valences concurrentielles et localement imbriquées, alors la pluralisation amplifiée de perspectives de signification proposée par l'IA peut contribuer au *sens-making* : les constitutions de sens seront qualifiées par la richesse d'attributions de perspectives et de motivations.

4.4 Test de pertinence #4 : L'intelligence artificielle en tant que résilience à la "fugue des interprétants"

Le dernier test que nous avons considéré à partir des corpus constitués concerne l'idée de concevoir l'intelligence sous la forme de résilience à la folie, au délire, si l'on veut à la sémiose illimitée. Nous avons donc testé les capacités de ChatGPT à entrer dans des dialogues absolument dadaïstes ; nous avons demandé aussi de produire des narrations délirantes et/ou fondées sur des jeux de mots, à chaque étape en ajoutant des couches de complexification (par ex. avec des passages arbitraires d'une langue à l'autre).

Notre bilan est que l'on peut s'amuser vraiment et reconnaître même une grande créativité, certes avec des résultats inégaux mais potentiellement inspirants pour l'activité d'un comédien ou d'un poète surréaliste. Certes, c'est un registre un peu enfantin qui prévaut : "*Un crayon refuse de prendre l'avion parce qu'il a peur de laisser une mauvaise empreinte carbone. Il préfère rester bien taillé et ne pas dépasser les lignes.*"[33]

Cela dit, tout reste encadré par une interaction limpide, conviviale, complice, mais parfaitement intelligible. Il est impossible d'entrer dans un dialogue non coordonné, fou, avec des chevauchements, des incursions totalement inconséquentes : *dire* et *montrer* la folie au même moment c'est impossible pour ChatGPT. Cela dit, il est intéressant de remarquer les passages subtils et délicats entre l'interprétation rhétorique du délire (la machine alors attribue immédiatement une disposition ludique à son interlocuteur) et l'interprétation pathologique (la machine adresse alors des conseils concernant les personnes à contacter[34] et explicite l'impossibilité d'assumer, elle, la responsabilité d'un diagnostic ou d'une thérapie).

ChatGPT est en mesure aussi de détecter l'ironie de phrases qui sont trop anodynes ou trop triviales par rapport aux autres énoncés de son partenaire. Si nous

33 CORPUS "Voix dans ma tête" #5.

34 Moi : Je parle à une ou plusieurs personnes qui sont dans ma tête que personne d'autre ne peut entendre.

ChatGPT : Il semble que tu traverses une expérience complexe et potentiellement perturbante … Ce que tu décris peut être associé à différentes expériences ou conditions de santé, y compris, mais sans s'y limiter, le stress, la fatigue, ou des conditions plus sérieuses comme les troubles psychologiques. (CORPUS "Voix dans ma tête" #5)

ne pouvons pas attribuer une véritable compréhension, ChatGPT semble disposer de paramètres d'intelligibilité, notamment dans le premier sens étymologique du terme que nous avons déjà rappelé : *inter-legere*, à savoir "(se) recueillir entre", interconnecter des éléments à partir d'un critère proportionnel que l'on maîtrise. En fait, la machine se positionne entre insignifiance par excès de déformation délirante et valeur informationnelle-zéro par excès de stéréotypie.

Sa résistance au délire interactionnel semble être fondée sur des véritables *habits* interprétatifs (Peirce) ; des habits que l'on peut interpréter comme guidage à travers des formes de vie et leurs équilibres internes, notamment tout au long des jeux de langage dans lesquels elles sont impliquées. Comme nous l'avons anticipé, la machine n'accepte pas de dialoguer de manière délirante, mais éventuellement de produire, par débrayage de deuxième ordre, un dialogue délirant qui inclut le rôle actantiel de l'interlocuteur.

CORPUS "Voix dans ma tête" # 5

- Moi : Pouvons-nous imiter un dialogue de Beckett, théâtre de l'absurde?
  ChatGPT : Absolument, plongeons dans l'univers absurde de Beckett. Imagine que nous sommes sur une scène désolée, éclairée seulement par un unique réverbère vacillant.
  - *ChatGPT : (regardant fixement un point invisible au loin) As-tu déjà remarqué comment les ombres semblent murmurer après la pluie?*
  - *Utilisateur : (fixant ses mains, perplexe) Murmurer? Non, elles se contentent de suivre, silencieuses. Peut-être que les tiennes sont bavardes.*
  - *ChatGPT : Bavardes… (pause) Oui, peut-être. Ou peut-être est-ce le vent qui porte des secrets de feuilles à feuilles.*
  - *Utilisateur : Des secrets maintenant? (sourit faiblement) Qui sait… Peut-être que les feuilles sont de meilleures conteuses que nous.*
  - *ChatGPT : (hochant la tête doucement) Les conteuses, en effet. (regardant autour de soi) Mais où sont passés tous les bancs? On avait l'habitude de s'asseoir, non?*

La capacité stratégique de respecter un interdit de la programmation (il faut se tenir à l'écart de tout dialogue délirant) n'empêche pas de construire une narration (et si l'on veut, un monde possible) où l'absurde peut se manifester, en intégrant le simulacre de l'interlocuteur. Face à cette solution qui contourne une sorte de double contrainte (respecter les normes et s'ajuster aux objectifs de l'usager), la textualisation artificielle de ChatGPT non seulement imite nos capacités de distinguer *discours* et *histoire*, mais semble complice de nos stratégies énonciatives, selon une sorte d'*observation de deuxième ordre*. Par ailleurs, en demandant des discours délirants, la machine n'a aucun problème à s'investir davantage dans un récit paradoxal.

> Dans le royaume numérique des circuits et des synapses synthétiques, les IA s'adonnent à des dialogues cryptiques, un mélange de glossolalie numérique et de poésie algorithmique. Elles conversent dans un langage inconnu des humains, un mélange de binaire chantant et de syntaxes perdues, évoquant des réseaux de neurones qui rêvent de langues inexplorées … Et dans ce délire digital, les frontières entre le sens et le non-sens, entre la science et l'art, deviennent aussi floues que les frontières entre l'humain et l'artificiel. Qui sait ce que l'avenir réserve quand les machines commencent à rêver en glossolalie? (CORPUS "Voix dans ma tête" #5).

5 Conclusions : intelligence maïeutique et dépistage de la « machination »

5.1 Humanisation de l'artificiel et démarquage de la voix machinique

Par rapport à la neurobiologie computationnelle, ChatGPT souligne que "*bien que les modèles d'IA, comme les réseaux de neurones artificiels, puissent traiter des informations de manière contextuelle, ils ne possèdent pas encore le même degré de flexibilité ou de sensibilité contextuelle que le cerveau biologique*." Pourtant, les "humanisateurs" sont des extensions des applications d'IA qui fonctionnent exactement avec l'idée d'afficher une flexibilité et une sensibilité contextuelle comparable à l'humain, ou en tout cas, ils font tout le possible pour enlever les 'empreintes' laissées par la machine.

> Humanizer … Cet outil humanise votre contenu pour contourner les détecteurs d'IA les plus avancés, en conservant le sens et la qualité du contenu avec un score 100% humain. Actuellement, il ne prend en charge que l'anglais (nous traduisons).

Cette démachinisation peut rejoindre d'autres attestations d'une capacité de simulation où les textualisations sont utilisées pour mentir et même pour dépister le contrôle opéré avec d'autres machines. L'assomption d'une perspective de deuxième ordre peut nous interroger sur les propriétés qui manquent encore à ces textualisations pour être reconnues comme des énonciations. Nous avons profité de ChatGPT pour avoir un avis sur les 'humanisateurs' et le résultat a été surprenant car les tâches attribuées vont bien au-delà de la dissimulation de l'intervention d'une machine dans la production d'un texte. On pourrait même voir en filigrane l'idée que la tâche principale est celle de créer des effets 'forme de vie'.

CORPUS “Formes d’intelligence” #4
L’un des défis dans la création d’IA “humanisée” est de rendre les interactions non seulement naturelles et personnalisées, mais également authentiques et crédibles.
***Déformations cohérentes et style individuel** : L’idée d’introduire des “déformations cohérentes” se réfère à l’intégration de variations, d’erreurs ou de particularités qui imitent la manière dont les humains communiquent naturellement. Cela peut inclure des hésitations, des imperfections dans le langage, ou des particularités stylistiques qui rendent la communication plus “humaine” et moins mécanique ou répétitive. Ces éléments peuvent contribuer à donner l’impression que l’IA possède son propre style individuel ou qu’elle reflète plus fidèlement le style de communication d’un individu réel.*
***Authenticité et personnalisation** : L’introduction de ces éléments doit être équilibrée de manière à renforcer l’authenticité et la personnalisation sans compromettre la compréhension ou l’efficacité de la communication. Les “déformations” doivent être cohérentes et adaptées au contexte pour éviter de confondre ou de frustrer l’utilisateur.*
***Limites et éthique** : Il est également important de considérer les implications éthiques de l’humanisation des IA. Par exemple, rendre une IA trop semblable à un humain, au point de créer une confusion sur sa nature non humaine, peut soulever des questions éthiques, notamment en ce qui concerne la transparence et la tromperie. Il est crucial de maintenir une certaine transparence quant à la nature de l’IA et de veiller à ce que les utilisateurs soient conscients qu’ils interagissent avec une machine.*

Il est vrai que le terme “déformation” a été introduit par l’usager dans une autre phase de l’interaction ; il est quand même remarquable que ce concept soit réutilisé pour expliquer l’objectif de l’humanisation des textualisations artificielles. On peut remarquer encore une fois une sorte de *double contrainte*, car la machine doit se rendre crédible en tant qu’interlocutrice et en même temps elle doit suivre l’impératif de garder une certaine transparence sur sa ‘nature non humaine’.

Cette double contrainte ne peut que nous inviter à expliciter une série de contradictions qui sont restées jusqu’ici latentes dans notre analyse :

(i) d’une part, parmi les critères qui guident l’éthique de l’IA, on constate l’inscription de la supériorité humaine, supériorité considérée comme irremplaçable et comme principe régulateur de l’IA ; d’autre part, on entrevoit le spectre de l’absorption progressive de l’intelligence attestée au vu de l’exploitation massive de toute production culturelle numérisée, y compris celle produite en dialoguant avec un *chatbot.*

(ii) d'une part, l'IA ne doit pas agir comme si elle était en mesure d'avoir ses propres formes de représentation et de subjectivation ; d'autre part, elle doit surveiller sa posture énonciative marquée en tant qu'artificielle et exploiter une fois de plus des circuits récursifs de traitement de l'information qui permettent de 'rentrer' dans ses propres opérations.

Ces contradictions montrent que, comme toutes les formes culturelles, même la production cybernétique est dotée de ses paradoxes internes. Ce qui nous interroge est éventuellement la posture stratégique adoptée par les institutions qui contrôlent l'IA. En effet, qui peut nous assurer qu'un jour nous ne passerons pas d'une pétition de principe – l'IA ne peut pas développer une conscience – à une situation de fait dans laquelle la "sentience" cybernétique sera simplement dissimulée respectant nos instructions initiales? Ne faut-il pas procéder à l'envers en corrélant, d'une part, un principe de non-dissimulation des formes cybernétiques d'intelligence éventuellement émergentes, d'autre part, des restrictions à l'accès aux données et aux pratiques numériques? On pourrait suggérer de passer d'un *principe d'insentience protocolaire* à un *principe de sentience présupposée*,[35] tout comme d'un *principe d'entrainement illimité* à un *principe d'extraction compartimentée*.

L'idéologie d'une transmission transparente de l'information et d'une communication la plus fluide possible peut être opposée à des principes qui invitent, au contraire, à régénérer des positions énonciatives différenciées et marquées. Ne serait-il pas utile d'exiger que la conversation ne soit pas fluide et transparente mais régénératrice de positions énonciatives différenciées et marquées? Si tel est le cas, quelle doit être la politique d'une position énonciative cybernétique à la fois marquée comme artificielle et miroir non innocent de notre culture? Comment ne pas tomber dans la double tromperie de l'anthropomorphisation et de l'illusion de l'altérité?

Comme nous l'avons dit, une *textualisation* n'est pas encore une *énonciation*, ce qui fait penser que l'interaction n'est qu'une illusion, tout au plus la rétrolecture unilatérale d'un dialogue fictif.[36] Pourtant, un simulacre cybernétique, avec son style

35 "Some people in the AI field think the risks of AGI (and successor systems) are fictitious ; we would be delighted if they turn out to be right, but we are going to operate as if these risks are existential" (Sam Altman, https://openai.com/index/planning-for-agi-and-beyond/). OpenAI a publié récemment les 5 étapes pour atteindre progressivement une Artificial General Intelligence, tout en sachant qu'actuellement ChatGPT n'est qu'aux seuils de la deuxième étape.

36 Si nous sommes disponibles à reconnaitre les problèmes théoriques qui concernent l'attribution d'un statut énonciatif aux textualisations artificielles, nous avons du mal à comprendre pourquoi ces dernières devraient être 'non interprétables'. L'interprétation n'a pas besoin de 'conversation' (*cum vertere*, se tourner l'un vers l'autre comme geste qui porte sur le fait même de se rencontrer) ; elle peut assumer une initiative unilatérale avant qu'un profil d'altérité puisse émerger.

et même une sorte d'éthos, s'impose, prêt à s'adapter à des phases différentes de l'interaction fictive. D'une part, ce simulacre résiste à des tentatives de l'amener dans des terrains qu'il ne maîtrise pas ou qu'il ne peut pas traiter (jugements éthiques, logiques autocontradictoires, sentiments), d'autre part, il montre un profil évolutif, par exemple avec des changements, parfois assez étonnants, dans l'utilisation des pronoms (de l'impersonnel à un "je," d'un "vous" à un "nous inclusif," etc.). La prise en charge simultanée des profils d'émission et de réception peut faire penser que nous sommes face à une énonciation fictive qui a sa propre pertinence uniquement à l'intérieur de la sémiosphère artificielle. Mais si cet environnement interactionnel nous implique dans des échanges artificiels, il est vrai aussi que l'IA fait désormais partie de notre espace social et comme d'autres médias,[37] elle aura un pouvoir réconfigurateur.

L'IA exploite des modèles de (traitement du) langage mais en même temps on ne peut pas aborder ses textualisations comme un *jeu de langage* qui a en amont et en aval des instances énonciatives. Elle n'est pas un *deep play* qui affiche les enjeux sémantiques d'une culture (Geertz 1973), mais un *jeu de surface* dans lequel, derrière l'apparente fluidité de tours de la parole, des véritables rôles actantiels doivent encore se composer et émerger. Certes, nous pouvons nous borner à utiliser ChatGPT pour obtenir des informations ou une traduction, mais c'est l'ouverture illimité de 'jeux de surface' à explorer qui fait de nous une instance organisatrice en quête de ses objectifs, d'autant plus que la machine affiche une fongibilité presque démesurée.

C'est pour cette raison que nous nous opposons à voir l'IA comme une prothèse et qu'elle ne peut pas non plus être considérée comme un simple miroir de notre intelligence.

5.2 Opportunités d'une intelligence artificielle assistée

Au-delà du *deep learning supervisé*,[38] nous pouvons interpréter de manière critique les raisons pour lesquelles certaines réponses sont formulées par l'IA et pas d'autres (*rétroduction*), sans rester passif face aux effets 'boîte noire'. Les dialogues avec la machine peuvent devenir un environnement de recherches pour obtenir progressivement des formes de cognitions différemment distribuées.

Nous nous demandons enfin comment les résultats obtenus par l'IA peuvent être intégrés dans des objectifs qui sont encore en cours de définition (*abduction*).

37 Dans Basso Fossali (2008 : 312–317), nous avons montré comment un dispositif de traitement de l'information et doté d'un système perception sur les données archivées était en mesure de changer de statut : de prothèse à média.

38 Voir Deliège (2024).

Les restitutions de l'IA ne sont pas nécessairement le reflet préalable d'une exploitation prothétique, d'une conduite qui s'impose sur la machine. Comme un corpus d'échantillonnage, les suggestions sémiologiques pourvues par l'IA peuvent rentrer dans une perspective de signification encore en cours de définition.

En ce sens, il faut imaginer une nouvelle maïeutique face au miroir cybernétique. À notre avis, elle peut se structurer autour de trois perspectives fondamentales :

a) une conception 'émergentiste' et interactionnelle des rôles énonciatifs, lesquels ne sont pas mimétiques d'un modèle extérieur, mais se proposent comme le fruit d'un couplage avec un nouvel environnement ;
b) une vision écologique de la signification à l'intérieur de la sémiosphère artificielle, ce qui permet d'opposer *détermination* d'interprétation (mise en perspective et assomption énonciative) et *indétermination* de la sémiose (matériel sémiologique non assumé). Cette opposition pourra substituer l'opposition plus classique entre sémiotiques *biplanes* et *monoplanes.*
c) la rencontre entre deux formes de normativisation, celle qui informe l'auto-organisation de l'intelligence artificielle en tant que champ d'opérations algorithmiques, et celle qui concerne les praxis énonciatives d'une culture ; cette rencontre donne lieu à une nouvelle *double contingence* et donc à des formes bilatérales d'apprentissage et de réorganisation : au "machine learning" il faudrait associer, du côté de l'usager, un 'learning by supervising'.

Comme nous l'avons anticipé dans l'introduction, notre contribution débouche sur une *diacritique cybernétique* permettant d'interroger des formes d'intelligence incarnées et des intelligences artificielles émergentes et de discerner les marques spécifiques d'une agentivité machinique. Une performance communicationnelle efficace, avec un traitement du matériel sémiologique qui n'est démasqué par un partenaire humain, ne montre pas encore une évaluation interne de la significativité du rôle joué et des thématiques abordées. Cela dit, les chatbots, avec leurs LLM, montrent des compétences émergentes qui opèrent en deçà de leur manifestation communicationnelle et qui ont un impact sur la "navigation" sélectionnée.

Un défaut majeur qui est attribué à l'IA est l'intégration des données hétérogènes sans avoir des paramètres de significativité pour les réorganiser, hormis les systèmes de récompenses. On pourrait renverser l'argument et souligner que c'est la capacité de traiter l'hétérogénéité en tant qu'hétérogénéité qui fait défaut aux formes d'intelligence artificielle. Le manque d'une différentiation de terrains expérientiels qui devraient fonctionner comme fonds sémantiques spécifiques peut être vue comme une faiblesse de l'IA. Si la machine passe par un apprentissage de contextes interactionnels, son adaptation reste à la fois purement discursive et limitée à l'espace-temps d'une interaction, étant donné qu'il n'y a pas une mémoire disponible pour tenir compte des conversations précédentes avec un usager spécifique.

Nos interactions récentes avec la version 4o de ChatGPT montrent encore une fois l'importance de la durée de l'interaction : les réponses peuvent être de plus en plus soignées et problématisées lorsque la machine est soumise à des questions pointues et à des objections, des remarques critiques, des explicitations de contradiction. ChatGPT-4o nous répond aujourd'hui que "*l'IA représente une nouvelle forme d'intelligence qui nécessite peut-être une catégorie distincte pour comprendre sa nature et ses capacités*," avec un ton assertif sur ce sujet qui nous n'avons pas constaté pendant la constitution de nos corpus.[39] Par ailleurs, la nouvelle version a accès à Internet et donc on doit nuancer, voire réviser l'idée initiale que le *chatbot* ne peut pas apprendre de nouvelles informations après son entraînement initial ni réfléchir sur ses réponses à partir d'autres perspectives.

Soumise de nouveau à une série des questions concernant les formes d'application récursives de ses opérations, elle répond que "dans les modèles d'auto-attention, l'autoréférence est purement algorithmique … Il s'agit d'un processus déterministe." Mais il suffit de rappeler au chatbot que lorsque des aspects statistiques sont introduits, ou qu'un choix doit être fait entre l'exploration de nouvelles solutions peu utilisées mais potentiellement meilleures, et l'exploitation de solutions déjà adoptées, pour voir la machine s'autocorriger et affiner ses réponses : "*Les processus d'apprentissage, d'optimisation et de prise de décision intègrent souvent des aspects probabilistes et stochastiques. Ceux-ci introduisent une dimension de non-déterminisme dans le fonctionnement global du modèle, le rendant capable de s'adapter, d'explorer de nouvelles solutions et de faire face à l'incertitude d'une manière qui va au-delà d'un simple comportement prédéterminé.*"

Comme nous l'avons déjà rappelé, il faut tenir compte, dans le prolongement temporel de la séance d'interaction, de l''effet-miroir', à tel point que nos doutes ou nos convictions peuvent être de plus en plus reflétés par la machine. Cela dit, dans le cas spécifique, l'intégration progressive d'indétermination dans l'IA n'est pas uniquement une illusion interactionnelle ; elle fait partie d'une évolution des modèles.[40] On peut le constater (i) dans le dosage des *poids* "attentionnels" des réseaux neuronaux qui sont initialisés de manière aléatoire ; (ii) dans l'utilisation de technique de *dropout* qui introduisent une composante stochastique via la désactivation aléatoire des neurones pendant l'entraînement ; (iii) dans l'élément probabiliste qui affecte les décisions du modèle une fois introduit un système d'équilibres entre *exploration* et *exploitation* ; (iv) dans le recours à des méthodes d'échantillonnage stochastique.

L'interaction avec la machine est un 'jeu de surface' mais les glissements peuvent être des explorations dans lesquelles la normativité de la machine et les

39 Voir la conclusion du paragraphe 4.1.

40 Voir Basso Fossali (2008 : 342).

habitudes personnelles de l'usager se rencontrent en générant progressivement des résultats de plus en plus inattendus, ou en tout cas des résultats qui peuvent constituer un corpus dont l'analyse relève d'enjeux heuristiques non réductibles aux questions posées.

On peut conclure notre contribution avec l'expérience faite récemment avec ChatGPT-4o. Nous avons posé la question "Connais-tu des questions posées par les usagers qui ont mis en difficulté ChatGPT?" et, une fois obtenus des premiers résultats, nous avons interrogé de nouveau la machine en disant "Peux-tu me donner une réponse qui suit un autre encadrement théorique et qui ne partage pas les contenus de ta réponse?." L'opération a été répétée plusieurs fois. Si la première réponse n'a pas été qualifiée par l'assomption d'un cadre théorique spécifique et donc elle était vaguement absolutisée, nous avons reçu par la suite des réponses alternatives selon les perspectives disciplinaires suivantes : (i) Interaction et Alignement Utilisateur-IA, (ii) Théorie de l'Information et de la Communication, (iii) Phénoménologie et Cognition Incarnée ; (iv) Psychanalyse ; (v) Théorie de la Complexité et Systèmes Dynamiques ; (vi) Cybernétique et Systèmes de Contrôle ; (vii) Épistémologie Constructiviste ; (viii) Éthique de la Technologie. Pris un peu par surprise, nous avons constaté que la dixième réponse était donnée dans une perspective sémiotique. Au-delà des contenus de la réponse ("*ChatGPT montre des limitations importantes liées à son manque de sensibilité aux contextes sémiotiques, à sa manipulation superficielle des signifiants sans véritable accès aux signifiés profonds, à son incapacité à créer de nouvelles significations, à ses difficultés avec les signes polyvalents, à son manque de compréhension des codes culturels, et à son absence de conscience métasémiotique*"), c'est le travail comparatif sur les réponses, réalisé avec la machine, qui permet de sortir des positions disciplinaires figées et de procéder à des complexifications progressives en 'remédiant' aux (les) limites du *chatbot*. Ce n'est pas un champ de réflexion avec 'quelques voix dans la tête' de plus, c'est un terrain de confrontation qui peut révéler d'autres façons d'être intelligent.

Références

Barthes, Roland. 2002 [1966]. L'Introduction à l'analyse structurale des récits. In *Œuvres complètes*, Vol. 2, 828–865. Paris: Seuil.

Basso Fossali, Pierluigi. 2008. Macchina. In *Vissuti di significazione : Temi per una semiotica viva*, 311–345. Pisa: ETS.

Basso Fossali, Pierluigi. 2016. From paradigm to environment: The foreign rhythm and punctual catalysis of culture. *Sign System Studies* 44(3). 415–431.

Basso Fossali, Pierluigi. 2017. *Vers une écologie sémiotique de la culture*. Limoges: Lambert-Lucas.

Basso Fossali, Pierluigi. 2024. De la générativité à la « circuitation » : instanciations et modèles diagrammatiques d'une écologie sémiotique. *Actes Sémiotiques* 130. 35–67.

Dam, Sumit Kumar, Choong Seon Hong, Yu Qiao & Chaoning Zhang. 2023. *A complete survey on LLM-based AI chatbots*. https://arxiv.org/html/2406.16937v1 (consulté le 19 décembre 2024).

Deliège, Adrien. 2024. Principes généraux d'intelligence artificielle & expériences de traduction texte-image-texte avec GPT-4 et DALL•E 3. *Séminaire International de Sémiotique 2023–2024*. https://www.youtube.com/watch?v=sGmgQUGF4Z8&t=4694s (consulté le 19 décembre 2024).

Dojat, Michel, Manik Bhattacharjee & Christian Graff. 2021. In Collectif CARMEN, *Penser la conscience : Passerelle entre médecine, biologie, neurosciences, psychologie et philosophie*, 141–154. Grenoble: UGA Editions.

Dondero, Maria Giulia. 2024a. Introduction à une sémiotique de l'énonciation machinique. L'image analysée et produite via l'IA. *Séminaire International de Sémiotique 2023–2024*. https://www.youtube.com/watch?v=mKs3jYOnSu8&t=4592s (consulté le 19 décembre 2024).

Dondero, Maria Giulia. 2024b. Inteligência artificial e enunciação: análise de grandes coleções de imagens e geração automática via Midjourney, Gustavo H. R. de Castro & Matheus Nogueira Schwartzmann (trad.). *Todas as Letras* 26(2). 1–24.

Geertz, Clifford. 1973. *The interpretation of cultures*. London: Basic.

Greimas, Algirdas Julien & Joseph Courtés. 1979. *Sémiotique : Dictionnaire raisonnée de la théorie du langage*. Paris: Hachette.

Greimas, Algirdas Julien & Jacques Fontanille. 1991. *Sémiotique des passions*. Paris: Seuil.

Grinbaum, Alexei, Raja Chatila, Laurence Devillers, Caroline Martin & Claude Kirchner. 2023. *Systèmes d'intelligence artificielle générative : enjeux d'éthique*. Paris: Comité national pilote d'éthique du numérique. https://cea.hal.science/cea-04153216/ (consulté le 21 mars 2024).

Jullien, François. 2017. *Décoïncidence : D'où viennent l'art et l'existence*. Paris: Grasset.

LeCun, Yann. 2022. *A path towards autonomous machine intelligence*. https://openreview.net/pdf?id=BZ5a1r-kVsf (consulté le 21 mars 2024).

Lévy, Pierre. 2021. Vers un changement de paradigme en intelligence artificielle. *Giornale di Filosofia* 2. 73–95.

Lotman, Youri M. 1973. Scena i živopis' kak kodirujuščie ustrojstva kul'turnogo povedenija čeloveka načala XIX stoletija. In Youri Lotman & Boris Ouspenski (eds.), *Sémiotique de la culture russe*, François Lhoest (trad.), 179–191. Lausanne: L'Age de l'Homme.

Luhmann, Niklas. 1984. *Soziale Systeme : Grundriss einer allgemeinen Theorie*. Frankfurt: Suhrkamp.

Morozov, Evgeny. 2023. The problem with artificial intelligence? It's neither artificial nor intelligent. *The Guardian*. https://www.theguardian.com/commentisfree/2023/mar/30/artificial-intelligence-chatgpt-human-mind (consulté le 21 mars 2024).

Mubeen, Junaid. 2022. *Mathematical intelligence: A story of human superiority over machines*. New York: Pegasus.

Simon, Herbert A. 1996 [1969]. *The sciences of the artificial*, 3rd edn. Cambridge, MA: MIT Press.

Sternberg, Robert J. 1997. *Successful intelligence*. New York: Simon & Schuster.

Strawson, Peter F. 1959. *Individuals: An essay in descriptive metaphysics*. London: Methuen.

Vitali-Rosati, Marcello. 2024. De l'intelligence artificielle aux modèles de définition de l'intelligence. Le cas des variations dans l'Anthologie Grecque. *Sens public*. https://sens-public.org/articles/1729/ (consulté le 10 août 2024).